Curve Fitting and Stock Price Prediction Using Least Square Method

Asep Juarna, Adi Kuswanto, Mohammad Abdul Mukhyi, and Raden Supriyanto

***Abstract*—**Although the stocks price, including its index, seemed changes in random way, actually in long enough periods it has certain trend. The objective of this research is to study the trend of stock prices changing by means of prediction or extrapolating an index of stocks preceded by curve fitting computation along some periods. Curve fitting computation uses 62 days data (in accordance with value interval of 341 - 365) of IDX30 stock index published by Indonesian Stock Exchange (IDX). Curve fitting computation uses the least square method; this computation gives optimal polynomial having degree of 58 with RMSE (root mean square error) of 0.400052 or approximately 0.1133% of median of the IDX30. Furthermore, this polynomial is used to predict the value of IDX30 on 19 days ahead; the value resulted value has giving RMSE (root mean square error) of 2.401567 or approximately 0.6625% of median of the IDX30.

***Keywords*—S**tock prices changing, curve fitting, polynomial, extrapolation, RMSE (root mean square error).

I. Introduction

STOCK is a share in the ownership of a company. Stock represents a claim on company's assets and earnings. As an investor acquires more stock, his or her ownership stakes in the company become greater [1]. However there are many investors keep the stock for profit taking purpose. The stock prices changes, up or down, is a very important indicator for the company management telling that something should (or should not) be carried out to preserve (or increase if possible) the company performance. For a profit taker, in the other side, this changing is a sign telling that it is the time to sell his/her stock or to buy more stocks in order to get more earnings from the gain capital.

There are some previous studies about prices or indexes prediction, especially the ones using computation methods. Accuracy of computation is showed by its MSE (mean square error) or the root RMSE (root mean square error). Kihoro and Okango [2] developed an ANN (artificial neural network) in order to predict the banking stock prices in Republic of Kenya with ANN training RMSE in interval [0.057628, 0.091336]; there is no information about ANN testing RMSE in their paper. Kaur and Mangat [3] improved the accuracy of SVM (*Support Vector Machine*) algorithm such and got the training RMSE in the interval of [0.907730, 0.687524] and the testing RMSE in the interval of [1.966749, 1.266228].

The objective of this research is to predict an index of stock in IDX (Indonesia Stock Exchange), namely IDX30, i.e., the stock prices index refers to the Indonesian Composite Index (*IHSG, Indeks Harga Saham Gabungan*) on December 30-th, 2004 which is valued of 100 [4]. The prediction computation is preceded by a curve fitting one based on least square method giving the optimal polynomial having certain degree in interval of [1,100] with minimal RMSE (*root mean square error*). This polynomial is developed based on stock prices in interval along 62 days over June until August 2012. Furthermore, this polynomial is used to predict or extrapolate an index value on 19 days ahead on September 2012.

II. Research Methodology

Given m pairs of data (x_1, y_1), (x_2, y_2), .., (x_m, y_m); *curve fitting* over that m data is a methode to get polynomial having degree of n, i.e., $p(x) = a_0 + a_1x + a_2x^2 + ... a_nx^n$, that across as close as possible to the m points of data. The measure of closeness is showed by the value of MSE (*mean square error*) or the root RMSE (*root mean square error*). There are some methods of curve fitting such as *Lagrangian polynomial*, *Newton divided difference*, and *least square*; this research uses *least square method*. To get a polynomial having degree of n, i.e., $p(x) = a_0 + a_1x + a_2x^2 + ... a_nx^n$, over m points, there are n+1 coefficients a_0, a_1, ... a_n must be determined from solving of the following system of linear equations [5]:

$$\begin{pmatrix} m & \Sigma x_i & \Sigma x_i^2 & \cdots & \Sigma x_i^n \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 & \cdots & \Sigma x_i^{n+1} \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 & \cdots & \Sigma x_i^{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Sigma x_i^n & \Sigma x_i^{n+1} & \Sigma x_i^{n+2} & & \Sigma x_i^{2n} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \Sigma y_i \\ \Sigma x_i y_i \\ \Sigma x_i^2 y_i \\ \vdots \\ \Sigma x_i^n y_i \end{pmatrix} \quad (1)$$

The detail of solving the system (1), then, is constructed as the following steps of research methodology:

Asep Juarna is with the Gunadarma University (corresponding author's e-mail: ajuarna@staff.gunadarma.ac.id).

This material is based upon research funded by Directorate of Higher Education, Department of Education and Culture, Republic of Indonesia, under Grant No. 187/K3/KM/2014.

1. Create the following table:

i	x_i	y_i	$x_i y_i$	x_i^2	$x_i^2 y_i$	x_i^3	$\cdots$	x_i^{2n}
1	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
2	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
m = 62	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	Σx_i	Σy_i	$\Sigma x_i y_i$	Σx_i^2	$\Sigma x_i^2 y_i$	Σx_i^3	$\cdots$	Σx_i^{2n}

2. Put the values of m, Σx_i, Σy_i , Σx_i^2 , and so on from the table into the matrix coefficient and the right-hand side of system (1). Solve the resulted system and get the n+1 coefficients of a_0, a_1, ... a_n of polynomial $p(x) = a_0 + a_1x + a_2x^2 + \ldots a_nx^n$.
3. For each value of x_i, i = 1, 2, ..., m, determine $z_i = a_0 + a_1x_i + a_2 x_i^2 + \ldots a_n x_i^{2n}$
4. Determine the MSE of curve fitting, i.e., $\text{MSE}_{\text{CF}} = \frac{1}{m}\Sigma(z_i - x_i)^2 = \frac{1}{62}\Sigma(z_i - x_i)^2$ for i = 1, 2, ..., 62, since in this result m = 62.
5. Based on test values of x_{63}, x_{64}, ..., x_{81}, determine the values of z_{63}, z_{64}, ..., z_{81}.
6. Determine the RMSE of testing, i.e., $\text{RMSE}_{\text{Test}} = \frac{1}{19}\Sigma(z_i - x_i)^2$ for i = 1, 2, ..., 19.

III. Result and Discussion

Research methodology, as is stated in the previous article, is implemented (encoded) in MATLAB language programming environment. To get the best degree of polynomial, that is the polynomial with minimal RMSE, the steps 3 and 4 of the research methodology, is repeated for degree 1 until 100. The computation result shows that the best polynomial is the one having degree of 58, with RMSE of of 0.400052 or approximately 0.1133% of median of the IDX30 as is shown in Figure 1.

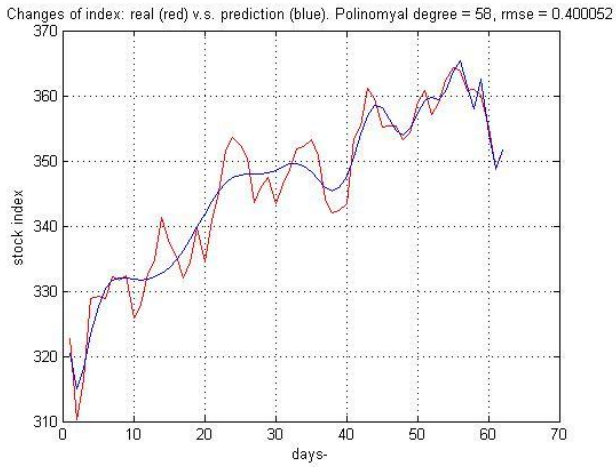

Fig. 1 Output of curve fitting computation

Extrapolation of the polynomial on the data test gives RMSE of 2.401567 or approximately 0.6625% of median of the IDX30 as is shown by Figure 2; it means that the accuracy of prediction is of 99.3375%.

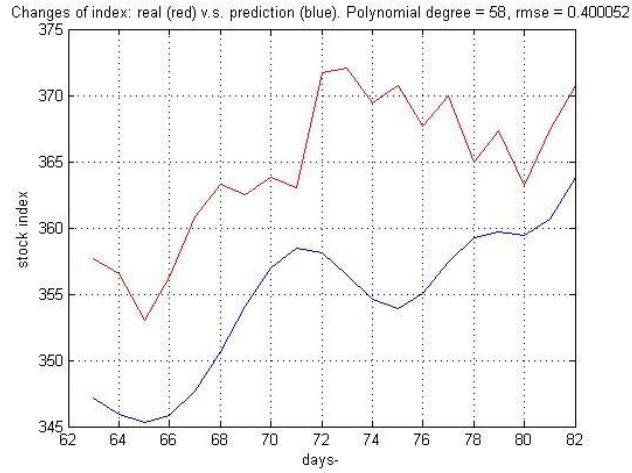

Fig. 2 Output of testing computation

IV. Conclusion

Computation of curve fitting has been performed on the data of IDX30 along 62 days hari over June - August 2012 resulting polynomial having degree of 58 with RMSE of 0.400052 or approximately 0.1133% of median over the interval of 62 days. Furthermore, testing (extrapolation) on the polynomial also has been performed on the data of 19 days ahead over September 2012 RMSE of 2.401567 or approximately 0.6625% of median over the interval of 19 days; it means that the accuracy of prediction is of 99.3375%.

References

[1] http://www.investopedia.com/university/stocks/stocks1.asp (downloaded at March 12, 2015)

[2] J.M. Kihoro and E.L. Okango, "Stock Market Price Prediction Using Artificial Neural Network: An Application to the Kenyan Bank Share Prices", *JAGST* Vol. 16(1) 2014.

[3] Savinderjit Kaur and Veenu Mangat., "Improved Accuracy of PSO and DE using Normalization: An Application to Stock Price Prediction", *IJACSA* (*International Journal of Advanced Computer Science and Applications*) Vol.3 No.9 2012.

[4] IDX, Indonesian Stock Exchange. *IDX Monthly Statistics June-September 2012*. Vol 21 No. 06-09, 2012.

[5] Francis Sched, *Numerical Analysis, 2-nd ed*, Schaum's Outline Series, McGraw-Hill – New York, 1989.