Nonlinear Autoregressive Exogenous Model in Stock Price Index's Prediction

Antoni Wibowo, Harry Pujianto, Dewi Retno Sari Saputro

Abstract**—**The stock market can provide huge profits in a relatively short time in financial sector; however, it also has a high risk for investors and traders if they are not careful to look the factors that affect the stock market. Therefore, they should give attention to the dynamic fluctuations and movements of the stock market to optimize profits from their investment. In this paper, we present a nonlinear autoregressive exogenous model (NARX) to predict the movements of stock market; especially, the movements of the closing price index. As case study, we consider to predict the movement of the closing price in Indonesia composite index (IHSG) and choose the best structures of NARX for IHSG's prediction.

Keywords**—**NARX, prediction, stock market, time series.

I. Introduction

In the capital market, it is well known that there are many offered forms and opportunities of investment offered. The investments may be in the form of debt securities, buying and selling foreign exchange, buying and selling precious metals, and other instruments. However, buying and selling activities on the stock market is one of the most interesting investments in the financial sector due to its capability to provide huge benefits in a relatively short time [1]. The great interest in the stock market cannot be separated from its ability to provide high profits in a relatively short [2]. However, the risks of this kind of investment are also very high for investors and traders if they are not careful to look the factors that affect the stock market.

The dynamic fluctuations and movements of the stock market will make a difficulty in determining the right time for making a transaction [1]. Therefore, the investors and traders need a proper prediction of stock market's fluctuations and movements to optimize their investments. For the investors and traders, predicting stock prices in future is an important activity in order to get a chance to transact early so that greater profits can be achieved [3].

Autoregressive moving average (ARIMA) and multiple linear regression (MLR) are two famous techniques in regression analysis. ARIMA is constructed based on a moving average, while MLR is developed based on ordinary least squares (OLS). Nevertheless, these techniques produce linear models that can be inappropriate in reality [4]. Besides, we must be careful with the presence of multicollinearity in MLR [5].

Instead of the above techniques, a NARX is conducted to predict the movements of stock market. NARX is a type of artificial neural network involving time series model regression. One of difficulties in NARX is how to select the best structure for prediction of a response variable. We usually perform the cross validation technique to select an appropriate regression model for the proposed prediction model.

In this paper, we select the best NARX and use it to predict the movement of the closing price in IHSG in which the original data are collected from Yahoo Finance. In these data, the variables of date, open price index, highest price index, lowest price index, and volume are assigned as regressor variables and variable closing is assigned as response variable, respectively. We assign the closing price index as our response variable since the price determines the performance of the fund manager, and as well, this number is followed by numerous investors. Furthermore, this is the most common price used by academic researchers [6].

The rest of the manuscript is organized as follows. Theory and method of the stock market and NARX are discussed in Section II. In Section III, the model selection of nonlinear time series regression models based on NARX is presented. Finally, we close this manuscript by conclusion and future work in Section IV.

II. Theory and Method

A. Capital Market

The capital markets are defined as "activities related to the Public Offering and Securities Trading, Public Companies, and those related to the Securities it publishes, as well as Securities-related institutions and professions" with respect to Capital Market Law No. 8 of 1995. In capital market, the stock price index becomes an important indicator to see the movement of stock prices. The function of the price index is as an indicator of market trends. Index movement describes the state of the market at a time. From the stock price index, we can know the trend of movement, whether it is going up, down, or stable. Index movement becomes an important indicator for investors to determine their action whether to sell or buy [7].

B. NARX

NARX is a time series regression model that developed through the linear autoregressive network with exogenous inputs (ARX) model. In NARX, the two tapped delay d_1 and

Antoni Wibowo is with Binus Graduate Program – Computer Science Department, Bina Nusantara University, Indonesia (phone: +62-21-53696969 ext 1803; e-mail: anwibowo@ binus.edu, a.wibowo2003@gmail.com).

Harry Pujianto is a master student at Binus Graduate Program – Computer Science Department, Bina Nusantara University, Indonesia (phone: +62-21-53696969 ext 1803; e-mail: harpujh@gmail.com).

Dewi Retno Sari Saputro is with Dept. of Mathematics, Faculty of Mathematics and Natural Sciences, Sebelas Maret University, Indonesia (phone: +62-271-669376ext 386/387; e-mail: dewiretnoss@staff.uns.ac.id).

d_2 are used to store previous values of $\mathbf{x}(t)$ and $y(t)$ sequences. The output of $y(t)$ sequence is considered a feedback signal, which is an input and also an output. Mathematically, NARX's model is given as follows:

$$y(t) = f(y(t-1), y(t-2), \cdots, y(t-d_1); \mathbf{x}(t-1), \mathbf{x}(t-2), \cdots, \mathbf{x}(t-d_{21})) \quad (1)$$

where f is a nonlinear function, $\mathbf{x}(t)$ is the input of NARX, $y(t)$ is the output and also feedback of NARX, d_1 is the feedback delay and d_2 is the input delay, respectively. From (1), it is evident that we should determine f, regression coefficients (or weights), the number of hidden neurons, d_1 and d_2 for the best NARX.

III. Model Selection

A. Original Data

The original data are collected from Yahoo Finance, in which Table I gives snapshot of the stock price index data [8].

TABLE I
The Snapshot of Stock Price Index in IHSG

Date (t)	Open Price Index (x_1)	Highest Price Index (x_2)	Lowest Price Index (x_2)	Volume Index (x_4)	Closing Price Index (y)
2011-01-03	3704.441	3738.589	3704.441	3.52E+09	3727.517
2011-01-04	3727.796	3763.383	3724.472	5.31E+09	3760.061
2011-01-05	3759.969	3786.055	3728.911	3.36E+09	3783.709
2011-01-06	3782.993	3789.473	3720.573	2.88E+09	3736.257
2011-01-07	3734.372	3734.372	3607.326	0	3631.453
2011-01-10	3631.267	3631.267	3449.58	4.55E+09	3478.549
2011-01-11	3478.079	3550.361	3434.27	3.84E+09	3455.127
2011-01-12	3455.409	3555.386	3455.409	5.04E+09	3554.766
2011-01-13	3555.33	3630.684	3549.008	4.33E+09	3564.937
2011-01-03	3704.441	3738.589	3704.441	3.52E+09	3727.517
2011-01-04	3727.796	3763.383	3724.472	5.31E+09	3760.061
2011-01-05	3759.969	3786.055	3728.911	3.36E+09	3783.709

B. Standardization Data

We perform the standardizing data by:

$$z_{new}^{i} = \frac{z_{\max} - z_i}{z_{\max} - z_{\min}} \quad (2)$$

with z_i is the i-th value of variable z in the original scale, $z_{\max}$ is the maximum value of variable z, $z_{\min}$ is the minimum value of variable z and z_{new}^{i} is the i-th value in the transformed scale.

C. Splitting Data

In our experiment, our data are divided into two subsets, namely learning data and evaluation data. The percentage of learning and evaluation data is 95% and 5%, respectively. The learning data is used to obtain the best NARX while the evaluation data is used to evaluate that the best NARX is valid in future data.

D. 10-Fold Cross Validation

Cross-validation (CV) is a statistical method that can be used to evaluate the performance of models or algorithms [9] using the above splitting data. We used *10-fold* CV, since it is recommended as the best model selection method [10]. In 10-fold CV, our data are divided into 10 folds of roughly equal size, says N_{cv} and we have 10 experiments to evaluate performance of models or algorithms. For each of 10 experiments, 10-fold CV uses nine folds for training and one fold for testing, respectively.

Let t_l^k and y_l^k (k=1, 2,3,…, 10 and l = 1, 2,, 3, …, N_{cv}) be the target value and the predicted value of the l-th data observation in the fold k. Then, the performance of models or algorithms can be estimated by the mean squared error of cross-validation (MSECV) as:

$$MSECV = \frac{1}{10}\frac{1}{N_{cv}}\sum_{k=1}^{10}\sum_{l=1}^{N_{cv}}(t_l^k - y_l^k)^2 \quad (3)$$

We use CV to choose an appropriate model by comparing the value of *mean squared error of cross-validation* (MSECV). The criteria of the best model or algorithm if it has the lowest value of MSECV compared to others. Moreover, let t_m^{val} and y_m^{val} be the target value and the predicted value of the validation data, and M be the number of the validation data (m=1, 2, 3, …, M). Then, the mean square error (MSE) for validation of the best model is given as:

$$MSE_{val} = \frac{1}{M}\sum_{m=1}^{M}(t_m^{val} - y_m^{val})^2 \quad (4)$$

IV. Discussion

We evaluate the performance the several models of NARX using 10-fold CV. The summary of our experiment is provided in Table II, which shows that model 6 produces the smallest MSECV compared the others. Therefore, model 6 is selected as the best IHSG's prediction based on NARX.

It is noticed that we made comparison between the best model of NARX (model 6) and the others models of NARX (model 1 and model 7) using our validation data. It is noticed that MSE for validation of model 1, model 6 and model 7 is 7.1205e-05, 2.2595e-05 and 5.6937e-05, respectively. It means that model 6 is fitter compared to model 1 and model 7. Fig. 1 also shows that model 6 gives better the stock price index's prediction since the predicted values of model 6 are closer to original values compared to the others.

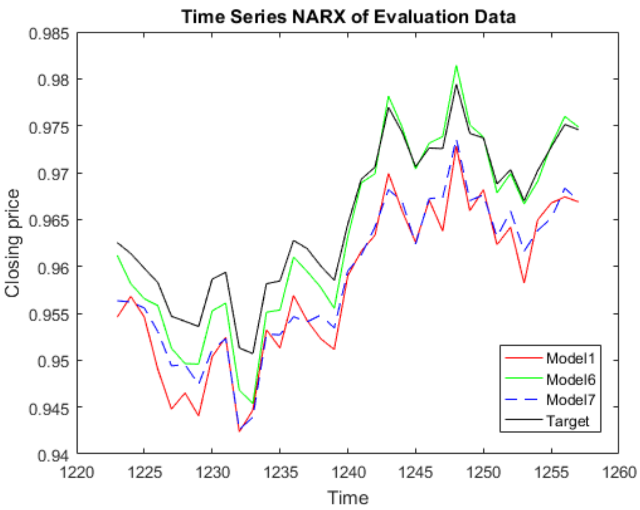

Fig. 1 Comparison of the best model (model 6), model 1 and model 7 for evaluation data

TABLE II
COMPARISON OF MSECV FOR SEVERAL NARX

Model	MSECV	Layer Size (Neuron #)	Feed-Back Delay (d_1)	Input Delay (d_2)
1	0.001289	5	3	2
2	0.001443	5	4	4
3	0.001366	5	4	6
4	0.001377	5	1	8
5	0.001574	5	3	10
6	0.001222	10	1	2
7	0.001350	10	2	4
8	0.001443	10	5	6
9	0.001454	10	1	8
10	0.001643	10	2	10
11	0.001289	20	2	1
12	0.001419	20	4	4
13	0.001577	20	6	2
14	0.001481	20	8	2
15	0.001708	20	10	4
16	0.001322	30	2	4
17	0.001332	30	4	3
18	0.001547	30	6	5

V. CONCLUSION AND FUTURE WORK

The main benefits of the stock market are that it can provide huge profits in a relatively short time; but, it also has a high risk for investors and traders. In this paper, we presented NARX as an alternative prediction model of the dynamic fluctuations and movements in IHSG. We selected several appropriate NARX models using 10-fold cross validation method in which model 6 is the best model prediction model in IHSG.

For our future works, we need to extend this work by employing hybrid of NARX and metaheuristic technique to increase the stability of the stock price index's prediction.

ACKNOWLEDGMENT

The authors would like to express a sincere gratitude to the anonymous reviewers for their valuable comments and suggestions to improve the quality of this manuscript. In addition, the authors would also like to thank Bina Nusantara University and Sebelas Maret University for supporting this research project.

REFERENCES

[1] R. Dash & P. K. Dash, "A Hybrid Stock Trading Framework Integrating Technical Analysis with Machine Learning Techniques," In *The Journal of Finance and Data Science*, Volume 2, Issue 1, March 2016, 42-57.

[2] M. B. Alam, M. Z. Hossain., A. M. Hossain, A. M. And M. M. Islam, "Price Prediction of Stock Market using Hybrid Model of Artificial Intelligence, "in *International Journal of Computer Applications*, 2015, *111*(3), 5–10.

[3] U. A. Umoh and U. G. Inyang, "A Fuzzy-Neural Intelligent Intelligent Trading Model for f or Stock Price Prediction," in *International Journal of Computer Science Issues*, 2015, *12*(3), 36–44.

[4] S. H. Arbain and A. Wibowo, "Neural Networks Based Nonlinear Time Series Regression for Water Level Forecasting of Dungun River," in *American Journal of Computer Science,* 2012, Science Publications.

[5] A. Wibowo and M. I. Desa, "Kernel Based Regression and Genetic algorithms for Estimating Cutting Conditions of Surface Roughness in End Milling Machining Process", in *Expert System with Applications*, 2012, Elsevier.

[6] P. Hillion and M. Suominen, "The manipulation of closing prices," in *Journal of Financial Markets*, 2004, 7, 351-375.

[7] T. Antolis and S. Dossugi, "Pengaruh Fluktuasi IHSG, Inflasi Dan Suku Bunga Terhadap Imbal Hasil Unitlink Berbasis Saham," in *Journal of Applied Finance and Accounting*, Binus Journal, 2008, *1*(1), 141–165 (in Bahasa).

[8] https://finance.yahoo.com/quote/^JKSE/history?period1=1343062800&period2=1488214800&interval=1d&filter=history&frequency=1d (Accessed: 1 July 2017).

[9] S. Cheng and M. Pecht," Using cross-validation for model parameter selection of sequencial probability ratio test," in *Expert Systems with Applications,* 2012, 39. pp. 8467-8473.

[10] P. Refaeilzadeh, L. Tang and H. Liu, *Cross-validation*, 2008, Arizona State University.

Antoni Wibowo has received my first degree of Applied Mathematics in 1995 and master degree of Computer Science in 2000. In 2003, He awarded a Japanese Government Scholarship (Monbukagakusho) to attend Master and PhD programs at Systems and Information Engineering in University of Tsukuba-Japan. He completed the second master degree in 2006 and PhD degree in 2009, respectively. His PhD research focused on machine learning, operations research, multivariate statistical analysis and mathematical programming, especially in developing nonlinear robust regressions using statistical learning theory. He has worked from 1997 to 2010 as a researcher in the Agency for the Assessment and Application of Technology – Indonesia. From April 2010 – September 2014, he worked as a senior lecturer in the Department of Computer Science - Faculty of Computing, and a researcher in the Operation Business Intelligence (OBI) Research Group, Universiti Teknologi Malaysia (UTM) – Malaysia. From October 2014 – October 2016, he was an Associate Professor at Department of Decision Sciences, School of Quantitative Sciences in Universiti Utara Malaysia (UUM). Dr. Eng. Wibowo is currently working at Binus Graduate Program (Master in Computer Science) in Bina Nusantara University-Indonesia as a Specialist Lecturer and continues his research activities in machine learning, optimization, operations research, multivariate data analysis, big data/data mining, data science, financial engineering, computational intelligence and artificial intelligence.

Harry Pujianto has received her first degree of Computer Science in 2014 from Bina Nusantara University. He is currently attending master programs at Magister in Computer Science – Bina Nusantara University. His master research focuses on machine learning, artificial intelligence and financial engineering.

Dewi Retno Sari Saputro is a lecturer in Department of Mathematics, Faculty of Mathematics and Natural Sciences – Sebelas Maret University. She holds a first degree in Mathematics from Sebelas Maret University and MSc in Computer Science Gajah Mada University, respectively. She has obtained her Ph.D. in Multivariate Statistics from Bogor Agricultural University. Her areas of research interests are time series regression, optimization, and modelling.