

The use of artificial neural network and multiple linear regressions for stock market forecasting

¹Abdu Masanawa Sagir and ²Saratha Sathasivam

^{1,2}Universiti Sains Malaysia
School of Mathematical Sciences, 11800 Gelugor, Penang, Malaysia
e-mail: ¹amsagir@yahoo.com, ²saratha@usm.my

Abstract In the recent economic crises, one of the precise uniqueness that all stock markets have in common is the uncertainty. An attempt was made to forecast future index of the Malaysia Stock Exchange Market using artificial neural network (ANN) model and a traditional forecasting tool – Multiple Linear Regressions (MLR). This paper starts with a brief introduction of stock exchange of Malaysia, an overview of artificial neural network and machine learning models used for prediction. System design and data normalization using MINITAB software were described. Training algorithm, MLR Model and network parameter models were presented. Best training graphs showing the training, validation, test and all regression values were analyzed.

Keywords Artificial neural networks; multiple linear regressions; stock market forecasting

2010 Mathematics Subject Classification 62J05; 62M10; 92B20; 97R20

1 Introduction

Stock market is a public entity for the trading of company stock (shares) and derivatives at an agreed price. Stock market forecasting or prediction is a very important aspect of economy that requires consideration due to re-occurring challenges in the stock market. Stock markets have become an integral part of the global economy. Personal and corporate financial lives and the economic health of a country are influenced by fluctuations in the market. Studies have shown that forecasting direction as compared to value can generate higher profit Chen *et al.* [1].

It is observed that there is need to develop efficient forecasting models to help the stock market strive better. This is necessary because of the uncertainties involved in the movement of stocks. Many factors affect the stock market including trader's expectations and general economic recession or conditions. Therefore, forecasting or predicting market price movements is very necessary Choudhry and Garg [2].

In this paper, we use artificial neural network in Matlab software and multiple linear regressions in Minitab version 17 for Malaysia Stock Market Forecasting. With the use of ANN and MLR, we were able to examine and analyzed the data which were collected from Bursa Malaysia. The data was pre-processed and trained. The ANN model developed was used for forecasting. The ability of the artificial neural network to forecast future trends of stock market indices was tested. The results were compared against a traditional forecasting method, multiple linear regressions for accuracy. The results from the experiment are represented in tabular form. Mean Squared Error (MSE) and Coefficients of determination (R^2) were used to determine the best performance model.

The remaining section of this paper are organized as follows: in Section 2, an overview of artificial neural network, machine learning models as well as the stock market exchange models for forecasting were presented. In Section 3, materials and methods of this research

work were described. This led us to Section 4 in which experiment and results were recorded and analyzed. Section 5 concludes the work.

2 An overview of artificial neural network, machine learning models and stock market models for forecasting

First we have to look at the Stock exchange which is a form of exchange that provides services for stock brokers and traders to trade stocks, bonds and other securities. Securities traded on a stock exchange include shares issued by companies, unit trusts, pool investment products and bonds. In other words, a stock exchange is an institution, organization or association that serves as a market for trading financial instruments such as stocks, bonds and their related derivatives.

Stock exchange of Malaysia was established in 1964 with an agreement of currency interchangeability between Malaysia and Singapore but later on in 1976 both are divided into Kuala Lumpur Stock Exchange and Stock Exchange of Singapore respectively. On April, 14, 2004, it has changed to Bursa Malaysia Berhad which is one of the largest bourses in Asia listing 1000 companies' worldwide. It has divided the listed companies according to its size and growth like the high growth companies listed in Malaysian Exchange of Securities Dealing and Automated Quotation (MESDAQ) market, the large size companies listed in the Bursa Malaysia Securities main board and the medium size companies listed in second board. The vision of Bursa Malaysia is "To be the preferred partner in Asia for fund raising, trading and investment" Smith [3].

The FTSE Bursa Malaysia Index Series is a broad range of real-time indices, which cover all eligible companies listed on the Bursa Malaysia Main and ACE Markets. The series divides the market into size segments (Large cap, Mid cap, Small cap and Fledgling), and includes Shariah-compliant indices, as well as themed indices representing the palm oil plantation sector. FTSE Bursa Malaysia mid 70 index, FTSE Bursa Malaysia top 100 index, FTSE Bursa Malaysia small cap index, FTSE Bursa Malaysia fledgling index and FTSE Bursa Malaysia EMAS index; it helps all stock sizes within the market. The FTSE Bursa Malaysia KLCI (FTSEKLCI) is a benchmark stock index for Malaysia Trading [4].

During the revamp of the Malaysian Stock Exchange in August 2009, the ACE Market replaced the Malaysian Exchange of Securities Dealing and Automated Quotation ("MESDAQ") Market and the Main and Second Boards unified to become the Main Market which consists of established companies with strong track records, the ACE Market facilitates the listing of emerging companies.

2.1 Artificial neural network

Artificial neural network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. It is divided into feed-forward and recurrent neural networks. ANN can be termed as a computer forecasting technique that originate from the work on artificial intelligent (AI) specifically aim to mimic the human ability to adapt to changing circumstances and the current environment. The general artificial neuron model has five components, which consists of input layer, weights schemes (W), threshold (u), transfer (activation) functions (f) and the layer output (y),

which depend on the activation function used:

$$y = f \left(\sum_{j=0}^n W_j x_j - u \right). \quad (1)$$

2.2 Machine learning models for stock forecasting

Stock market forecasting has been investigated by several researchers with different models and methods. They include:

Support Vector Machine (SVM) is a new technique within the field of computational intelligence which has recently been applied to many fields of studies. It is used for handwriting recognition Cortes and Vapnik [5]. However, Lean *et al.* [6] used Least squares support vector machine (LSSVM) for stock market trend to select input features for LSSVM learning. Support vector machines can easily be extended to the task of regression and time series prediction.

Particle Swarm Optimization (PSO) is one of the techniques widely used for the prediction of stock market movement. Particle swarm optimization is a population based stochastic optimization technique that was developed by Eberhart and Kennedy [7]. Also in their book, it described many philosophical aspects of Particle Swarm Optimization, Swarm Intelligence and their applications Poli *et al.* [8].

2.3 The stock market models for forecasting

There are several models used by various researchers for forecasting stock market. Stock market is defined as a mechanism that allows people to buy and sell financial securities and commodities, to facilitate the raising of capital markets Pilbeam [9].

Stock market is defined as a complex, non-stationary, noisy, chaotic, non-linear and dynamic system but it does not follow Random Walk Hypothesis (RWH) Lo and MacKinlay [10].

Forecasting of stock market is based on the finance hypothesis known as Efficient Market Hypothesis (EMH), which implies that the financial markets are “informationally efficient”. The Efficient Market Hypothesis states that all the information relevant to a market is contained in the prices and each time that new information arise; the market corrects itself and absorbs it. However, the prediction of movement of stock price will help investors to know when to invest, predict stock prices and index changes and to analyze price patterns.

3 Materials and methods

The details of how the data obtained from the Malaysia stock exchange market (Bursa Malaysia) for normalization, system design and the proposed model have been described in this section.

3.1 Data normalization

The values were rescaled, usually they are so scaled such that the time series data set has mean zero (0) and variance one (1). The data from Bursa Malaysia daily returns was pre-processed using normal transformation of a normal curve ($0 < x < 1$) using MINITAB

version 17 software, and was analyzed so that there are no leaking values (i.e. redundant stock data owing to technical suspension of the stock by Bursa Malaysia) and the pre-processed data was used for the experiment. The Normal curve equation is as follows:

$$\int_a^b f(x) = \int_0^1 e^{-\frac{1}{2}x^2} dx \quad (2)$$

where x is the standardized normal random variable.

The data set has been divided into two the training set and testing set. The training function used was Levenberg-Marquardt algorithm. The learning rate per layer and the training tolerance of maximum epoch size were set to be 0.001 and 1000, respectively

3.2 System design

The Malaysian Stock Market Data was collected from Bursa Malaysia based on daily basis return from April, 2015 to June 2015, having 63 days in all implying an initial sample size of $n = 63$. Artificial neural network model was trained with the data set collected for forecasting after pre-processing it.

3.2.1 Multi-layer perceptron network

A multi-layer perceptron (MLP) network is composed of several layers containing nodes. In the case of predicting the currency market, the inputs will be the past observations of the exchange market and the output will be the future value of the exchange market rate. Between the input and output, there can be one or more intermediate layers that are called the hidden layers. The main advantage of MLP is their ease of use and approximate of any input/output map.

3.2.2 Training algorithm

Training is a procedure where artificial neural network is adjusted to do a particular job. In other words, it is a process by which the free parameters of the networks (i.e. weights) get optimal values. There are several ways in which training can take place. Supervised, unsupervised and reinforcement learning models. For multilayer perceptron network, which uses supervised learning models, train certain output nodes to respond to certain input patterns and the changes in connection weights, due to learning, cause those same nodes to respond to more general classes of patterns.

The training algorithm of back propagation involves four stages:

- (i) Initialization of weights to small random values
- (ii) Forward pass:

$$S_i = \left(\sum_{j=0}^n W_{ij} a_j \right) \quad (3)$$

where S_i - the total input at layer i , W_{ij} - weight of the input from node j in layer i , a_j - activation function defined as: $a_j = f(S_i)$, where $f(S_i) = 1/(1 + e^{-x})$

- (iii) Backward pass:

This is the error back propagation and adjustment of weights. Gradient descent approach with a constant step length, also referred to as learning rate is used to train

the network. This method minimizes the sum of squares errors, contain many local minima.

The output of each processing unit for the backward pass will be defined as follows:

$$f' S_i = a_i * (1 - a_i) \quad (4)$$

- (iv) Updating of the weights and biases: Weights are then updated by the gradient descent method:

$$\delta_i = - \left(\frac{\partial \varepsilon}{\partial S_i} \right) \quad (5)$$

$$W_{ij}^* = W_{ij} + \alpha \delta_i a_i \quad (6)$$

where ε is the mean squares error (MSE) and α is the step size (learning rate).

3.2.3 Network parameter model

The topology used is 3-3-1, using MATLAB software package. ANN model performance was measured by the coefficient of determination (R^2) and the difference in the performance of the Mean Squared Error. The coefficient of determination is a measure of the accuracy of prediction of the trained network models. Higher (R^2) values indicate better prediction than lower values.

3.3 Modelling of the FTSEKLCI bursa Malaysia indices with ANN

The components of the daily Malaysia Stock market indices are as follows:

- (i) The FTSE Bursa Malaysia KLCI (FTSEKLCI) is a benchmark stock index for Malaysia. It consists of KLCI, FTSE Bursa Malaysia 100 and FTSE Bursa Malaysia EMAS. The ACE Market replaced the "MESDAQ" Market and the Main and Second Boards unified to become the Main Market which consists of established companies with strong track records. This shows the behavior of the shares transaction in the Malaysia stock market. The daily calculation shows the movement of the prices.
- (ii) The Market Valuation (*MarketVAL*) is also commonly used to refer to the market capitalization of a publicly-traded company, and is obtained by multiplying the number of its outstanding shares by the current share price.
- (iii) The Volume (*Vol@close*) is the volume traded at the close of the day transaction
- (iv) The Value (*Val@close*) is the value in Ringgit of the shares traded

The closing value (*Val@close*) is taken as a function of the following input variables for this model as: *FTSEKLCI*, *MarketVAL* and *Vol@close*, while the function of the variable is the *Val@close*. The following system model is considered for the forecasting of stock exchange market index value. Hence, the model equation is defined as:

$$Val@close = f (Vol@close, FTSEKLCI, MarketVAL) \quad (7)$$

Equation (7) is explicitly expressed in 3.3 using MLP with Backpropagation algorithm and pre-determined topology.

3.4 Multiple linear regressions model

This method uses the sum of the least squared errors to fit a curve to a data set. Using Minitab 17, a linear regression analysis was performed on the data set. Three variables were used as the independent variables (X_i), that is, X_1 , X_2 and X_3 be the volume at the day close of the market (*Vol@close*), the Financial Times Stock Exchange Bursa Malaysia Index (FTSEKLCI) and Market Valuation (*Market VAL*), respectively. The independent variable is termed Y is the value at the day close of the market (*Val@close*).

Four assumptions checks were carried out on the data set: (i). Normality (ii.) Homoscedasticity (iii) The Independence of errors and (iv). Linearly The Multiple Linear Regressions formula is as follows:

$$Y = \beta_0 + \beta_i X_i + \varepsilon_i \quad (8)$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \quad (9)$$

4 Results and discussion

From the methods and materials that were experimented with the data collected, the results from the two different models were recorded and analyzed.

4.1 Implementation of the ANN model

The neural network model using Nonlinear Autoregressive with External input (NARX) neural network type with topology 3-3-1 was implemented and the training function Levenberg-Marquardt Backpropagation was used. The results from its first training and fifteenth other retraining are shown in Table 1.

Table 1: Result of the best convergence training of ANN model

| Data set | Target | Mean squared error | Regression |
|------------|--------|--------------------|------------|
| Training | 43 | 1.06194e-2 | 9.2179e-1 |
| Validation | 10 | 1.09351e-2 | 9.3084e-1 |
| Testing | 10 | 1.03598e-2 | 9.2843e-1 |

The data set was divided into training set of 68%, the validation set of 16% and the testing set 16%. The testing set performed best in the mean square error and the validation had the best regression.

From the plot of Figure 1, plot from the fifteenth retrain of ANN model, had its best validation performance as 0.010935 at epoch 5, and the test and validation curves converged to the best line, the training curve was close to the best line. Hence, it was considered to be the best plot and the trained that had the best convergence. Figure 2, had best training graphs showing the training, validation testing and all regression values.

From Table 1, it implies that the train which converge to the best line has its regression above 0.86 gives better prediction equation. Hence, the Mathematical model for ANN is:

$$A \cong 0.86 * T + 0.059 \quad (10)$$

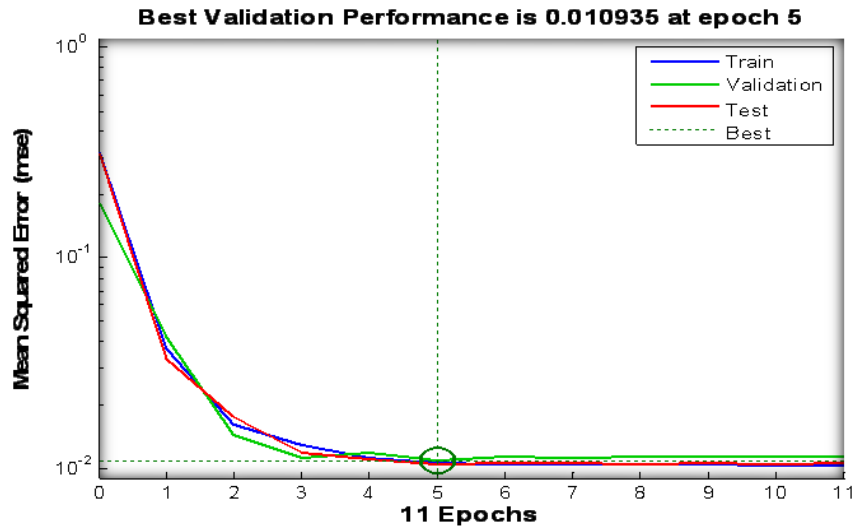


Figure 1: Plot for fifteenth retrain of ANN model

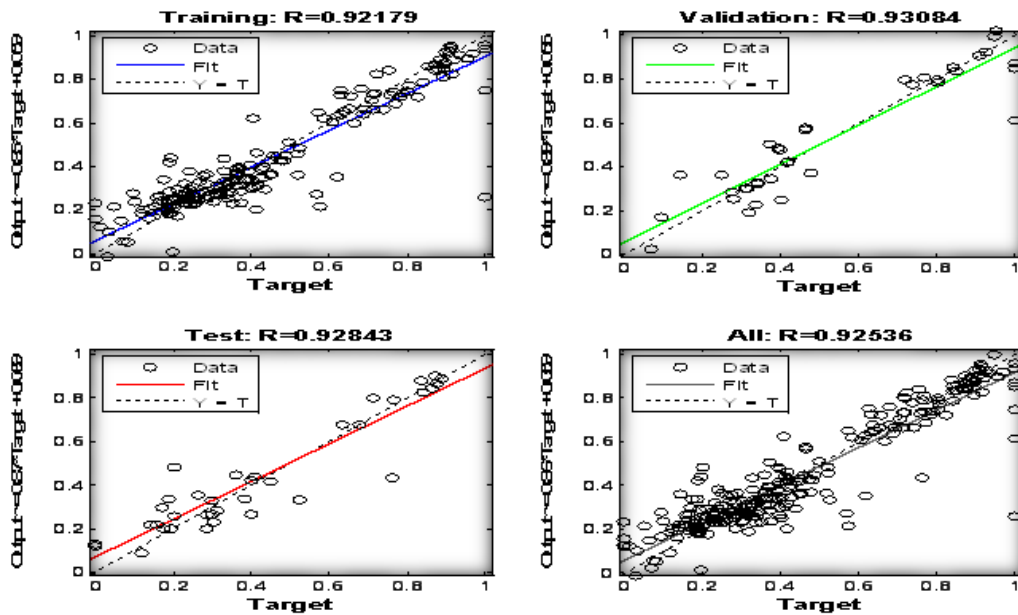


Figure 2: The best training graphs showing the training, validation testing and all regression values

The letter A is the actual stock price to be predicted while T is used as the training data (for today's stock price) which will be used for making prediction.

4.2 Multiple Linear Regressions analysis

Multiple linear regressions analysis was performed on the normalized data using Minitab version 17. The output of this process for the final model is presented in below.

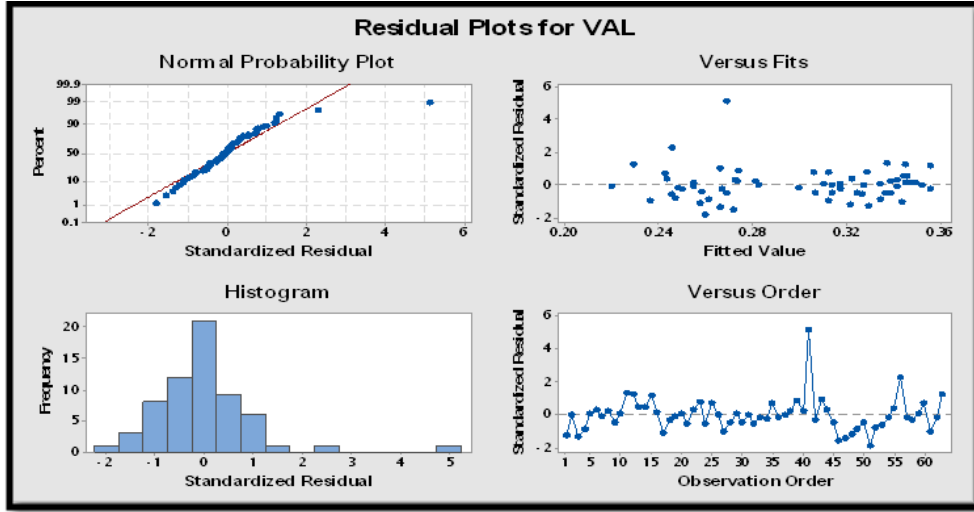


Figure 3: The verification of the regression assumptions in plot forms

Figure 3, shows the verification of the regression assumptions in plot forms. That is, a normal probability plot of the residuals and plots of the standardized residuals against the fitted (predicted) values.

Table 2, shows the overall goodness of fit measures. The correlation coefficient multiple R indicates the strongest or weakness the linear relations is between dependent and independent variables. When the value of R is 1 or closer to 1, the linear relationship is positively perfect, and a value zero means no relationship at all. In Table 2, multiple $R = 0.553904311$ shows that there is good relationship, the standard error, MSE gave an adequate error as 0.126913401 and the number of observation $n = 63$ was used in the regression.

Table 2: Multiple linear regressions statistical outcomes

| Regression parameter | Result |
|----------------------|-------------|
| Multiple R | 0.553904311 |
| R^2 | 0.306809986 |
| Adjusted R^2 | 0.271563036 |
| MSE | 0.126913401 |
| Observations | 63 |

All the assumptions of regression analysis were verified; therefore the linear regression model seemed appropriate. Hence, the Regression model equation is:

$$VAL@CLOSE = 0.1892 + 0.4089VOL@CLOSE + 0.0774FTSEKLCI - 0.0182MarketVAL$$

Table 3, show that the results obtained between the two models. ANN result outperformed that of MLR. ANN produced high value of R^2 with less performance error of 0.0731 than MLR with a very low of R^2 and high performance error of 0.1269. In prediction models, the closer the values of R^2 to 1 with lower performance error, the better prediction results.

Table 3: Ranking results of the MSE and the R^2 of ANN and MLP models

| Models | MSE | R^2 |
|--------|--------|---------|
| ANN | 0.0731 | 0.92536 |
| MLR | 0.1269 | 0.55390 |

5 Conclusion

Artificial intelligence technique under the machine learning phenomenon was applied for the prediction of the future index of the Malaysia Stock market. The prediction model based on ANN was more accurate than MLR. Investors can make use of ANN to aid them in maximizing profit in the stock market.

For future research work, metaheuristic algorithms and other traditional forecasting tools such as R statistical software can be used to develop an effective model on the possibilities of using a combination of the basic market indices and other micro-economic indicial (for example, fiscal monetary policies, Inflation rate, budget figures on critical infrastructure and growth drivers, e.t.c) to predict the future index of the Malaysia Stock Exchange Market.

Acknowledgment

This research is partly financed by FRGS grant (203/ PMATHS/6711368) from the Ministry of Higher Education, Malaysia.

References

- [1] Chen, A. S., Leung, M. T. and Daouk, H. Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. *Computers & Operations Research*. 2003. 30(6): 901-923.
- [2] Choudhry, R. and Garg, K. A hybrid machine learning system for stock market forecasting. *World Academy of Science, Engineering and Technology*. 2008. 39(3): 315-318.

- [3] Smith, C. Stock Market Malaysia History and Operation, world village. Retrieved from: www.ukessays.co.uk/essays/business/the-stock-exchange-of-malaysia.php. 2015, March 13.
- [4] Trading, E. Malaysia Stock Market. Retrieved from: <http://www.tradingeconomics.com/malaysia/stock-market>. 2015, April 6.
- [5] Cortes, C. and Vapnik, V. Support-vector networks. *Machine Learning*. 1995. 20(3): 273-297.
- [6] Yu, L., Chen, H., Wang, S. and Lai, K. K. Evolving least squares support vector machines for stock market trend mining. *IEEE Transactions on evolutionary computation*. 2009. 13(1): 87-102.
- [7] Eberhart, R. C. and Kennedy, J. A new optimizer using particle swarm theory. In *Proceedings of the sixth international symposium on micro machine and human science*. 1995. 1: 39-43.
- [8] Poli, R., Kennedy, J. and Blackwell, T. Particle swarm optimization. *Swarm intelligence*. 2007. 1(1): 33-57.
- [9] Pilbeam, K. *Finance and Financial Markets*. Palgrave Macmillan. 2010.
- [10] Lo, A. W. and MacKinlay, A. C. Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*. 1988. 1(1): 41-66.