International Journal of Engineering & Technology, 7 (3.15) (2018) 36-39

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Predicting Stock Market Index Using Hybrid Intelligence Model

Sarah Nadirah Mohd Johari[1]*, Fairuz Husna Muhamad Farid[2], Nur Afifah Enara Binti Nasrudin[3], Nur Sarah Liyana Bistamam[4], Nur Syamira Syamimi Muhammad Shuhaili[5]

[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia*
[2] *Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia*
[3] *Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia*
[4] *Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia*
[5] *Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia*
**Corresponding author E-mail:sarahnadirah@tmsk.uitm.edu.my*

Abstract

Predicting financial market changes is an important issue in time series analysis, receiving an increasing attention due to financial crisis. Autoregressive integrated moving average (ARIMA) model has been one of the most widely used linear models in time series forecasting but ARIMA model cannot capture nonlinear patterns easily. Generalized autoregressive conditional heteroscedasticity (GARCH) model applied understanding of volatility depending to the estimation of previous forecast error and current volatility, improving ARIMA model. Support vector machine (SVM) and artificial neural network (ANN) have been successfully applied in solving nonlinear regression estimation problems. This study proposes hybrid methodology that exploits unique strength of GARCH + SVM model, and GARCH + ANN model in forecasting stock index. Real data sets of stock prices FTSE Bursa Malaysia KLCI were used to examine the forecasting accuracy of the proposed model. The results shows that the proposed hybrid model achieves best forecasting compared to other model.

Keywords: *stock prediction; GARCH, ANN; SVM; Hybrid model*

1. Introduction

The most depressing phenomenon that gives negative impact on restructuring and reforming policies to every field in market are when financial crisis, stock market crash and economic collapse. When stock market crash occurs, investors are possible to face financial loss because of the declination value of stock and gold in the market globally [1]. It leads to unfavorable situation on economic environment.

Therefore, it is important to analyse the markets past trending order to reduce the probability of a market crash to arise in the future [2]. The FTSE Bursa Malaysia KLCI will be examined to understand and determined the pattern of stock market. Before selecting a model for this paper, the stock market's chaotic behaviour will be studied initially. Chaotic behaviour is an unpredictable nature that will reveal the changes of hidden pattern. From studying the chaotic behaviour, we want to find the best model in forecasting the stock market index. As a result, we found that Hybrid Intelligence Model is widely used [3].

The model is divided into two categories, linear models and nonlinear models. Artificial Neural Network (ANN) and Support Vector Machine (SVM) under linear models and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) under nonlinear model has been proposed for this paper. The comparative model that combined with two different combinations which are GARCH + ANN and GARCH + SVM has been considered. A combination of linear and nonlinear method as hybrid intelligence model can produce a more accurate result of stock market index.

2. Literature Review

The financial crisis of the global bank affected FTSE Bursa Malaysia KLCI, a fall of 670 points as the biggest decline after the East Asian Financial Crisis in 1997. But Malaysia as the developing country had made its wise move by having safety precautions to have reserve and other reliable source of dependency during the financial crisis. Still, having such precautions cannot restraint a crash if the condition of the economical world is not good.

An illustration of a small system projecting patterned and unpredictable outcomes is a Chaos Theory, where it is the study of chaos, dynamic and nonlinear systems [4]. As cited in [5], they have applied this theory in financial field in other to analyse the changes in the bill rate and stock market index. [6] too, have applied chaos theory in determining the changing process of technical environment dynamic cycle. This has been proven that the chaos theory is widely used in various fields of studies.

Artificial Neural Networks (ANNs) is a common technique used in predicting financial time series analysis. [7] used the ANNs model to predict Bombay Stock Exchange (BSE) market returns using ANNs. They have found that ANNs model is more superior to the autoregressive model due to their ability to interpret both linear and nonlinear patterns. According to Refenes et al., S.I. Wu et al., Schoeneburg, E., as cited in [8], ANNs is a better techniques to capture the relationship between stock's performance and the determinant's factors.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

When modelling the Support Vector Machine (SVM), the objective is to choose the most optimal hyperplane in high dimensional space to ensure the upper bound error is minimal. There are two categories for SVM which is support vector classification and support vector regression. SVM is to detect maximum margin hyperplane, separating positive and negative data [9]. The SVM that have been developed provide a good prediction with respect of ANNs and other regression model.

3. Research Methodology

Starting by collecting data, the data will be processed. Once the data has been processed, the pattern of the FTSE Bursa Malaysia KLCI is detected by using the cumulative multi-year on differences in adjusted close price. The chaotic has to be detected too by using the Largest Lyapunov Exponent. A nonlinear process which appears to be random is called chaos. [3] Then the data for year 2018 will be forecasted on two combinations of methods (ARIMA+ANN and ARIMA+SVM) for linear and nonlinear. FTSE Bursa Malaysia KLCI was also known as FBM KLCI which comprises 30 largest companies on Bursa Malaysia that capitalized on weighted stock market index. The 30 companies must meet eligibility requirements of FTSE Bursa Malaysia Index Ground Rules (Bloomberg).The data series consists of closing index of FTSE Bursa Malaysia KLCI from January 1993 to December 2017 for forecasting period where all the data obtained from The Wall Street Journal.

3.1. Detecting Chaos by Using the Largest Lyapunov Exponent

The method will be used to detect the chaotic behaviour in this research. Three initial steps are needed in the largest Lyapunov exponent which are;

Step 1: Any data can be modelled by equation below;

$$\ln \lambda = \lim(n\rightarrow\infty) [1/n \sum \ln (df(x_1)/dx)] \quad (1)$$

The rule is developed by finding the nearest point of x_1 and x_n with embedding dimension m and the average logarithmic rate of separation between the two nearest points with time interval k is computed as follows;

$$L_k = 1/(2(N-k-m+1)) \sum \log \sum (x_{l-j+k} - x_{n-j+k}^2) \quad (2)$$

The largest Lyapunov exponent is defined as [10];

$$\lambda_1 = (dL_k / dK) \quad (3)$$

Small values of L_k are ignored since the point is not arranged along the direction of maximum expansion and there may be noise and error in calculating the separation. The large number of L_k are ignored too as the separation approaching the attractor's size and cause deviations from exponential [10].

Step 2: The gradient of the slope is found using curve fitting tool for linear in MATLAB and λ is obtained.

Step 3: The findings are analysed using the following interpretations.

Table 1: Interpretations of Findings the Largest Lyapunov Exponent

Value of λ	Interpretation
λ > 0	Chaos exists. The system has a greater stability and the orbit attracts to a stable and fixed point.
λ < 0	The data is not chaotic and it is stable.

3.2. Forecasting Stock Market Index

The method that will be used to predict the stock market index is hybrid intelligence method. The model that will be used for linear model is Generalized Autoregressive Conditional Heteroscedasticity (GARCH) and for nonlinear are ANN and SVM. We will combine the GARCH with ANN and SVM respectively.

3.2.1. Generalized Autoregressive Conditional Heteroscedasticity (GARCH)

GARCH models are often used to model volatility. The model applies the understanding of volatility depending on the estimation of previous forecast error and current volatility. The general GARCH (p,q) model has the following form:

$$h_t = \gamma_0 + \sum \delta_i h_{t-i} + \sum \gamma_j u_{t-j}^2 \quad (4)$$

The value of the variance scaling parameter h_t now depends both on past values of the stocks. If p = 0, the model reduces to ARCH (q), the simplest form of the GARCH (p, q) model is the GARCH (1, 1) model which the variance equation is:

$$h_t = \gamma_0 + \delta_1 h_{t-1} + \gamma_1 u_{t-1}^2 \quad (5)$$

Hence, GARCH estimation model has been employed by [11] which introduced for estimating time-varying risk and provides effective tools for estimating the volatilities of stock price.

3.2.2. Artificial Neural Network (ANN)

ANNs are widely used and have been proven to be one of the most successful methods in solving prediction problem [12]. It is a fit instrument to capture the nonlinear behaviour with a high degree of accuracy and competitive computational time [12]. Inside ANNs, the system is called Multilayer Perceptron (MLP) that contains neuron which is organized in layers. Each neuron performs a simple calculation and a scalar function.

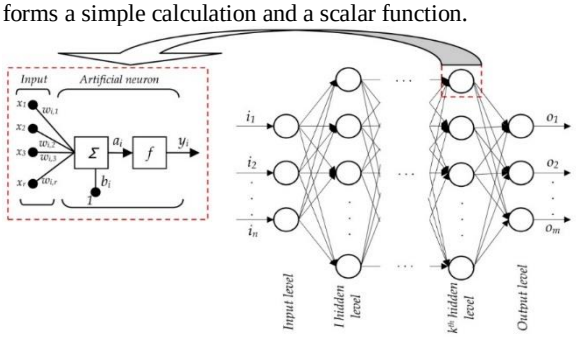


Fig. 1: A simple neuron diagram [13].

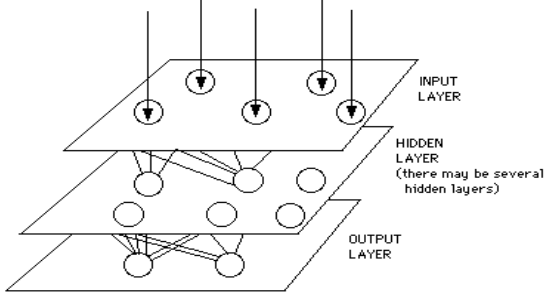


Fig. 2: Neural Network Diagram [7].

Each neuron contains a summation function and an activation function.

$$S = \sum w_i x_i \quad (6)$$

$$\emptyset(S) = 1/(1+e^{-S}) \quad (7)$$

S = output of the neuron.

w_i = weight for each input.
x_i = inputs to the neuron.

The parameters should be confirmed including number of layers in the hidden layer, type of sigmoid function for the neurons and the learning algorithm.

3.2.2. Support Vector Machine (SVM)

SVM create a boundary such that most points in one category fall on one side of boundary and vice versa. Assuming the linear boundary (hyperplane) as [14]:

$$\beta_0 + \beta_1 X_1 + \beta_n X_n = \beta_0 + \sum \beta_i \ X_i = 0 \quad (8)$$

β_n= Variables affecting closing stock market index.

The optimal hyperplane is the maximize distance from the plane to any point. This is called margin. It is needed to add error variables $\varepsilon_1, \ldots, \varepsilon_n$ and keep the sum since the differentiation is not perfect. The specific kernel function used in this study is the radial kernel and its advantage is it can handle diverse input sets of the input data:

$$(x_i, x_k) = \exp -1/\delta \wedge 2 \sum (x_{ij} - x_{kj})^2 \quad (9)$$

δ = bandwidth of the kernel function
x_i = Opening Stock Market
x_k = Closing Stock Market

3.3. Propose of Hybrid Intelligence Model

3.3.1. Generalized Autoregressive Conditional Heteroscedasticity (GARCH) + Artificial Neural Network (ANN).

In order to applies both of linear and nonlinear, [15] has developed a hybrid approach that applies GARCH and ANN,

$$y_t = L_t + N_t \quad (10)$$

y_t = the observation of the stock market.
N_t = nonlinear element from ANN.
L_t = the predicted value of closing stock market from linear model.

When GARCH is fitted to the linear component and the corresponding forecast $(L_t)\hat{}$ at time t is obtained. The residual at time t will be by $e_t = y_t - (L_t)\hat{}$. The residual dataset after fitting GARCH only contains nonlinear component. This can be modelled through ANN. Using p input nodes, the ANN for residuals will be in the form:

$$e_t = f(e_{t-1}, e_{t-2}, .., e_{t-p}) + \varepsilon_t \quad (11)$$

f = nonlinear function, result value of ANN
ε_t = error term

If $N\hat{}_t$ is the forecast of this ANN, the hybrid forecast at time t is obtained as;

$$y\hat{}_t = L\hat{}_t + N\hat{}_t \quad (12)$$

$L\hat{}_t$ = the forecasted closing stock market at time t
$N\hat{}_t$ = the error value of ANN at time t

3.3.2. Generalized Autoregressive Conditional Heteroscedasticity (GARCH) + Support Vector Machine (SVM).

It is reasonable to consider a time series to be composed of a linear autocorrelation structure and a nonlinear component. The formula is shown below:

$$y_t = L_t + N_t \quad (13)$$

L = linear component of closing stock market.
N = nonlinear component closing stock market.

Let e_t denote the residual at time t from the linear model and let $L\hat{}$ denotes the forecast value from GARCH model.

$$e_t = y_t - L\hat{}_t \quad (14)$$

Linear model is not sufficient because there are still linear correlation structures left in the residuals [13]. By modelling residuals using SVM, nonlinear relationships can be discovered.

$$e_t = f(e_{t-1}, e_{t-2}, .., e_{t-p}) + \varepsilon_t \quad (15)$$

f = nonlinear function of closing stock market
ε_t = error term

This brings to;

$$y\hat{}_t = L\hat{}_t + N\hat{}_t \quad (16)$$

$L\hat{}_t$ = the forecasted value of closing stock market
$N\hat{}_t$ = forecast value of SVM

4. Results and Analysis

By using cumulative multi-year on differences in adjusted close price FTSE Bursa Malaysia KLCI, the pattern for stock market to crash can be seen every 10 years. Comparing to the historical data of FTSE Bursa Malaysia KLCI, it happens to be crashed in 1998 and 2008 as the cumulative multi-year on the difference in adjusted price. After several simulations in MATLAB to find the largest Lyapunov exponent, we have found that the λ is equal to 0.004361. This indicates that a positive value of λ equivalent to an existence of chaos and the divergence of the nearby trajectories. To find the best model, we have ran the simulations of all models in R Studio and MATLAB. The mean squared error (MSE) was observed for all models to determine which one is the best model. This can be seen in Table 2.

Table 2: Mean squared error (MSE) for all models

Model	MSE	RMSE	R Squared
GARCH	116.9023	10.8121	0.99
ANN	74.2000	8.6139	0.99
SVM	1205.4000	34.719	0.99
GARCH + ANN	116.9007	10.8120	0.99
GARCH + SVM	59.8757	7.7379	0.99

The mean squared error of the hybrid intelligence model both GARCH + ANN and GARCH + SVM are 116.9007 and 59.8757 respectively. The value error in the hybrid model of GARCH + SVM has outperformed the entire model individually of another hybrid method which is GARCH + ANN.

5. Conclusion

Since the λ is more than 0, this means that the stock market is chaotic. This shows that the system has a greater stability with the orbit attract to a fixed and stable point. Proving that the data are chaotic will is to proceed in forecasting the stock market. Achieving a highly accurate forecast is a challenging task. GARCH, ANN and SVM are widely known in their effectiveness in predicting stock market. It is impossible to really capture the exact nature of a time series because it often contains both linear and nonlinear component. In this paper, we have proposed a hybrid intelligence model combining linear and nonlinear model. SVM and ANN are used for nonlinear model as it were the best single method for stock market forecasting [9]. In accordance to Sheta, SVM and ANN has many advantages over classical classification approaches where SVM has good performance in high dimensional space and ANN has the ability to deal with uncertainty and insufficient

data sets which change rapidly in very short period of time. We have found that the combination of GARCH + SVM has the least mean squared error. Hence, it is the best model in predicting stock market index. Besides using models in this paper, it is best to have further studies using on Value at Risk (VAR) model in determining the performance of a stock as in [16]. It can also be carried on further research regarding gold price in hedging strategies and causality linkages [1][17].

Acknowledgement

This research is funded by Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, Malaysia.

References

[1] Nawawi, A. H., Radzali, N. H., Hussin, S. A. S., & Mohd, M. A. (2016). Gold hedging strategies in the Asian Markets. Jurnal Teknologi, 78(4-4), 89-98.

[2] Barro, R.J., &Ursu´a, J. F. (2017). Stock-market crashes and depressions. Research in Economics, 71(3), 384-398. https://doi.org/10.1016/j.rie.2017.04.001

[3] Miah, M.B.A, Hossainm M.Z, Hossain, M.A, Islam. (2015). M.M. Index prediction of stock market using hybrid model of artificial intelligence. International Journal of Computer Applications 111(3):5-9.

[4] Levy, D. (1994). Chaos theory and strategy: Theory, application and managerial implications. Strategic management journal, 15, 167-167

[5] Wang, S-C., Wang, P-L., & Fang, K. (2011). To investigate the development process of most popular blog through chaos theory. Paper presented at the ICSSSM11

[6] Hu, J., Tung, W.-w., Gao, J., & Cao, Y. (2005). Reliability of the 0-1 test for chaos. Physical Review E, 72(5), 056207

[7] Perwej, Y., Perwej, A.(2012). Prediction of the Bombay stock exchange (BSE) market returns using artificial neural network and genetic algorithm. Journal of Intelligent Learning Systems and Applications, Vol. 4 No. 2, pp. 108-119. doi: 10.4236/jilsa.2012.42010.

[8] Majumder, M., Hussian MD. (2009). Forecasting of Indian stock market index using artificial neural network. Retrieved from www.nseindia.com/content/research/FinalPaper206.pdf

[9] Sheta, A.F, Ahmed, S.E.M, Faris, H. (2015). A comparison between regression, artificial neural networks and support vector machines for predicting stock market index. International Journal of Advanced Research in Artificial Intelligence (IJARAI), 4(7).

[10] Sapini, M. L., Rahim, N. Z. b. A., Noorani, M. S. M., Mohamed, I., How, L. T., Mui, A. C. Y., & Bin, W. K. (2015). The behavior of PM10 and the ozone in Malaysia through nonlinear dynamical systems. Paper presented at the AIP Conference Proceedings

[11] Shahruddin, S. N. S. S. M., Tafri, F. H. B., & Lazam, N. M. (2012, September). Optimal portfolio:"Empirical investigation of food industry listed in KLSE". In Statistics in Science, Business, and Engineering (ICSSBE), 2012 International Conference on (pp. 1-6). IEEE.

[12] Azmir, Zahrul & Tafri, Fauziah Hanim. (2012). Islamic options pricing model via artificial neural network: "Benchmarking to Black-Scholes". 1-6. 10.1109/ICSSBE.2012.6396562.

[13] Tanikic, D., Despotovic, V. (2012). Artificial intelligence techniques for modeling of temperature in the metal cutting process. Metallurgy – Advances in Materials and Processes, Chapter 7

[14] Patil, S.S, Patidar, K., Jain, M. (2016). Stock market prediction using support vector machine. International Journal of Current Trends in Engineering & Technology, Vol. 02 (1).

[15] Zhang, G., P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159-175.

[16] Siti, S. M., Mohd, M. A., Shima, S. N., & Adzhar, Z. A. A. K. (2017). Value-at-Risk on different economic sectors in Malaysia. Journal of Engineering and Applied Sciences, 12(9), 2289-2293.

[17] Mohd, M. A. (2011, September). Causality linkages between USA and Asian Islamic stock markets. In Business, Engineering and Industrial Applications (ISBEIA), 2011 IEEE Symposium on (pp. 123-128). IEEE.