Modelling and Forecasting Four Market Indices: Autoregressive Integrated Moving Average Model versus Artificial Neural Network Model

Kok Thim CHAN, Pei Yee PHAN and Voon Choong YAP
Faculty of Management, Multimedia University, Malaysia
*Corresponding author

Keywords: Forecasting, Index, Modelling, Neural network, ARIMA.

Abstract. This study model and forecast the stock prices of 4 stock market indices: the FTSE KLCI, Dow Jones Industrial Average, NASDAQ Composite and S&P 500. We utilize two distinct approaches which are time series analysis and artificial intelligence system. We model the data with Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Network (ANN) applying the time series data of the 4 selected stock market indices. The time frame was five years, starting on 3 January 2012 to 29 December 2017. After the modelling, the outputs are subsequently compared and contrasted in terms of forecast accuracy, such as MAE, MAPE, RMSE and MSE, and the model with the lowest forecast error is sought to be the best-fitted model. We found ANN model to be superior as it outperformed the time series model by generating the lowest forecast error across four datasets.

Introduction

In this era of globalization, stock market indices have typically garnered a drastic attention across the worldwide as they offer the measurements of the total market return and they act as the benchmarks of portfolio performance in addition. Therefore, stock market indices are emphasized in numerous of research works as the insights of a particular stock market are likely to be revealed and discovered through the study on stock market indices.

According to [1], investment in stock market indices has garnered extraordinary popularity in primary financial markets around the globe. Almost every market participants, particularly traders or investors, are paying an extremely high degree of attention to the daily movements in equity market which is significantly volatile and hence making the traded stocks risky yet greatly rewarding.

Practically, the market players strive for predicting the stock market movements, specifically the shifts of a particular stock price, either rises or drops will greatly impact the strategies or actions which will be taken in order to achieve return maximization or loss minimization. As claimed by [2], investors are granted the opportunities to yield from the equity market, given that they are able to forecast the changes in stock index precisely. Therefore, market players have been engaging in forecasting the stock market and attempting to gain prosperity with its application. Due to the significance of forecasting, a vast amount of forecasting techniques has been developed and introduced over the years.

Literature Reviews

Stock market, which is one of the major components in composing the capital market, is an open market, where the market players freely trade for the securities of public listed firms. [3] have stated that securities have been perceived as a highly rewarding investment over a long time; while according to [4], the prerequisite for a successful financial market trading is firmly believed to be having the ability to precisely forecast the succeeding market movements. Relatively, stock market forecasting has been gaining substantial attention over the years as traders are possibly to reap a huge wealth from stock investments.
According to [5], stock price index is the primary indicator which is adopted in analyzing the performance of the stock market as it can represent some harmonious factors, as well as measure and compare the significant events which contribute to the fluctuations in stock prices. In addition, stock market index also acts as the reflection of the economic circumstances of a nation, whereby the recession in a country is signified when the stock market index declines while the economic growth is implied if there is a rise in the stock market index on the contradictory.

Time series forecasting has long been heavily researched for a few decades [6]. Overall, time series forecasting is regarded as a vital context of forecasting, whereby the historical data of a variable of interest is gathered and analyzed thus to develop a model which tends to explain the underlying relationship and estimate the time series into the future [7]. A variety of time series model has been introduced and developed over the years, such as Naïve model, moving average, exponential smoothing, and autoregressive integrated moving average (ARIMA) which are the most widely applied classical or traditional forecasting model.

According to [8], artificial intelligence (AI) is capable of collaborating with agents in optimizing the communication and enhancement of the performance and thus improve the probability of success. Among the AI models, Artificial Neural Network (ANN) and Genetic Algorithm (GA) are the two widely used models for nonlinear time series prediction [15]. Practically, ANN is an information processing system which has homogeneous attributes with biological neural networks [9] whereas [10] has defined the ANN as a mathematical model which is modeled according to the way the human brain functions in determining the data trends.

According to Zhang [7], ANN has been widely investigated and applied in time series forecasting in the late decades and it has been positively contributing to many fields of business, industry and science. In conjunction with that, [11] have also mentioned about the success of the application of ANN in a variety of industry, particularly for forecasting. Generally, ANN possesses the capability in analyzing complex data patterns efficiently and relatively accurately and it simultaneously offers a certain degree of flexibility in its application [12]. Therefore, ANN is preferred in stock market forecasting due to its ability in diagnosing the complex relationship between variables.

Besides, ANN has basically provided numerous advantages over the time series models, specifically its fundamental ability of non-linear modeling which is extensively practical yet offering the ability of linear modeling simultaneously [13]. Besides, the major strength of ANN also includes its capability of dealing parallel with input variable and handling with extensively huge data sets, according to [7] and [13]. Furthermore, ANN is a data-driven and self-adaptive model as it is structured according to the characteristics portrayed in the selected data [11]. Relatively, this data-driven attribute is perceived to be competent when handling data sets that are lack of theoretical guidance for recommending an adequate data generating process.

Nonetheless, the robustness of ANN model in pattern classification and prediction performance have been supported by a vast amount of researches, for instance, [3], [14], [15] and [16].

**Research Design**

In general, the methodology or procedures pertaining to the use of ARIMA model and ANN model are complicate and complex, but we can only briefly describe in this section. In the nutshell, ARIMA and ANN models are similar as they both have three distinct stages in modeling whereby ARIMA model needs to go through model identification, model estimation and model evaluation whereas ANN model must undergo training, testing and validating of data. The best model under each of the approach is determined primarily through AIC and SIC and MSE respectively. The settings for ANN model has been defined and the tools used in this study include Microsoft Excel, Eviews and MATLAB Neural Network Toolbox.

The study period of this research has spanned five years from 2012 to 2017, precisely the time series data of each stock market index was gathered. A total sample size of 1509 for American stock market indices, 1474 observations for BM KLCI. The data used in this research are daily historical
price of four different stock market indices and it is retrieved from reputable and reliable sources such as Bloomberg. Therefore, in an ARIMA model, the impacts of both components have been captured collectively and generally expressed as:

\[ Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon_t - \gamma_1 \varepsilon_{t-1} - \gamma_2 \varepsilon_{t-2} - \cdots - \gamma_q \varepsilon_{t-q} \]  

where,

- \( Y_t \) represents the forecast value,
- \( \varepsilon_t \) represents the random error at \( t \),
- \( \beta_i \) and \( \gamma_i \) represent the coefficients identified by linear regressions,
- \( p \) and \( q \) represent the order of AR and MA component respectively.

Typically, an ANN model is comprised of one input layer, numerous hidden layers and one output layer and the single hidden layer feedforward network is often adopted in time series forecasting (Zhang, 2003). Each layer has several units which output is a function of the weighted sum of the inputs and it is computed using the equation as below:

\[ Y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \cdot \delta(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} \cdot Y_{t-1}) + \varepsilon_t \]  

where,

- \( p \) refers to the number of input neurons,
- \( q \) refers to the number of hidden neurons,
- \( \alpha_j \) refers to the connection weights where \( j = 1,2,\ldots,q \),
- \( \beta_{ij} \) refers to the connection weights where \( i = 0,1,2,\ldots,p \) and \( j = 1,2,\ldots,q \).

Principally, the feed-forward neural network describes the circumstances where the information or data only moves in one single direction, precisely moves forward from the input layer to the output layer via hidden layers. Before the data series can be trained using ANN, the data, either input or output, shall necessarily undergo a process called normalization, whereby the data is transformed and scaled down to be within a specified range of -1 to 1. After the process of normalization has been settled, the training process of neural network begins. Essentially, training process involves the optimization of network’s performance through adjusting the connection weights repetitively, as explained in the back-propagation algorithm.

**Findings**

The empirical results are consistent across four sets of data in spite of the geographic differential which may exert certain degree of impacts to the fluctuations in data, hence making them to have different data patterns or characteristics.

| Table 1. Descriptive statistics of each stock market index. |
|-----------------|----------------|----------------|----------------|----------------|
| S&P             | NASDAQ         | KLCI           | DOWJONES       |
| Mean            | 1926.933       | 4509.334       | 1719.808       | 17005.01       |
| Median          | 1986.51        | 4620.723       | 1718.16        | 17009.69       |
| Maximum         | 2690.16        | 6994.759       | 1892.65        | 24837.51       |
| Minimum         | 1277.06        | 2648.36        | 1504.22        | 12101.46       |
| Std. Dev.       | 351.3563       | 1091.067       | 89.3294        | 2822.743       |
| Skewness        | -0.091868      | 0.175412       | -0.02481       | 0.433203       |
| Kurtosis        | 2.219434       | 2.255346       | 2.113809       | 2.867514       |

From the Table 1, it shown the general descriptive statistics of each stock market index, which is comprised of mean, median, maximum value, minimum value, and standard deviation. In general, FTSE Bursa Malaysia KLCI portrayed a daily average value of 1719.808 over the years while its
standard deviation of 89.32940 has explained about the dispersion of the data to the mean. The Dow Jones Industrial Average has an average daily price of 17005.01 throughout the years with a relatively high standard deviation which means that the data is relatively dispersed from the mean. On average, the daily closing price of NASDAQ Composite accounted for 4509.334 with its standard deviation of 1091.067. Overall, S&P 500 has an average daily closing price of 1926.933 and a standard deviation relatively low at 351.

Basically, the best model under each approach is initially selected based on the four forecast measurements, which are MAE, MSE, RMSE and MAPE and each forecasting method are subsequently compared to each other based on the same measurements while the model with the lowest errors is determined as the best forecasting model.

Table 2. Forecast Errors for FTSE Bursa Malaysia KLCI.

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE</th>
<th>MSE</th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA(2,1,1)</td>
<td>6.5293</td>
<td>79.9230</td>
<td>8.9400</td>
<td>0.3819</td>
</tr>
<tr>
<td>ANN(4-9-1)</td>
<td>2.6334</td>
<td>13.4519</td>
<td>3.6677</td>
<td>0.1540</td>
</tr>
</tbody>
</table>

Table 3. Forecast Errors for Dow Jones Industrial Averages.

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE</th>
<th>MSE</th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA(2,1,2)</td>
<td>86.8433</td>
<td>14307.7608</td>
<td>119.6151</td>
<td>0.5243</td>
</tr>
<tr>
<td>ANN(4-6-1)</td>
<td>33.1636</td>
<td>2345.5510</td>
<td>48.4309</td>
<td>0.2009</td>
</tr>
</tbody>
</table>

Table 4. Forecast Errors for NASDAQ Composite.

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE</th>
<th>MSE</th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA(2,1,1)</td>
<td>28.4616</td>
<td>1567.7201</td>
<td>39.5944</td>
<td>0.6497</td>
</tr>
<tr>
<td>ANN(4-4-1)</td>
<td>10.0558</td>
<td>200.9861</td>
<td>14.1770</td>
<td>0.2289</td>
</tr>
</tbody>
</table>

Table 5. Forecast Errors for S&P 500.

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE</th>
<th>MSE</th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA(1,1,1)</td>
<td>10.1018</td>
<td>200.8337</td>
<td>14.1716</td>
<td>0.5402</td>
</tr>
<tr>
<td>ANN(4-9-1)</td>
<td>3.8029</td>
<td>27.3680</td>
<td>5.2314</td>
<td>0.2052</td>
</tr>
</tbody>
</table>

ANN model appears to be the best forecasting model across the four different set of times series data simultaneously as ANN model exhibits the lowest forecast error in MAE, MSE, RMSE and MAPE. The forecast accuracy depicted in ANN model is drastically superior to the other forecasting models, at least 2 to 3 times better than the worst forecasting model. This result is in compliance with the expected result of the research, as well as the other past researches worldwide.

A research carried out by [9] which is relevant to forecasting the stock prices of Indonesian market (PT Aneka Tambang) using ARIMA model and ANN model, has shown to support our findings, where the forecasted values produced by ANN model through the application of MATLAB generated a smaller forecast error than ARIMA model.

Summary

Apparently, ANN model is deemed to be the best forecasting model as compare to the time series models like ARIMA. ANN model has the high precision in forecasting future values of every stock market index. The extremely strong predictive power of ANN model has been revealed in the forecasting measurements used in this research, such as MAE, MSE, RMSE and MAPE. The substantially low error yielded by ANN model in each of the forecast measurements has supported the claim that ANN model is the best model in providing relatively accurate and precise predictions when dealing with the non-linear problem. ARIMA model and ANN model require tedious and proper design of its architecture, specifically the order of AR and MA terms for ARIMA model while the number of hidden neurons for ANN model are essentially critical to the forecast performance of ARIMA and ANN model. If the models are well-designed with no errors, prediction will be absolute accurate, businesses will be profitable.
Acknowledgement

This research was financially supported by the Malaysian Fundamental Research Grant Scheme by the Ministry of Education.

References


