

PETRONAS's Stock Price Forecasting: An Application of Neural Network Model

Nur Aina Syafiq Zulkefli¹, Maria Elena Nor^{1,*}

¹ Department of Mathematics and Statistics,
Faculty of Applied Sciences and Technology,
Universiti Tun Hussein Onn Malaysia, Pagoh Edu Hub, 84600, Johor, Malaysia.

*Corresponding Author Designation

DOI: <https://doi.org/10.30880/ekst.2021.01.02.003>

Received 07 May 2021; Accepted 03 July 2021; Available online 29 July 2021

Abstract: Many researches have been carried out to investigate the influences of oil prices on stock market by using oil and gas as the indicator. The study of the stock market is necessary since it gives a better insight into the future economic growth in the country. However, stock price forecasting somewhat can be quite challenging in some circumstances due to the existing nature of the relationship of the non-linearity and unpredictability in financial market. Since neural network have been proven reliable in many studies to overcome the non-linearity relationships, this study focuses on PETRONAS's stock prices and decided to come out with ANN model to forecast the stock prices against ARIMA (3,1,2) as the benchmark model. Besides, the common error magnitude in most forecast evaluations has been ignoring the importance to evaluate the directional accuracy as well. Therefore, this study had also considered the directional movement to measure the forecast accuracy and the result indicated that both models successfully predicted the directional movements with their MAPE values were below than 5% margin of error.

Keywords: Neural network, ARIMA, error magnitude, directional accuracy

1. Introduction

Stock research is crucial because the time consumed to study over financial history of companies is often helpful to give investors a clear view of their portfolios in future. As a leading economic indicator, the stock market and the causal relationships were evaluated and explored by researchers to study if the price fluctuations in economic activity were directly occurred by the changes in stock prices [1]. The oil and gas industry is doubtlessly to be the important contributors to drive the global economy. A large share in economy is mainly provided from oil and gas production and consumption energy [2]. According to [3], competitors' investment and production activities influenced the future market conditions that required decisions making as well as the prices of oil and natural gas were affected by the production decisions of other firms. . There was a research examined if oil and gas risk components contributed in the returns of nation's oil and gas stocks that engaged in the strength of pricing model.

The results suggested that there was a negative return produced in Malaysian oil and gas industry and investors received excessive profits when investing in growth based in oil and gas stocks [4].

Based on many researchers, traditional time series analysis can be challenging to predict stock data that is due to the data intensity, noise, non-stationary, unfixed nature, high level of undetermined and unseen interconnection behind the process. In the study of stock market, [5] demonstrated a NN model and learning techniques hence study the nonlinear regularities of asset price movements and predicted IBM daily stock prices. However, the result was not able to find an appropriate evidence against the simple market hypothesis when using simple network which only a three layers feedforward network with the same five inputs and five hidden units was being used. In addition, some authors suggested that ANN can satisfactorily adjust both seasonality and the linear trend of a time series, because that ANN was capable of modelling any inconsistent function [6]. Besides, back-propagation neural network models to predict the daily Shanghai Stock Exchange Composite Index were being applied [7]. The models were able to construct the learning algorithm but however inefficiently predicted the return rate. The hidden layer and the output layer function options produced a significant effect towards the accuracy of back-propagation neural network due to the network structure was the same as the weights and the thresholds [8]. Furthermore, ANN was widely used to model a univariate time series using a certain number of lagged terms as the input and the forecasts as the output [9].

Many forecasting methods only considered the magnitude of the error to evaluate the models and did not acknowledged the importance of directional accuracy as well [10]. Previous researcher [11], pointed out that directional accuracy had a major meaning behind compared to the magnitude of accuracy. According to [12], direction of change error criterion had produced consequently the most accurate result by using econometric models over a one-year time horizon. Although there are already lots of existing models in predicting stock prices, there are always rooms for model improvements that can increase the accuracy and reduce the error. The result of this study can support and strengthen the evidence that neural network model approach is suitable to be used and more efficient in predicting stock prices compared to another forecasting approach.

1.1 Problem statement

Based on many researchers, traditional time series analysis can be challenging to predict stock data that is due to the data intensity, noise, non-stationary, unfixed nature, high level of undetermined and unseen interconnection behind the process. The technical analysis in forecasting supposed that all the past stock prices is an essential information to predict subsequent stock prices, which means one should only analyse historical data in order to predict them [13].

Although there are already lots of existing models in predicting stock prices, there are always rooms for model improvements that can increase the accuracy and reduce the error. This study focused on the technical analysis by using the daily data of PETRONAS's stock price to build the forecasting model. The non-linearity and the non-stationary of the data can be overcome by using both ANN and Box-Jenkins approach.

Besides, evaluation by using error magnitude have been ignoring the directional accuracy in forecasting. Therefore, another alternative used in this study to evaluate the forecast performance was by considering the directional movements between observed and predicted value. The application of neural network model can boost the forecasting performance evaluation of error magnitude measurement by taking into consideration about the directional movements.

2. Methodology

In this study, the daily data of PETRONAS's Gas stock price were extracted from a global financial portal and internet brand, which is <https://www.investing.com/equities/petronas-gas-bhd-historical-data>. The data started from March 2014 until September 2020 were used in this study. The dataset was partitioned into a training and testing set to create the model for each method used in this study. The

ratio used for both sets was 80:20. The data for the testing set were being predicted and the values for the outcome of the forecasts were used to compare with the actual testing data.

2.1 Forecast accuracy measurement

2.1.1 Error magnitude

The model which had the least error or had the better accuracy were the best fit model for this stock price dataset. The formulas for all three accuracy measurements of error magnitude are as in the following equations:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_t - \hat{y}_t| \tag{Eq. 1}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_t - \hat{y}_t)^2} \tag{Eq. 2}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| (100) \tag{Eq. 3}$$

where

y_t	=	the actual values at time t
\hat{y}_t	=	the predicted values at time t
$ y_t - \hat{y}_t $	=	the absolute value of the forecast error for period t
n	=	the number period of evaluation

2.1.2 Directional accuracy

A direction of change error occurs when the forecast overlook the true direction of change. For example, if the forecast change was positive and the actual change was negative, or if the forecast change was negative and the actual change was positive. The direction of change error was mentioning to the upward and the downward movement in the data which the downward can be similar as the fall.

Let the out-sample data are at time $T^* = t + 1, t + 2, \dots, t + n$. Then, the directional of actual and forecast data between time T^* and $T^* + 1$ is given by A_{T^*} and F_{T^*} individually, which also can be written as follow:

$$A_{T^*} = y_{T^*+1} - y_{T^*} \tag{Eq. 4}$$

$$F_{T^*} = \hat{y}_{T^*+1} - y_{T^*} \tag{Eq. 5}$$

where the signs of A_{T^*} and F_{T^*} indicate the downward or no downward movement.

i) Chi – Square Test

Chi–Square test of independency using 2x2 contingency table had been shown by previous researchers and it can be used to analyze forecasting performance by assuming actual and predicted values as dichotomous random variables [14].

Table 1: Contingency Table of Directional Between Actual and Predicted Values

F_{T^*}	A_{T^*}		Total
	> 0	≤ 0	
≥ 0	$n_{0,0}$	$n_{0,1}$	$n_{0\cdot}$
< 0	$n_{1,0}$	$n_{1,1}$	$n_{1\cdot}$
Total	$n_{\cdot 0}$	$n_{\cdot 1}$	N

i) Fisher’s exact test

Fisher’s exact test can be computed calculated by finding the p -value. However, it did not calculate a test statistic like the Chi-square test. The p -value obtained by using the hypergeometric distribution to calculate the probability of independency between the observed and the forecast in terms of directional movements. From the contingency table in Table 1, the Fisher’s p -value was calculated by using the equation 6.

$$p\text{-value} = \frac{\binom{n_{0\cdot}}{n_{0,0}} \binom{n_{1\cdot}}{n_{1,0}}}{\binom{N}{n_{0\cdot}}} \tag{Eq. 6}$$

3. Results and Discussion

3.1 ARIMA model

The non-seasonal ARIMA had three tentative models with different AIC values. The best fitted model which was ARIMA (3,1,2) had AIC value of 86.778 and it was the lowest compared to ARIMA (0,1,3) and ARIMA (3,1,0). Therefore, only ARIMA (3,1,2) was chosen as the benchmark model to compare its performance with the ANN model.

Table 2: Comparison of AIC Value

Model	AIC value
ARIMA (0,1,3)	92.065
ARIMA (3,1,0)	90.703
ARIMA (3,1,2)	86.778

Diagnostic checking was required for the model adequacy by using the chosen benchmark model. In order for the chosen model to be adequate, the p -value for AR and MA must be smaller than 0.05 but the p -value in Ljung-Box must be greater than 0.05. Thus, the coefficient of error and the p -value in Table 3 indicate that the model was confirmed to be adequate.

Table 3: The coefficients of error and p -value for model chosen

Model parameter	Parameter significant	Ljung-Box Chi-Square statistics				
		Lag	12	24	36	48
ARIMA (3,1,2)	0.000	p -value	0.694	0.753	0.838	0.764
	0.003					
	0.000					
	0.000					
	0.025					

It was important to initially identify the best fitted model for ARIMA to assist in the selection of lagged observations as the relevant number of inputs for ANN later. The underlying processes that generated the ARIMA model were given in equation 7.

$$\begin{aligned} \phi_p(B) \nabla^d y_t &= \theta_q(B) a_t && \text{Eq. 7} \\ \phi_3(B) \nabla^1 y_t &= \theta_2(B) a_t \\ (1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B) y_t &= \theta_2(B) a_t \\ (1 - B - \phi_1 B + \phi_1 B^2 - \phi_2 B^2 + \phi_2 B^3 - \phi_3 B^3 + \phi_3 B^4) y_t &= \theta_2(B) a_t \\ y_t - B y_t - \phi_1 B y_t + \phi_1 B^2 y_t - \phi_2 B^2 y_t + \phi_2 B^3 y_t - \phi_3 B^3 y_t + \phi_3 B^4 y_t &= \theta_2(B) a_t \\ y_t - y_{t-1} - \phi_1 y_{t-1} + \phi_1 y_{t-2} - \phi_2 y_{t-2} + \phi_2 y_{t-3} - \phi_3 y_{t-3} + \phi_3 y_{t-4} &= \theta(B) a_t \\ y_t &= (1 + \phi_1) y_{t-1} + (\phi_1 - \phi_2) y_{t-2} + (\phi_2 - \phi_3) y_{t-3} + \phi_3 y_{t-4} + \theta_2(B) a_{t-1} \end{aligned}$$

From the derivation, ARIMA (3,1,2) produced lags y_{t-1} , y_{t-2} , y_{t-3} and y_{t-4} . These lags were used as the input layers for ANN model to forecast the stock prices by using Python programming language.

3.2 Neural network architecture

Since the lagged values were already identified by using ARIMA (3,1,2), the lags were used as the input layers in the ANN. There were a total of four lags observations which were also equivalent to four input layers used in the ANN model. The architecture of the ANN was built based on the performance after running the model.

Table 4: Selection for the best number of neurons in ANN based on R^2 value

Number of neurons	R^2 value
10	0.8803
20	0.8754
30	0.8604
40	0.8833
50	0.8824

The number of neurons in the hidden layer was varied to get the best number of neurons. The optimal number of hidden neurons based on the R^2 value since R^2 value measures statistically the closeness of fitness of the data towards the regression line [15]. The hidden neurons with R^2 value nearest to one was selected as the ANN model used in this study. Table 4 presents the selection stage for choosing the number of neurons and 40 hidden neurons were selected as the best number of neurons since the R^2 value achieved was nearest to 1 which was at 0.8833.

3.3 Forecast accuracy measurement

Table 5: Comparison for each forecast accuracy measurement for ARIMA (3,1,2) and ANN

Forecast accuracy measurement	Model	
	ARIMA	Neural Network
MSE	0.075	0.072
MAE	0.170	0.169
RMSE	0.273	0.268
MAPE	1.031	1.007
Fisher's	< 0.00001	< 0.00001

Table 5 indicates that ANN was able to outperform the benchmark model which was ARIMA in terms of error magnitude. In addition, by using directional of accuracy, both models have successfully predicted the directional of the data very well since the significant p -values indicated that the null hypothesis was rejected for its independency between observed and predicted values. By other means, it can be concluded that the forecasts have meaningful value in terms of its direction. The results convinced that it is necessary to consider both error magnitudes and directional movements in forecasting the economics value.

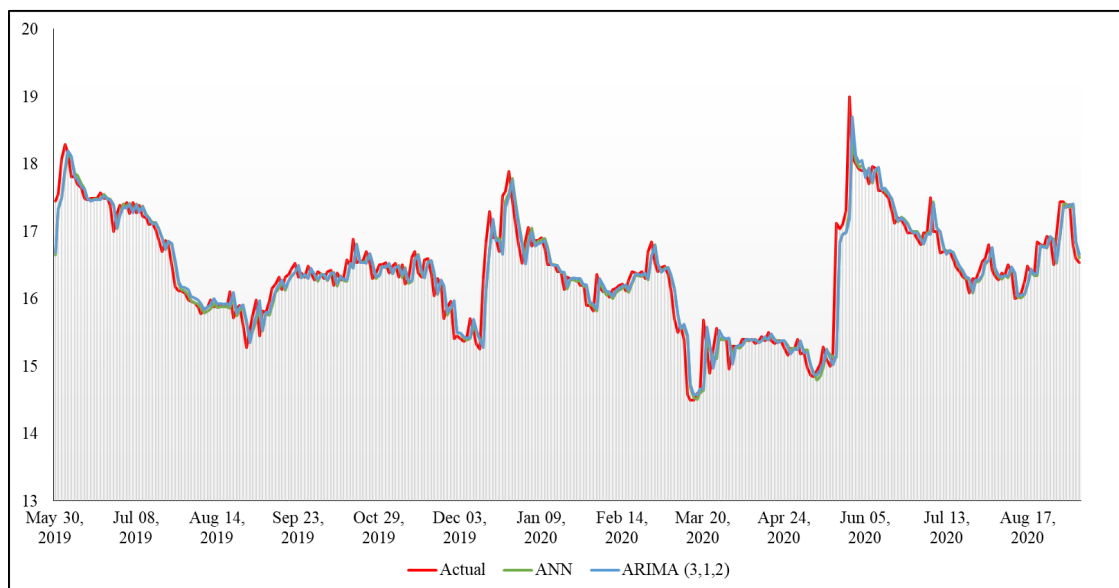


Figure 1: The time series plot of movements for actual and forecast values

Figure 1 shows that the forecast values were extremely fitted to the actual value of the testing set. Therefore, this study assumed that the programming language was not able to produce the real forecast by only using the training set, but instead, it can only generate the forecast output after the values were re-fitted to the testing set too since the plot was nearly perfect.

4. Conclusion

This study had successfully analyzed both Box-Jenkins and ANN model in predicting PETRONAS's stock prices by using Minitab software and Python programming language. The forecast performance between the two models also has been evaluated by using Microsoft Excel and have been compared by using the error magnitude and directional accuracy. The result of error magnitude achieved in this study showed that the ANN model was able to outperform which the values for MAE, RMSE and MAPE were consistently lower than ARIMA (3,1,2) since it is acknowledged that the performance

of ANN has been recognized due to its ability in learning the non-linear regularities of asset price movements in forecasting the stock prices.

In addition, error magnitude was already very commonly used in many researches to evaluate the forecast performance but the information was not enough for decision making especially when it comes to the economics field. Therefore, this study had also considered the directional movements to find if the forecasts were able to predict the directional change excellently. Thus, the overall result showed that both ARIMA (3,1,2) and ANN models had sufficient information to reject the null hypothesis of the independency between observed and predicted value. The significant p -values indicated that the forecasts from both models managed to predict the directional successfully.

Since this study assumed that the Python programming language was not able to produce the output for real forecasts and the result is expected to differ when analyse the same model in different machine language, it is recommended to train the same model by using other software that able to do the forecasting without using the testing set at all. Other limitations in this study were that only the historical prices that were used for the predictions problem which it was recommended to consider the macroeconomic factors as well as the input variables in the neural network.

Besides, the current pandemic of COVID-19 recently had influenced the performance of PETRONAS's company which was affected by the volatility of oil prices and continuously worsen by the unpredictability brought by the ongoing pandemic [16]. The statement also convinced that the historical data was not sufficient to forecast the stock prices due to the other uncertainty factors that should be considered as well. In addition, future study should be concerned to use another advanced machine learning algorithm in forecasting to find more vigorous systems in the stock markets. More technical indicators such as the stochastics or the money flow index can also be used to increase the training performance and to achieve higher accuracy for the forecasting model.

5. Acknowledgment

This research was supported by Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS/1/2019/STG06/UTHM/02/7).

References

- [1] Har, W. M., Ee, C. S., & Tan, C. T. (2008), Stock market and economic growth in Malaysia: Casuality test. *Asian Social Science*, 4(4), 86-92.
- [2] Finley (2012), "The Oil Market to 2030–Implications for Investment and Policy," *Economics of Energy & Environmental Policy*, 1, 25–36.
- [3] Kheiravar, K. H. & Lawell, C. Y. L. (2020), The effects of fuel subsidies on air quality: Evidence from the Iranian subsidy reform. *Working paper, Cornell University*.
- [4] Hoque, M. E., Wah, L. S. & Zaidi, M. A. S. (2020), Do Oil and Gas Risk Factors Matter in the Malaysian Oil and Gas Industry? A Fama-MacBeth Two Stage Panel Regression Approach. *Energies*, 13(5), pp. 1154.
- [5] White, H. (1988), Economic prediction using neural networks: The case of IBM daily stock returns. In *ICNN 2*(2), 451-458.
- [6] Franses, P. H., & Draisma, G. (1997), Recognizing changing seasonal patterns using artificial neural networks. *Journal of Econometrics*, 81(1), 273-280.

- [7] Bing, Y., Hao, J. K., & Zhang, S. C. (2012), Stock market prediction using artificial neural networks. In *Advanced Engineering Forum*, 6, 1055-1060.
- [8] Long, J., Chen, Z., He, W., Wu, T. & Ren, J. (2020), An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market. *Applied Soft Computing*, 91, 106205.
- [9] Bishop (1995), Neural networks for pattern recognition. *Oxford: Oxford University Press*.
- [10] Nor, M. E. (2014), Error magnitude and directional accuracy for time series forecasting evaluation (*Doctoral dissertation, Universiti Teknologi Malaysia*).
- [11] Cicarelli (1982), A new method of evaluating the accuracy of economic forecasts, *Journal of Macroeconomics*, 4(4), 469-475.
- [12] Martin, C. A., & Witt, S. F. (1989), Accuracy of econometric forecasts of tourism. *Annals of Tourism Research*, 16(3), 407-428.
- [13] Lawrence, R. (1997), Using neural networks to forecast stock market prices. *University of Manitoba*, 333, 2006-2013.
- [14] Cumby, R. E., & Modest, D. M. (1987), Testing for market timing ability: A framework for forecast evaluation. *Journal of Financial Economics*, 19(1), 169-189.
- [15] CFI (2020, June 17), What is R-Squared? *Corporate Finance Institute* <https://corporatefinanceinstitute.com/resources/knowledge/other/r-squared/>.
- [16] Allison Lai (2020, September 4), Petronas expects performance to be severely affected this year, *The Star Online*, <https://www.thestar.com.my/business/businessnews/2020/09/04/petronas-expectperformance-to-be-severely-affected-thisyear>.