Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia

Vol. 3, 2021, page 172-180

# Application of Hybrid Forecasting Model in Forecasting Airline Stock Price in Singapore

**Tay Ee Thung [a], Siti Rohani Mohd Nor [b]***

[a,b]Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia.

*Corresponding author: sitirohani@utm.my

**Abstract**

In this research, a hybrid forecasting ARIMA-ANN model is proposed to capture the heteroscedasticity that exists in nonlinear dataset. This research is carried out to forecast airline stock price in Singapore using hybrid ARIMA-ANN model and to compare the forecasting performance of ARIMA model and hybrid ARIMA-ANN model for airline stock price in Singapore. 2508 observations of Singapore Airline Limited stock prices which is from 1st January 2010 to 31st December 2019 have been used in building forecasting model. The result shown that the hybrid ARIMA-ANN model outperformed the ARIMA model in forecasting the airline stock prices in Singapore.

**Keywords:** Airline Industry; Stock Price; ARIMA; Artificial neural networks; Time series forecasting; Hybrid forecasting model

## 1 Introduction

Airline industry is an important sector which offer air transport services for paying consumers and cargo along regularly scheduled routes via air, usually by jets but occasionally by helicopters. The airline industry may be viewed as a subset of the broader aviation business. It is a vital part of the economy due to strong inter-industry linkages with both upstream and downstream sectors. In this demanding industry environment, Singapore Airlines has achieved remarkable results and maintained its competitive edge by successfully adopting a dual strategy of differentiation through service excellence and innovation, as well as cost leadership in its peer group. Since stock prices are affected by the supply and demand in the market at the most fundamental level, therefore the incredible performance of Singapore Airlines contribute to the consistency of Singapore Airline Limited (SGX:C6L) data while the price volatility and instability is caused by dramatic disaster such the September 11 attacks, SARS and Covid-19 happened in the past 10 years. In the past 10 years, the airline industry undergoes impressive change constantly. According to He, Liu, Wang, and Yu (2020), those disasters have huge effect on the economies of the countries impacted. They have negative but short-term influence on the stock markets of the impacted countries and has a bidirectional spillover effect between Asian countries, European and American countries.

The price fluctuation of a stock reflects the investors' valuation of a company and the expected growth of a company in future by investors. According to Bustos and Pomares-Quimbaya (2020), a precise stock market forecasts can offer opportunities for investors to make smarter data-based choices. Inaccurate analysis and forecasting outcome can be a waste of time and money for investors while a good forecast of stock's future price which lead to informed investment decisions will yield significant profit to investors. In these situations of potential influencing factor and high risks that are dependent on changes in needs, an outstanding forecasting model should be developed to help in better decision-making.

Stock markets are nonlinear with unpredictable behavior that varies with time. Stock market forecasting is a very demanding and tough assignment for investors, professional analysts, and researchers in the financial sector due to volatile, complicated, non-linear, and dynamic nature of stock price time series. Conventional approaches such as Auto Regressive Integrated Moving Average (ARIMA) are normally used to generate most stock market analysis in history. However, these models are only ideal for linear data. Therefore, artificial intelligence models such as neural networks, genetic algorithms, and statistically dependent support vector machines have been introduced for the forecasting in stock market. Artificial neural networks (ANN) are ideal for predicting non- stationary data but lack of capabilities in dealing with linear pattern of data. . Thus, the approaches proposed in this research is the hybrid model to fit and to forecast the daily stock prices of Singapore Airline Limited (SGX:C6L) data.

In this paper, the remaining parts will be organized as follows. Dataset and the research methodology will be reviewed in Section 2 while analysis and forecasting results of ARIMA and hybrid ARIMA-ANN models will be introduced in Section 3. Lastly, Section 4 provides the summary and conclusion of this paper.

## 2 Methodology

### *2.1 The Dataset*

The data used is the stock prices of Singapore Airline Limited (Singapore Exchange: C6L) from 4th January 2010 to 31st December 2019. All 2605 observations are obtained from the Yahoo Finance website. Figure 1 show the overall time series plot of daily closing price of Singapore Airline Limited (Singapore Exchange: C6L).


Figure 1 Time series plot of daily closing price of Singapore Airline Limited

### *2.2 ARIMA Model*

Autoregressive Integrated Moving Average with orders $p, q, d$ is used as the mean process. The ARIMA $(p, q, d)$ can refer as in Equation (1):

$$(1 - \sum_{k=1}^{p} \alpha_k B^k)(1 - B)^d X_t = (1 + \sum_{k=1}^{q} \beta_k B^k)\varepsilon_t \quad (1)$$

where $X_t$ is the time series, $\propto$ and $\beta$ are the parameters of autoregressive and moving average terms with order p and q respectively, $\varepsilon_t$ are error terms generally assumed to be independent distributed variables sampled from a normal distribution. $B$ is the difference operator defined as follows by Equation (2), where d is the order of the difference operator.

$$\Delta X_t = X_t - X_{t-1} = (1 - B)X_t \quad (2)$$

Before modelling the dataset with ARIMA$(p, q, d)$, the time series is determined by Augmented Dickey Fuller (ADF) test to verify the stationary of time series. If it is non-stationary, then appropriate order of degree of differencing $d$ is applied on time series as transformation. The autocorrelation function (ACF) and partial autocorrelation function (PACF) of the time series are constructed to determine appropriate values of autoregressive order *p* and moving average *q*.

### 2.3 Hybrid ARIMA-ANN Model

There is a two-phase method in ARIMA-ANN hybrid model system. The best model of the ARIMA models will be used in the first step to predict the linear time series data. By modelling the residuals of ARIMA, ANN is used to discover nonlinearity in the purpose of modelling and forecasting in the second phase. Time series which include linear component and non-linear component is shown as Equation (3)

$$y_t = l_t + n_t \quad (3)$$

where $l_t$ is the linear part and $n_t$ is the nonlinear part. These two parts can be estimated separately from the data. First, the linear part is determined that is separated from the time series. This residual is fitted with the ANN model. Let $e_t$ be the residuals which can be obtained from the time series by subtracting forecasted value $l_t$ from ARIMA model as shown in Equation (4):

$$e_t = y_t - \hat{l}_t \quad (4)$$

The residual analysis is sufficient to capture the nonlinear patterns in the data. By modeling residuals using ANNs, nonlinear relationships can be discovered. Thus, a hybrid methodology that has both linear and nonlinear modeling capabilities can be a good strategy for practical use. By combining different models, different aspects of the underlying patterns may be captured.

## 3 Results and Discussion

### *3.1 Data Description*

In Table 1, the descriptive statistics of Singapore Airline Limited stock prices are presented.

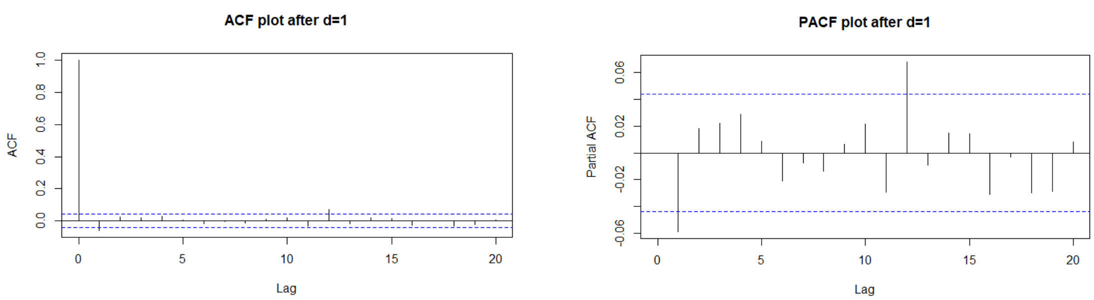
Table 1 Descriptive Statistics of Daily Closing Price of Singapore Airline Limited

| Mean | Median | Standard deviation | Variance | Minimum | Maximum |
|---|---|---|---|---|---|
| 10.6458 | 10.55 | 2.3864 | 5.6947 | 6.13 | 16.51 |

Table 1 show that the data obtained has a small variance, 5.6497. Due to variance, minimum and maximum value obtained from the data, this data potentially features the nature of majority stock price time series which is volatile and non-linear. The nonlinearity of the time series is examined from the residual of ARIMA model since ARIMA unable to capture the nonlinear structure of the time series. Thus, two methods were applied in this study: conventional method based on ARIMA model and hybrid ARIMA-ANN model.

### *3.2 Fitting of ARIMA Model*

ACF plot shows there are two significant spikes at lag 1 and 12 which reflects the order of Moving Average (MA) while PACF shows significant spikes at lag 1 and 11 which indicates the order of Autoregressive (AR). Thus, the suggested tentative models are ARIMA (1,1,0), ARIMA (1,1,1), ARIMA (1,1,12), ARIMA (1,1,0), ARIMA (1,1,1) and ARIMA (1,1,12). To find the best fitted model for this time series, trial and error need to be applied and tested in RStudio.




(a) ACF plot (b) PACF plot

Figure 2 Plot of the ACF and PACF after the first difference

According to the lowest AICc value obtained, the best fitted model is ARIMA(1,1,0) model. ARIMA(1,1,0) model obtain AICc value at -2398.04 and the equation for ARIMA(1,1,0) model is shown below:

$$y_t = y_{t-1} - 0.0583 y_{t-2} + \varepsilon_t \tag{5}$$

Mean Absolute Error, Root Mean Square Error and Mean Average Percentage Error of ARIMA(1,1,0) were obtained with the aid of RStudio as shown in Table 2 below.

Table 2 Evaluation criteria of in-sample data using ARIMA(1,1,0)

| | |
|---|---|
| MAE | 0.0900 |
| MAPE | 0.7754 |
| RMSE | 0.1329 |

For the checking of adequacy of ARIMA(1,1,0) in fitting the time series, Ljung-Box test is used to check the randomness of the residuals from ARIMA(1,1,0) model. From the result computed by RStudio, the *p*-value is 0.0161 which is below critical value, 5%. Therefore, there is evidence to reject null hypothesis at significance level of 5%. Ljung-Box test prove that the residuals are not independently distributed, and they are not white noise, so the residuals need to be further modeling for higher accuracy of forecasting performance. The diagnostics for the fit are displayed in Figure 3 and Figure 4 below.

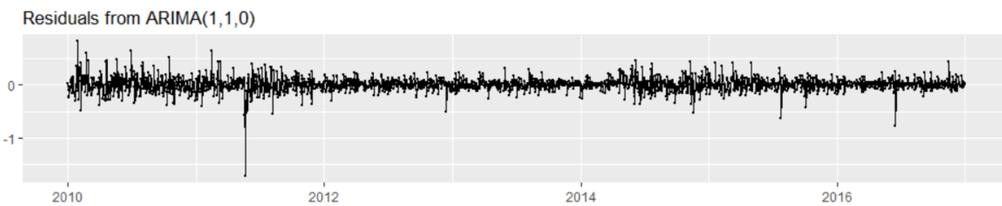



Figure 3 Residuals plot from ARIMA(1,1,0) model

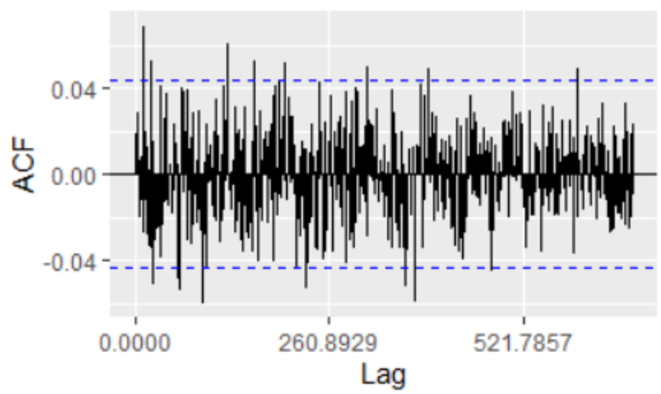



Figure 4 ACF plot of the residuals from ARIMA(1,1,0) model

According to Figure 3, the residual plot is abnormal as the time plot of residuals shows that the variation of residual is not consistent across the data while ACF plot of the residuals in Figure 4shows some of the autocorrelations are out of the threshold limits. Figure 4.6 prove that the residuals do not behave like white noise. Thus, even ARIMA(1,1,0) is chosen as the best fitted model, but the presence of heteroscedasticity in the residuals reflects that ARIMA(1,1,0) failed to capture the non-linearity behaviour in time series of the daily closing price of Singapore Airline Limited.

### *3.3 Fitting of Hybrid ARIMA-ANN Model*

Hybrid ARIMA-ANN model will focus on modelling the residuals from fitted value and forecasting value of ARIMA(1,1,0) by artificial neural network (ANN). Only single layer feedforward network is considered in this research to avoid the overfitting effect. The residuals are modelling by the training set and forecasted by test set for the next two years (2018-2019). After the forecasting output from ANN is computed, the forecasted residuals and the forecasted value of ARIMA(1,1,0) from 2018 to 2019 are combined to form a new forecasting output.

To develop Artificial Neural Network (ANN) for modelling the residuals, the residuals are scaled within the range from 0 to 1 by the min-max technique. The output is divided into two subsets at the ratio of 8:2. , the number of hidden nodes is data dependent, there is not certain formula to determine the number of hidden nodes should be applied in the network of ANN. The number of hidden neurons is decided by the square root of the sum of neurons used in input layer and output layer by referring to previous research. Thus, the number of hidden neurons is estimated approximately to 63 and more or less than 63 are examined to compare the performances by the technique of trial and error.

Table 3 Evaluation criteria of in-sample data using hybrid ARIMA-ANN model.

| Hybrid ARIMA(1,1,0)-ANN(63) model | |
|---|---|
| MAE | 0.0895 |
| MAPE | 0.0077 |
| RMSE | 0.1323 |

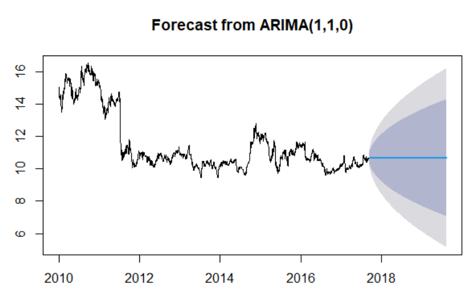
### *3.4 Forecasting*

The evaluation criteria of hybrid ARIMA-ANN model with ARIMA(1,1,0) and ANN with eleven selections of hidden neurons are presented in Table 4. All of trial and error is programmed in RStudio.

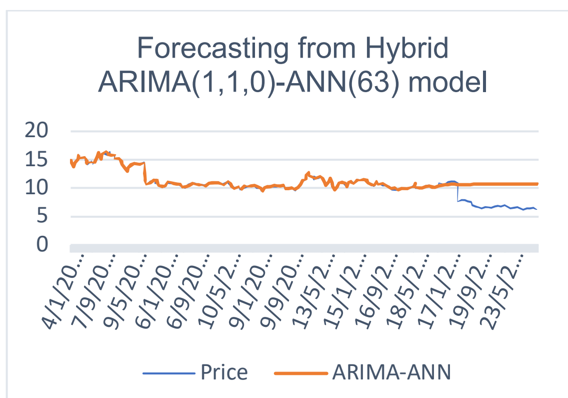
Table 4 Evaluation Criterion of Forecasting from Hybrid model.

| Hybrid model | MAE | RMSE | MAPE |
|---|---|---|---|
| 57 | 3.3075 | 3.4833 | 0.4121 |
| 58 | 3.3124 | 3.4921 | 0.4131 |
| 59 | 3.3251 | 3.5134 | 0.4421 |
| 60 | 3.3268 | 3.5861 | 0.4935 |
| 61 | 3.3518 | 3.5281 | 0.4501 |
| 62 | 3.2557 | 3.4922 | 0.3945 |
| 63 | 3.2446 | 3.4825 | 0.3874 |
| 64 | 3.2851 | 3.4832 | 0.3946 |
| 65 | 3.2763 | 3.5018 | 0.3967 |
| 66 | 3.2682 | 3.5201 | 0.3951 |
| 67 | 3.2651 | 3.5378 | 0.3987 |

The out-sample forecast is generated using ARIMA (1,1,0) and hybrid ARIMA (1,1,0)-ANN (63) for the next two years are plotted in Figure 5 below.




(a) ARIMA model




(b) ARIMA-ANN model

Figure 5 Forecasting results of (a) and (b)

To measure the performances of modelling and forecasting, we need to use accuracy criterions. There are a few types of accuracy measures are considered in this study, that are MAE, RMSE and MAPE. The purpose of using these evaluation criteria is to confirm the conclusions made regarding the best model. Table 5 tabulates the comparison values of MAE, RMSE and MAPE for ARIMA model and hybrid ARIMA-ANN models for out-sample from January 2018 to December 2019.

Table 5 Comparison Table between ARIMA model and Hybrid ARIMA-ANN model.

| Type of sample | Evaluation Criteria | ARIMA(1,1,0) | Hybrid ARIMA(1,1,0)-ANN(63) |
|---|---|---|---|
| Out sample | MAE | 3.3556 | 3.2446 |
| | RMSE | 3.5927 | 3.4825 |
| | MAPE | 0.4944 | 0.3874 |

Table 5 shows the results of RMSE, MAE and MAPE for ARIMA (1,1,0) and ARIMA(1,1,0)-ANN(63). The results present that the values of MAE, RMSE and MAPE for forecasting of ARIMA(1,1,0)-ANN(63) model is smaller than ARIMA (1,1,0) model. Hence, from the results, it can be concluded that hybrid ARIMA(1,1,0)-ANN(63) is the best model to be used to forecast the closing price of Singapore Airline Limited if compare with ARIMA(1,1,0) model and eleven suggested hybrid ARIMA-ANN model.

## 4 Conclusion and Recommendation

### *4.1 Summary and Conclusion*

Hybridization is proposed in this research to develop a model which have high capabilities in forecasting. Hybrid ARIMA-ANN model is introduced to overcome the limitation of a component model. First, ARIMA model is an approach to fit the ten years' time series in the model and forecasting. There are six tentative ARIMA models is suggested to accomplish the task, which is ARIMA (1,1,0), ARIMA (1,1,1), ARIMA (1,1,12), ARIMA (1,1,0), ARIMA (1,1,1) and ARIMA (1,1,12). After diagnostic checking by Ljung-Box test, ARIMA model is improper model to fit daily closing price of Singapore Airline Limited. Then, this research is carried out by hybrid ARIMA-ANN model for better forecasting performance. The residuals from ARIMA(1,1,0) model in previous stage are extracted to divided into two subsets, which is training set and test set for developing artificial neural network (ANN). Mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) is applied as the forecasting accuracy measures.

In conclusion, hybrid ARIMA-ANN model with ARIMA(1,1,0) and neural network of 63 hidden nodes is selected as the best hybrid model in fitting the data and forecasting. Mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) of this hybrid model achieved the lowest value in comparison of ARIMA(1,1,0) model and eleven suggested hybrid ARIMA-ANN model.

### *4.2 Recommendation*

Future research is encouraged to explore more opportunity in modeling and forecasting data using hybridization, such as ARIMA-GARCH family. Various software such as Python and Matlab is also recommended for the experience in application and the advanced technique applied. Prevention is better than cure. The purpose of forecasting is to reduce the risk and help in make wiser decision. Thus, we should always strive for better forecasting output and eliminate the limitation we faced in current situation. Although we forecast the future values, however there is always undiscovered factors that might affect the fluctuation of data. Apply trial and error when conduct experiment and think over when making decision.

## 5 References

[1] Adebiyi, A.A., Adewumi, A.O. and Ayo, C.K., 2014. *Comparison of ARIMA and artificial neural networks models for stock price prediction.* Journal of Applied Mathematics.

[2] Armano, G., Marchesi, M., & Murru, A. (2002). NXCS: Hybrid approach to stock indexes forecasting. In *Genetic Algorithms and Genetic Programming in Computational Finance* (pp. 125-158). Springer, Boston, MA.

[3] Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International journal of forecasting*, *8*(1), 69-80.

[4] Babu, C.N. and Reddy, B.E. (2014). *A moving-average filter-based hybrid ARIMA– ANN model for forecasting time series data.* Applied Soft Computing, 23, pp.27-38.

[5] Banerjee, D. (2014, January). Forecasting of Indian stock market using time-series ARIMA model. In *2014 2nd International Conference on Business and Information Management (ICBIM)* (pp. 131-135). IEEE.

[6] Büyükşahin, Ü.Ç. and Ertekin, Ş. (2019). *Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition.* Neurocomputing, 361, pp.151-163.

[7] Bustos, O. and Pomares-Quimbaya, A., 2020. *Stock market movement forecast: A systematic review.* Expert Systems with Applications, p.113464.

[8] Chan, W., Yung, Y. F., & Bentler, P. M. (1995). A Note on Using and Unbiased Weight Matrix in the ADF Test Statistic. *Multivariate Behavioral Research*, *30*(4), 453-459.

[9] Cheng, H. and Shi, Y. (2020). *Forecasting China's stock market variance.* Pacific- Basin Finance Journal, 64, p.101421.

[10] Contreras, J., Espinola, R., Nogales, F. J., & Conejo, A. J. (2003). ARIMA models to predict next-day electricity prices. *IEEE transactions on power systems*, *18*(3), 1014-1020.

[11] Díaz-Robles, L.A., Ortega, J.C., Fu, J.S., Reed, G.D., Chow, J.C., Watson, J.G. and Moncada-Herrera, J.A. (2008). *A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile.* Atmospheric Environment, 42(35), pp.8331-8340.

[12] Efendi, R., Arbaiy, N. and Deris, M.M. (2018). *A new procedure in stock market forecasting based on fuzzy random auto-regression time series model.* Information Sciences, 441, pp.113-132.

[13] Eni, D. (2015). Seasonal ARIMA modeling and forecasting of rainfall in Warri Town, Nigeria. *Journal of Geoscience and Environment Protection*, *3*(06), 91.

[14] Fama, E. F., & French, K. R. (1995). Size and book-to-market factors in earnings and returns. *The journal of finance*, *50*(1), 131-155.

[15] Faruk, D. Ö. (2010). A hybrid neural network and ARIMA model for water quality time series prediction. *Engineering applications of artificial intelligence*, *23*(4), 586-594.

[16] Harris, R. I. (1992). Testing for unit roots using the augmented Dickey-Fuller test: Some issues relating to the size, power and the lag structure of the test. *Economics letters*, *38*(4), 381-386.

[17] Hassani, H., & Yeganegi, M. R. (2019). Sum of squared ACF and the Ljung–Box statistics. *Physica A: Statistical Mechanics and its Applications*, *520*, 81-86.

[18] He, Q., Liu, J., Wang, S. and Yu, J. (2020). *The impact of COVID-19 on stock markets.* Applied Economics, pp.275-288.

[19] Liu, H., Tian, H.Q., and Li, Y.F. (2012). *Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction.* Applied Energy, 98, pp.415-424.

[20] M Asrah, N., Nor, M.E., A Rahim, S.N. and Leng, W.K. (2018). *Time series forecasting of the number of Malaysia Airlines and AirAsia passengers.* In Journal of Physics: Conference Series (Vol. 995, No. 012006, pp. 1-15). IOP Publishing.

[21] Mijwil, M. M., & Abttan, R. A. (2018). Artificial Intelligence: A Survey on Evolution and Future Trends. *Asian Journal of Applied Sciences (ISSN: 2321–0893)*, *9*(2).

[22] Petrică, A. C., Stancu, S., & Tindeche, A. (2016). Limitation of ARIMA models in financial and monetary economics. *Theoretical & Applied Economics*, *23*(4).

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