Advances and Applications in Statistics
© 2025 Pushpa Publishing House, Prayagraj, India
https://pphmjopenaccess.com
https://doi.org/10.17654/0972361725031
Volume 92, Number 5, 2025, Pages 749-766
P-ISSN: 0972-3617
E-ISSN: 3049-2165

PREDICTION OF GOLD PRICES USING HYBRID MODEL ARIMA-LSTM

Woon Kah Mun[1,2], Suliadi Firaus Sufahani[2,*], Anton Abdulbasah Kamil[3] and Mohd Kamal Mohd Nawawi[4]

[1]Huawei Technologies
Level 45, Exchange 106, Lingkaran TRX
55188 Tun Razak Exchange
Kuala Lumpur, Malaysia
e-mail: kahmunw1999@gmail.com

[2]Department of Mathematics and Statistics
Faculty of Applied Sciences and Technology
Universiti Tun Hussein Onn Malaysia
Pagoh Campus, 84600 Pagoh
Johor, Malaysia
e-mail: suliadi@uthm.edu.my

Received: October 6, 2024; Revised: October 31, 2024; Accepted: December 30, 2024
2020 Mathematics Subject Classification: 62P20, 60G25, 91B84, 62A09, 46N30.
Keywords and phrases: forecasting, prediction, gold price, ARIMA, LSTM.
*Corresponding author
Communicated by Kuan-Fu Shen

How to cite this article: Woon Kah Mun, Suliadi Firaus Sufahani, Anton Abdulbasah Kamil and Mohd Kamal Mohd Nawawi, Prediction of gold prices using hybrid model ARIMA-LSTM, Advances and Applications in Statistics 92(5) (2025), 749-766.
https://doi.org/10.17654/0972361725031
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Published Online: March 12, 2025

[3]Faculty of Economics
Administration and Social Sciences
Istanbul Nisantasi University
Maslak Mahallesi, Turkey
e-mail: akamil@gelisim.edu.tr

[4]School of Quantitative Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok, Kedah, Malaysia
e-mail: mdkamal@uum.edu.my

Abstract

Time series forecasting has been gaining attention since the COVID-19 pandemic to predict sales, economics, and weather outcomes. In this research, an empirical study on a time series model was done to predict daily gold prices using historical everyday gold prices from 1st January 2020 to 31st December 2021 as the training and testing dataset. The performance of autoregressive integrated moving average (ARIMA), long short-term memory (LSTM) and hybrid-model ARIMA-LSTM was compared through their mean absolute percentage error (MAPE) and root mean-squared error (RMSE) values. The results showed that the model with the smallest RMSE was the hybrid ARIMA-LSTM, but the model with the smallest MAPE was LSTM.

1. Introduction

According to Parisi's paper published in 2008 [1], gold has been recognised as a 'frontier-less currency', which can be traded anytime under any given circumstance. However, gold is still considered the most volatile of all exchanges as the price fluctuates rapidly, even in short timeframes. Though gold prices are not controllable, technical and fundamental techniques can be applied to forecast future prices for future decisions [2]. The financial market is a complex system with non-linear properties

influenced by several interconnected economic, political, and psychological factors [3]. Traditionally, models such as autoregressive integrated moving average (ARIMA) and vector aggressive models (VAM) predict price movements in time-series data. However, it has been found that there are many disadvantages in doing so due to its incapability to follow non-linear trends [4]. Because of their better predictability, machine learning approaches are gaining more attention in various industries.

In December 2019, an infectious disease, COVID-19 spread worldwide. The impact of this disease was not only on human health and lifestyle but also on global financial markets and economics [5]. Due to that, being able to forecast daily gold prices using predictive modelling is immensely crucial for investors and businesses to make decisions and maximise their profits [6]. ARIMA, also known as the *Box-Jenkins approach*, is widely known as the traditional approach to predicting time-series data [7]. It is a forecasting framework developed by Box and Jenkins [8] in 1970 and was widely used on time series data. However, many researchers have found that this method struggles with projecting non-linear mechanisms, which may cause low prediction accuracy as real-world data are often non-linear [9].

In this current day and age, machine learning approaches such as RNN, convolutional neural network (CNN), and multilayer perceptron (MLP). LSTM and RNN models were first proposed by Hochreiter and Schmidhuber in 1997 [10]. Ever since then, many researchers have popularly equipped and improved this method due to the predictive performance of autoregressive integrated moving averages and deep learning algorithms [11]. The ARIMA-LSTM hybrid model was introduced in 2003 in an empirical study by Hansen and Nelson on various time-series data [12]. This hybrid model employs a combined approach, where the ARIMA method is applied first to capture linear tendencies in the data. In contrast, long short-term memory (LSTM) is used to capture non-linearity in the residuals [13].

This paper aims to investigate the trend of the gold prices' time series data by reviewing the accuracy of ARIMA, LSTM and hybrid ARIMA-

LSTM modelling predictions. By applying these models, the future daily gold price could be estimated and be a good indicator of the market situation. Lastly, the accuracy of all the methods mentioned will be compared to determine the best model.

2. Materials and Method

2.1. Dataset

The daily data for gold price is obtained from Yahoo Finance (https://finance.yahoo.com). The gold price data obtained ranges from 1st January 2020 to 31st December 2021, exactly two years of daily price data (730 data size) when COVID-19 was new to the world. No small dataset or other data involve. This duration was chosen because COVID-19 was recent and dramatically affected the price of gold, triggering unpredictable movements. Hence, taking this period as a study would provide better insight into the models' capability in uncertainty. For future research, it would be beneficial to incorporate data from additional years, ideally spanning both stable and volatile periods, to capture a more balanced view of market dynamics.

The price of gold is measured in dollars, also known as the United States Dollar (USD). Besides that, since the price of gold constantly fluctuates due to the continuous trading in the market, the daily closing price is used as the forecast target. After retrieving the data, the data will first be cleaned by removing missing data and would then be split into two parts, the training set and the testing set, by an 80:20 ratio to avoid overfitting and validate through *k*-fold cross-validation [14]. Missing values are handled by removing the affected rows, effectively cleaning the data. However, only a few rows are affected based on the received data. The validation process Furthermore, the statistical software used in this research would be Python. The trend is then observed by graphical plotting, which would provide a general view of the stability of the data set.

2.2. ARIMA

ARIMA method is a statistical model that is used for time-series data forecasting. The ARIMA equation is based on the regression type equation where the independent variables are lags of the dependent variable. The ARIMA is represented as $Y(t)$ in the following equation:

$$Y(t) = \alpha + \beta_1 Y_{t-2} + \cdots + \beta_p Y_{t-p} \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \cdots + \phi_q \varepsilon_{t-q}. \quad (1)$$

By deconstructing equation (1), the autoregressive (AR) function and the moving average (MA) function can be obtained and are represented as the following:

$$MA = \alpha + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \cdots + \phi_q \varepsilon_{t-q}, \quad (2)$$

$$AR = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon_t, \quad (3)$$

where α denotes the intercept constant term, and ε represents the standard errors in the equations (1), (2) and (3). Hence, ARIMA can be split into three parts based on the model's name: AR, Integrated (I), and MA, which can also be denoted as p, d, and q of the model.

AR regresses the time-series data with the previous value, and the order of this lag is denoted as p. To determine the value needed for p, the partial autocorrelation (PACF) plot will be analysed. Then, the integrated part determines the order of differencing required to ensure the data is stationary. This part is denoted as d. The Augmented Dickey-Fuller (ADF) test will be conducted to confirm the stationarity of the data. Lastly, MA is the part where the data is regressed one more time but in relation to the residuals of past observations. The order of the error log for this part is denoted as q. The values will be obtained to represent d and q by analysing the autocorrelation plot (ACF). After determining the model for ARIMA, the Akaike information criteria (AIC), Bayesian information criteria (BIC) and Ljung-Box would be used to determine if the ARIMA model is the best fit.

2.3. Long short-term memory

Long short-term memory (LSTM) (as in Figure 1) is an RNN model that uses previously hidden nodes as data to learn continuously. LSTM follows three steps: the forgetting layer, the input layer, and the output layer [15].

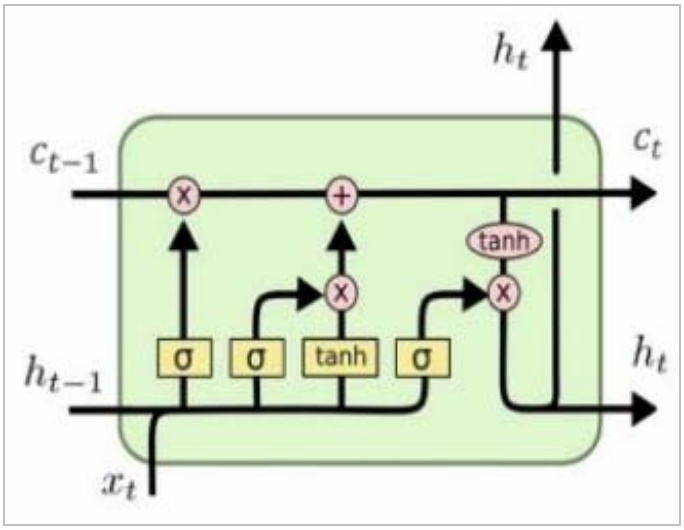


Figure 1. Structure of LSTM cell state.

The forgetting layer controls the information that will be thrown away from the cell state or memory, defined as the forget gate, f_t, with the equation:

$$f_t = \sigma(W_{xf} x_t + W_{hf} h_{t-1} + W_{cf} c_{t-1} + b_f), \tag{4}$$

where σ is the sigmoid activation function. The input layer determines whether the information is useful, acting as a storage for essential values, defined as the input gate, i_t, with the equation:

$$i_t = \sigma(W_{xf} x_t + W_{hf} h_{t-1} + W_{cf} c_{t-1} + b_i), \tag{5}$$

$$\tilde{c}_t = \tanh(W_{xc} x_t + W_{hc} h_{t-1} + b_c), \tag{6}$$

$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t, \tag{7}$$

where *tanh* is the tanh activation function, known as the hyperbolic tangent function, *W* refers to the weight matrix, *b* refers to the bias vector, and $\widetilde{c}_t$ is the matrix with a new candidate value that can be added to the cell state. Lastly, the output layer is where the cell determines the result of the sequence, which is defined as the output gate, o_t, with the equation:

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o), \tag{8}$$

$$h_t = o_t \times \tanh(c_t), \tag{9}$$

where h_t is denoted as the output vector of the LSTM unit.

2.4. ARIMA-LSTM

ARIMA-LSTM hybrid model is structured together to work on linearity and non-linearity in the data. The formula x_t can be sketched as the following:

$$x_t = L_t + N_t + \varepsilon_t, \tag{10}$$

$$L_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p}\varepsilon_t + \phi_1\varepsilon_{t-1} + \phi_2\varepsilon_{t-2} + \cdots + \phi_q\varepsilon_{t-q}, \tag{11}$$

$$N_t = o_t \times \tanh(c_t), \tag{12}$$

where L_t denotes the linearity of the data at time *t* using equation (1) while N_t refers to the non-linearity following equation (9). The model is created through equation (10) by obtaining the forecasted values from ARIMA as L_t. Then, the forecast residuals will be inserted into the LSTM model to determine the non-linearity.

2.5. Model evaluation

The forecasted results will be compared to determine the best model out of all the methods. To assess the performance of the proposed models, the root-mean-squared error (RMSE) and the mean absolute percentage error

(MAPE) of each model are obtained and compared. The mean square error (MSE) refers to the sum of all squared errors divided by the sample size, with the formula:

$$RMSE = \sqrt{\sum_{t=1}^{n} \frac{(y_t - \hat{y}_t)^2}{n}}. \tag{13}$$

The mean absolute error (MAE) refers to the accumulation of all errors divided by the number of days forecasted, with the formula:

$$MAPE = \sum_{t=1}^{n} \frac{| y_t - \hat{y}_t |}{y_t} \times \frac{100\%}{n}, \text{ where } y_t \neq 0. \tag{14}$$

Two evaluation metrics are done to better understand the forecasted values' accuracy. This is because MAPE explains the percentage of errors in the data while RMSE explains the average squared error.

RMSE and MAPE are popular evaluation metrics due to their simplicity, ease of interpretation, and ability to highlight different aspects of prediction error. RMSE measures the average error magnitude in the target's units, making large deviations more impactful, while MAPE provides a percentage-based error, which is useful for comparing performance across different scales. Both metrics are widely applicable and accessible, making them standard choices for quick assessments of prediction accuracy. However, they mainly focus on accuracy, and additional metrics are needed to evaluate generalization and prediction interval reliability comprehensively.

2.6. Flowchart

Figure 2 shows the overall plan of this study's procedure. First, the ARIMA and LSTM models will be constructed, and the MAPE and RMSE of the results will be tabulated. Then, the hybrid model would start with the residuals from the ARIMA model and continue with the LSTM model as before.

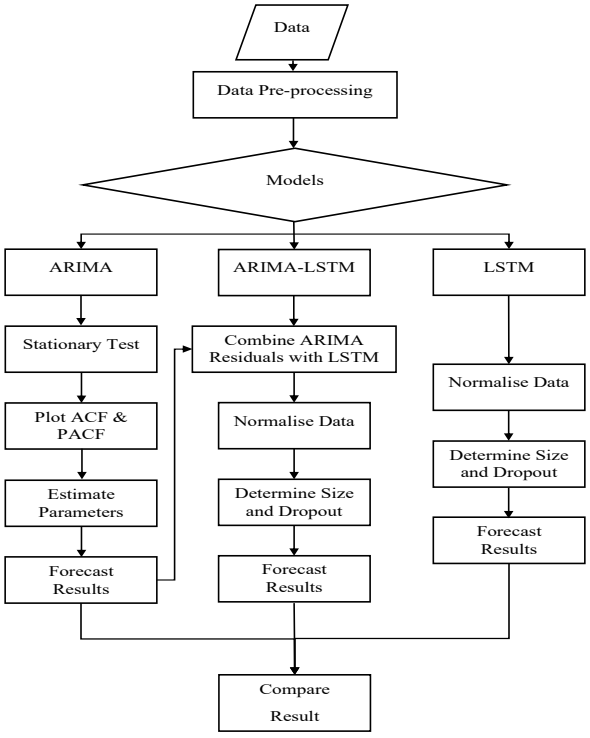


Figure 2. Flowchart of research.

3. Results and Discussion

Figure 3 shows the overview of daily gold prices for the two years. Based on Figure 3, the trend of gold price was linearly increasing from January to July 2020 before it dived until February 2021. In the following days, gold price movements did not fluctuate as fast as previously, but it was still slowly increasing.

Altogether, there are 517 data points in two years (as shown in Table 1). Hence, based on the 80/20 split, the training and testing datasets would have 414 and 103 data points, respectively. The range of the data is 17179, with a standard deviation of 3110.09. This indicates that the data contains non linear properties, making it very volatile. Thus, using this dataset as a comparative analysis for forecasting models is suitable.



Figure 3. Time series of daily gold price from 1st January 2020 to 31st December 2021 (USD).

Table 1. Descriptive statistics of the model

Descriptive statistics	
Count	517
Mean	47306.4
Standard deviation	3110.09
Min	38938
Max	56117

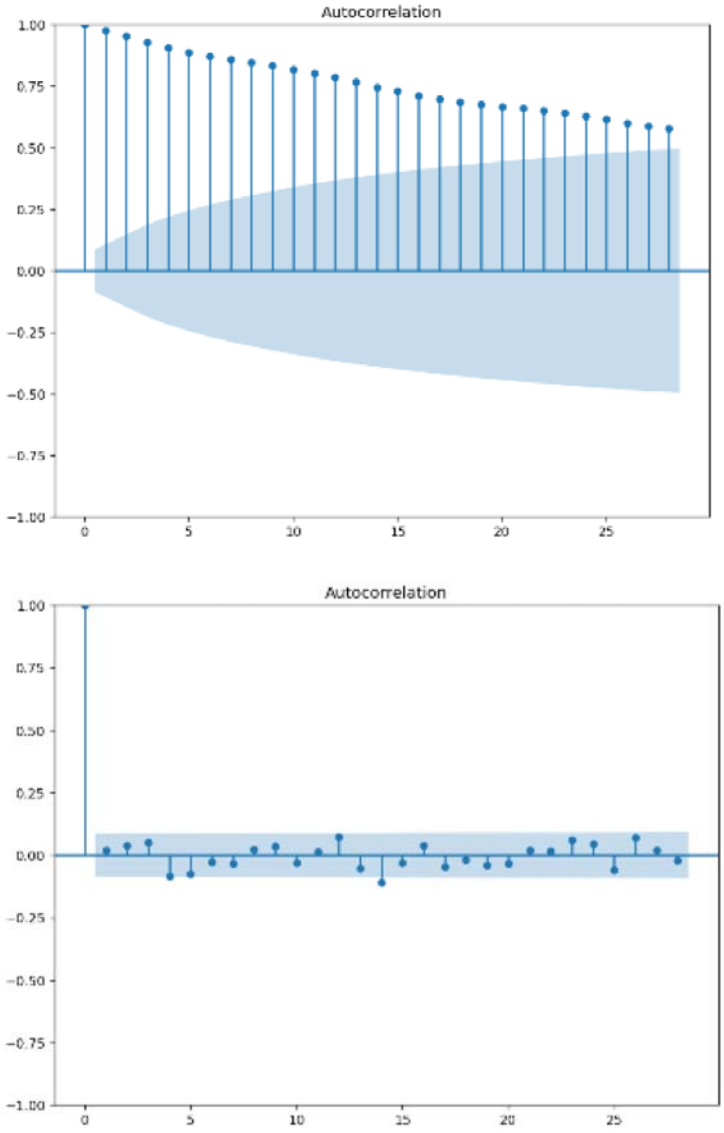
3.1. Forecast using ARIMA

Based on Table 2, the *P*-value of the original dataset is more than 0.05, indicating that the data is non-stationary. Hence, the first and second orders of differencing are placed as a focus since the *P*-value obtained is less than 0.05, indicating stationarity.

Table 2. ADF test

	ADF statistics	*P*-value
Original	-2.449	0.128
First difference	-11.095	0.000
Second difference	-4.417	0.000

In Figure 4, the autocorrelation plot is obtained. The first and second orders showed more consistent lags than the original datasets, confirming ADF's hypothesis. However, the first point in the second order of differencing shows a big jump in value, which indicates over-differencing. Hence, first-order differencing is ideal.



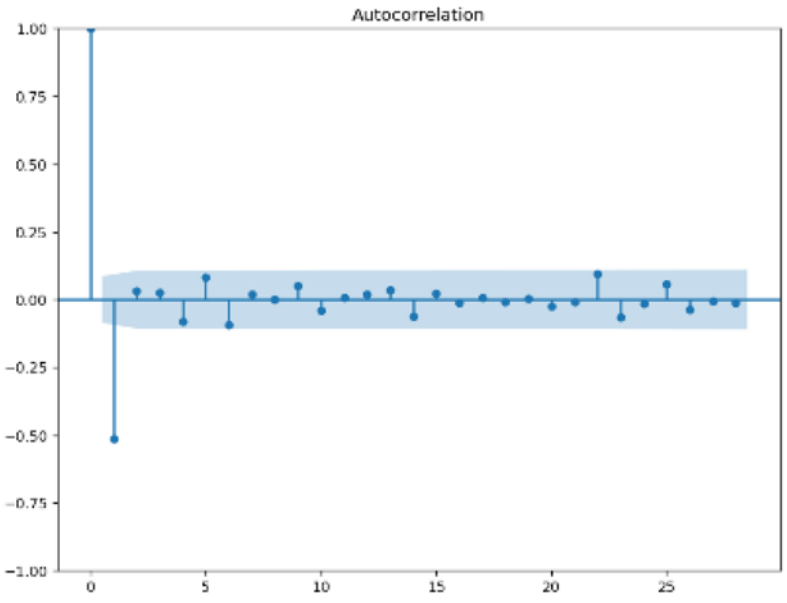


Figure 4. Autocorrelation plots.

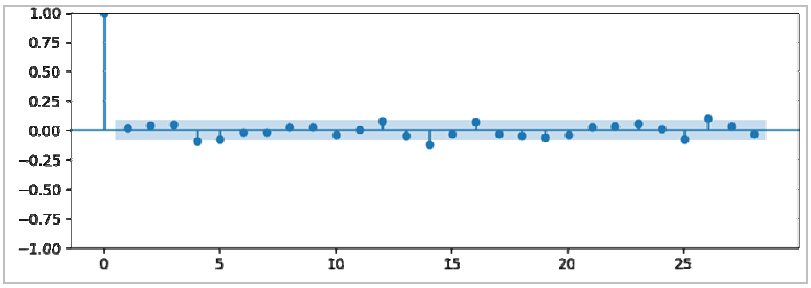


Figure 5. Partial autocorrelation plot of second-order differencing.

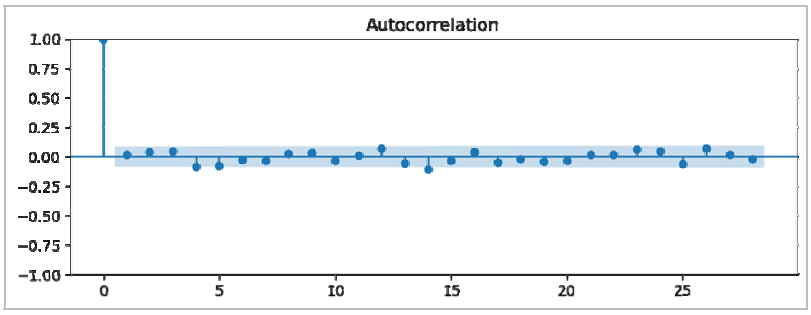


Figure 6. Autocorrelation plot of first-order differencing.

Based on Figures 5 and 6, ARIMA $(0, 1, 0)$, ARIMA $(1, 1, 1)$, and ARIMA $(2, 1, 2)$ are proposed, as the lag moves in the same direction before changing in both plots.

Table 3. ARIMA models

ARIMA model	AIC	BIC	Ljung-Box
(0, 1, 0)	7805.57	7809.82	0.21
(1, 1, 1)	7807.74	7820.48	0.36
(1, 1, 2)	7807.31	7824.299	0.02
(2, 1, 1)	7807.91	7824.89	0.05
(2, 1, 2)	7790.68	7811.91	0.00

To assess the feasibility of the proposed ARIMA models, the AIC, BIC, and Ljung-Box test are obtained to check the autocorrelation of the residuals. Since ARIMA $(2, 1, 2)$ has the lowest AIC, BIC, and Ljung-Box values, it is considered the best model, as shown in Table 3.

3.2. Forecast using LSTM

After testing the LSTM model with different units, epochs, and batch units, the best LSTM model was found to have layers of 100 units, two epochs, and 23 batch sizes (as shown in Figure 7).

```
model = keras.Sequential()
model.add(layers.LSTM(100,return_sequences=True,
input_shape=(x_train.shape[1],1)))
model.add(layers.LSTM(100, return_sequences = False))
model.add(layers.Dense(25))
model.add(layers.Dense(1))
model.summary()
```

Figure 7. LSTM model.

3.3. Forecast using ARIMA-LSTM

The residuals from the ARIMA model's prediction are used for this section. By using the residuals as the dataset, the data were split into two sets: a training set with 413 data points and a testing set with 103 data points. The training set has one less value compared to the input of ARIMA and LSTM models because the first value does not have a predicted value. Hence, the first row is removed. Similarly, the residuals are tested with different units, epochs, and batch units. Ultimately, the prediction with the smallest RMSE obtained is 200 units, three epochs, and 15 batch sizes (as shown in Figure 8).

```
model = keras.Sequential()
model.add(layers.LSTM(200,return_sequences=True,
input_shape=(x_train.shape[1],1)))
model.add(layers.LSTM(200, return_sequences=False))
model.add(layers.Dense(15))
model.add(layers.Dense(1))
model.summary()
```

Figure 8. ARIMA-LSTM model.

3.4. Models evaluation

Based on Table 4, ARIMA-LSTM has the lowest RMSE value, but LSTM has a lower MAPE value. By focusing on Figures 9, 10 and 11, LSTM can explain the data better than ARIMA-LSTM. Focusing on the MAPE value, ARIMA-LSTM's model computes the highest percentage error compared to the other two models. Hence, though the RMSE may be the lowest, it may not be a good model as ARIMA-LSTM performed worse than ARIMA in terms of MAPE value.

Table 4. Accuracy results of the models

Model	RMSE	MAPE
ARIMA	952.27	1.62
LSTM	560.71	0.01
ARIMA-LSTM	204.01	33.29

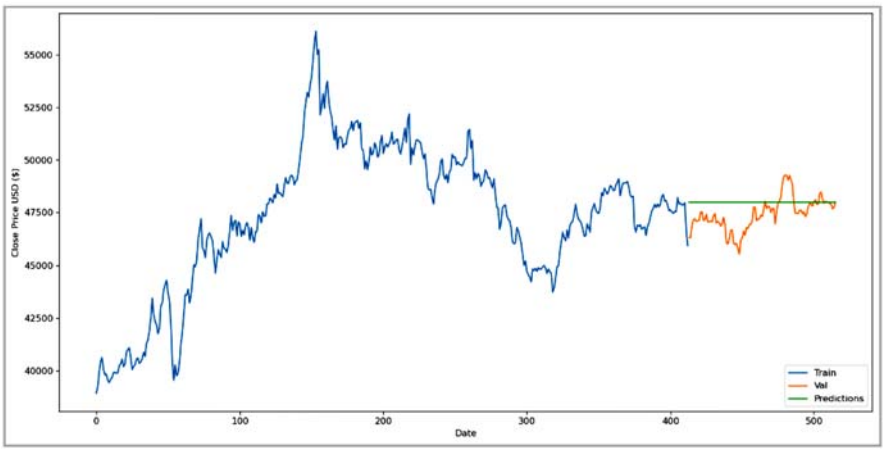


Figure 9. ARIMA forecast.

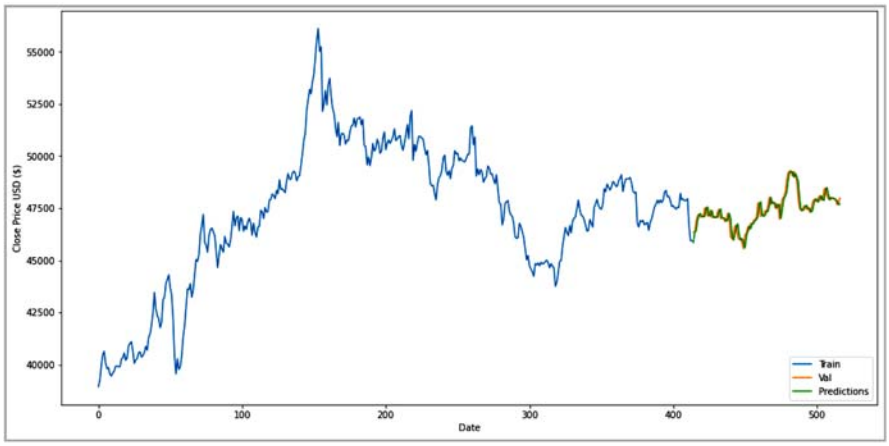


Figure 10. LSTM forecast.

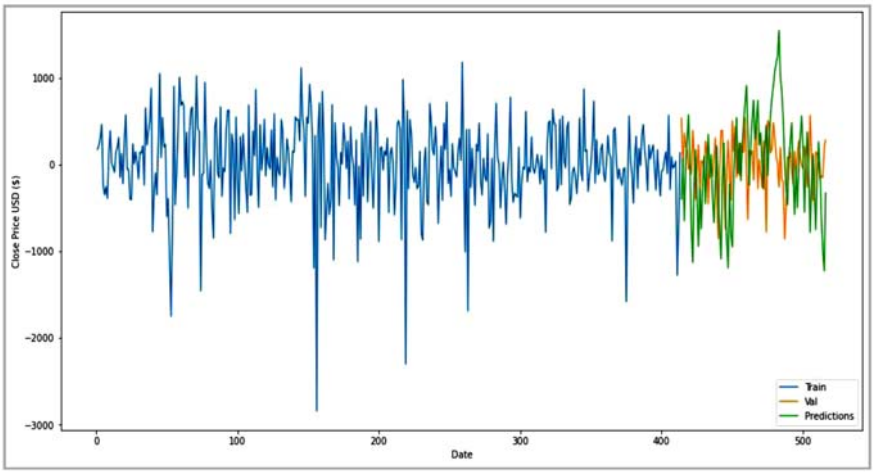


Figure 11. ARIMA-LSTM residual forecast.

4. Conclusion

Hybrid models have shown practical approaches to accurately tackling time series prediction. However, the performance of the models may not always be the same as the expected results. In this paper, though the hybrid model ARIMA-LSTM has a lower RMSE value, the model has a very high MAPE value, indicating that the predicted values fluctuate between over-positive and over-negative very quickly. Although it was proven that hybridisation could reduce the model uncertainty, the residual from the ARIMA model was similar to white noise, resulting into the LSTM model struggling with prediction. Thus, even though ARIMA-LSTM yields better results compared to the mere LSTM model and the steps of ARIMA-LSTM may be tedious and redundant. Nevertheless, LSTM models could be further used with sentiment analysis and financial indicators to improve prediction accuracy. To get the root of sentiments, financial astrology and behavioural bias could be used alongside neural network models to provide a more accurate and precise time forecast. Besides that, price prediction is tricky as it outputs specific values or data. Hence, forecasting price movement or direction would be more valuable and accurate.

Acknowledgements

The authors would like to thank Universiti Tun Hussein Onn Malaysia for the MTUN Research Grant Q622 and its support while completing this research.

The authors also thank the anonymous referees for their valuable suggestions and comments which led to the improvement of the paper.

References

[1] A. Parisi, F. Parisi and D. Díaz, Forecasting gold price changes: Rolling and recursive neural network models, Journal of Multinational Financial Management 18(5) (2008), 477-487. https://doi.org/10.1016/j.mulfin.2007.12.002.

[2] M. Massarrat, Forecasting of gold prices (Box Jenkins Approach), International Journal of Emerging Technology and Advanced Engineering 3(3) (2013), 662-670. http://dx.doi.org/10.18488/journal.1/2016.6.11/1.11.614.624.

[3] U. JuHyok, P. Lu, C. Kim, U. Ryu and K. Pak, A new LSTM based reversal point prediction method using upward/downward reversal point feature sets, Chaos, Solitons and Fractals 132 (2020), 109559. https://doi.org/10.1016/j.chaos.2019.109559.

[4] R. Zhang, C. Zhang and M. Yu, A similar day based short term load forecasting method using wavelet transform and LSTM, IEEJ Transactions on Electrical and Electronic Engineering 17(4) (2021), 506-513. http://dx.doi.org/10.1002/tee.23536.

[5] M. K. Mishra, The World after COVID-19 and its impact on Global Economy, Working Paper, 2020. https://hdl.handle.net/10419/215931.

[6] M. K. Ho, H. Darman and S. Musa, Stock price prediction using ARIMA, neural network and LSTM models, Journal of Physics: Conference Series, 1988(1) (2021), p. 012041. https://doi.org/10.1088/1742-6596/1988/1/012041.

[7] M. Khashei and M. Bijari, An artificial neural network (p, d, q) model for time series forecasting, Expert Systems with Applications 37(1) (2010), 479-489. https://doi.org/10.1016/j.eswa.2009.05.044.

[8] G. Box and G. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco, Reference Book, 1970.

[9] G. Zhang, B. E. Patuwo and M. Y. Hu, Forecasting with artificial neural networks: The state of the art, International Journal of Forecasting 14(1) (1998), 35-62. https://doi.org/10.1016/S0169-2070(97)00044-7.

[10] S. Hochreiter and J. Schmidhuber, Long short-term memory, Neural Computation MIT Press, 9(8) (1997), 1735-1780. http://dx.doi.org/10.1162/neco.1997.9.8.1735.

[11] P. Maneejuk and W. Srichaikul, Forecasting foreign exchange markets: further evidence using machine learning models, Soft Computing 25(12) (2021), 7887-7898. https://doi.org/10.1007/s00500-021-05830-1.

[12] J. V. Hansen and R. D. Nelson, Forecasting and recombining time-series components by using neural networks, The Journal of the Operational Research Society 54(3) (2003), 307-317. https://doi.org/10.1057/palgrave.jors.2601523.

[13] H. K. Choi, Stock price correlation coefficient prediction with ARIMA-LSTM hybrid model, 2018. arXiv preprint arXiv:1808.01560. https://doi.org/10.48550/arXiv.1808.01560.

[14] R. Kumar, P. Kumar and Y. Kumar, Three stage fusion for effective time series forecasting using Bi-LSTM-ARIMA and improved DE-ABC algorithm, Neural Computing and Applications 34(21) (2022), 18421-18437. https://doi.org/10.1007/s00521-022-07431-x.

[15] A. Mahendra, S. P. Mohanty and S. Sudalaimuthu, Financial astrology and behavioral bias: evidence from India, Asia-Pac Financial Markets 28 (2021), 3-17. https://doi.org/10.1007/s10690-020-09310-8.