

The Stock Price Forecasting Based on Time Series Model and Neural Network

Boming Zhang *

College of Arts and Science, New York University, New York, United States

*Corresponding author: bz2196@nyu.edu

Abstract. The ARIMA model and LSTM model are commonly seen in stock price forecasting, and their excellent performance in processing and forecasting time series data has made them widely used and popular. They are used to predict future values with relatively high accuracy and small error values. Based on the excellent performance of LSTM model and ARIMA model on time series, this paper discusses the use and findings of ARIMA model and LSTM model in empirical studies and indicates whether they can be used for stock forecasting. This paper highlights the advantages and disadvantages of ARIMA model and LSTM model by comparing and summarizing the different applications. This paper finally finds that ARIMA model and LSTM model have high accuracy in predicting stock market. In this paper, the performance results of the ARIMA model and the LSTM model are extensively illustrated by using different approaches to the use of these two models. It also shows the accuracy of ARIMA model and LSTM model.

Keywords: ARIMA model; LSTM model; stock price.

1. Introduction

The global stock market is growing at a rapid pace. It plays an important role in the market by creating wealth and promoting economic growth. In the past ten years, the global stock market capitalization increases from \$4.7 trillion to \$15.2 trillion. In the other words, the emerging market was soared from 4 percent to 13 percent of total global value [1]. The stock market promotes the role of developed and developing countries in economic performance. In the present, although the equity market might be susceptible with foreign capital spillovers, interest rate risk and foreign exchange shocks, every complex equity market can make the benefits or profit for the economies [2]. Furthermore, there are two foundation reasons that illustrate the significant of stock market forecasting. First, it can help investors to anticipate and avoid potential market risks. Second, it creates a new opportunity to profitability for arbitrageurs and speculators. This means that it has some ability to predict stock returns when the forecast strength is realistic [3].

Stock price forecasting has a significant impact in finance and economics, and better forecasting models have been developed over the years. So far, ARIMA model and LSTM model are the two classical forecasting models. In many studies, hybrid models have been used. Hyeong Kyu Choi [4] passed linear data filtered by ARIMA model into LSTM model and concluded that this hybrid model contains both linear and nonlinear and largely outperforms other prediction models in terms of predictive power. Zhiwen Zeng, Matloob Khushi [5]. An attention based ARNN and ARIMA model is proposed to stabilize the data by removing the noise using wavelets. The ARNN is used to capture the nonlinear relationships and then ARIMA is used to fit the linear relationships about these data and by mixing these models a model is constructed that can better predict the trend of the market. Ha Young Kim, Chang Hyun Won [6]. A new hybrid model is proposed that combines the LSTM model with the GARCH model. It has excellent sequential pattern learning, and excellent stock prediction performance. The proposed method combines more than one econometric model, this method is a good combination of neural network model and time series to predict stock market.

This paper is devoted to introducing ARIMA model and LSTM model, summarizing the methods and experiences of previous people using the models, and giving an analysis. How these models are good at predicting the trend of stocks in the stock market. In section 2, this study will describe the definition of time series and the meaning of price forecasting. In section 3, the ARIMA model will be discussed and several studies will be provided to summarize the previous applications and results.

In Section IV, this paper will introduce the use of the LSTM model and provide several studies that summarize the results generated in the study. Section 5. will summarize the current limitations of the ARIMA model and the LSTM model through the issues and conclusions generated in the study. In Section 6 will summarize all the conclusions drawn by the ARIMA model and the LSTM model and provide concluding remarks.

2. Time Series and Price Prediction

2.1 Data

Over the years, most studies have dealt with time series and price forecasting, and a time series is identified as a numerical organism that has continuity and continuity, and it is constantly updated with what has come before. In simple terms, it is a set of data arranged in a chronological order [7]. It can predict future values by existing values. The time series contains stock prices, house prices, and all the values that need to change with time. Moreover, these data are collected from different time periods. Additionally, these data may be correlated since they are collected in adjacent time periods. For novice investors, their experience is not enough to make an investment decision, the stock market is very risky, but these time series prediction models help novice investors to decide. Similarly, price forecasting is an algorithm that forecasts and analyzes based on markets, products, and demand. These prices are always changing, and it needs to forecast and set prices to levels that will attract customers while maximizing sales more accurately. Otherwise, it can determine market trends and predict the value of a product based on data. All of this is achieved through predictive models that enable the machine to better provide price forecasts and improve sales efficiency. It can reduce the risk posed by the market and give accurate forecasts, bringing a higher probability of profitability.

2.2 ARIMA model & LSTM model

ARIMA model is one of the most widely used linear models. It is a type of statistical algorithm that captures the common time dependence of time series data. ARIMA model is excellent in terms of precision and accuracy. It has accuracy and reliability in short-term forecasting. In comparison with the latent loss neural network, the ARIMA model performs more accurately, with better inversion detection and a long-term prediction error of 0.6. When choose the appropriate alpha values, which is a time series model that can predict and repair failures [8]. LSTM model is a variety of recurrent neural networks, on sequence prediction, it can continuously learn the dependencies. Besides, it can handle the whole data sequence. When the LSTM model is used to predict gold prices and trends, its predictions are widely accepted and used, it feeds back the regression problem with the best predictions and feeds back the best performance of the classification problem about gold trends, and its conclusions outperform the traditional time series [9].

3. Results & Discussion

3.1 ARIMA

ARIMA measures and predicts future value based on linear combinations of past given data sets. Particularly speaking, non-stationary data is applied that takes no account on time variables, but only time intervals [10]. Its capability to integrate multivariate factors, determine effects of the factors on the values that are not subject to seasonality and recognize the dependency of observed value to errors endows it with variety range of uses in real lives, e.g., the daily closing price of a stock, the height of ocean tides, and the number of sunspots [11]. The expression is given as follows:

$$y_t = \mu + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t \quad (1)$$

An application was to predict international gold price in 2018 by Xiaohui Yang using ARIMA model. He evaluated ARIMA's effectiveness and short comes by comparing actual values with

predicted values by implementing both in-sample and out-sample data. After achieving stationary data in first-order differential, the author regressed in-sample data from June 25th, 2018, to June 29th 2018 based on ARMA (3,1,2) with the following data expression:

$$y_t = 0.0035 + 1.425Y_{t-1} - 0.910Y_{t-2} + 0.029Y_{t-3} + \epsilon_t - 1.457\epsilon_{t-1} + 0.906\epsilon_{t-2} \quad (2)$$

By plotting the real and expected gold price, it is significantly noticed that the forecasting data is comparably lower than the actual data with large flows of variability. However, the relative error remains small, indicating a successful use of ARIMA models.

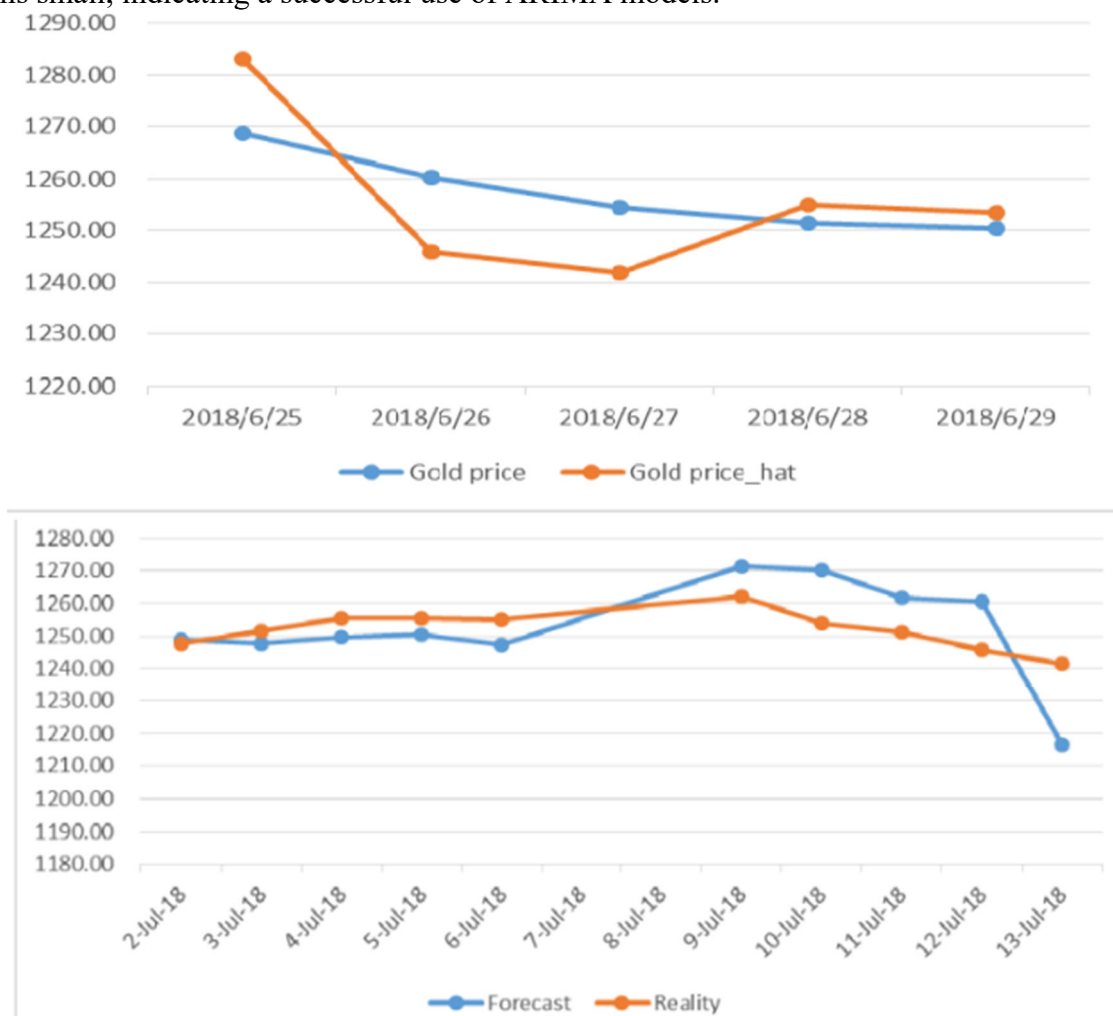


Fig. 1 Gold price from 2018/6/25 to 2018/6/29 (upper panel) and 2018/7/2 to 2018/7/13 (lower panel).

Now considering the out-sample data with a great range of time from July 2nd 2018 to July 13th 2018 (shown in Fig. 1). The forecasting price line remain relatively stable regarding to that of in-sample data. Taking a closer analysis, the relative residual gradually becomes larger especially after July 7th 2018 which indicates a potential disadvantage of ARIMA model – fails for long-term prediction. From the Fig. 1, noticeable gaps exist which mainly contribute to variants that are not considered in the linear model. Numerous factors including inflation, exchange rates and manufacturing capacity etc. or in combination can incur nonlinear impacts on gold price [12]. With ARIMA-invention strategy, it is possible to analyze Chinese market stock over lengthy period accompanying with determining the effectiveness fluctuated economic movements for future decisions. The inclusion of invention in the modelling means introducing a dummy variable equal to 1 when the corresponding factors are in effect and vice versa. In this research, data are collected over 10 years with relatively large samples, long enough to lessen uncontrollable market changes which serves a way to eliminate problem raised in the first case and big enough to make the prediction significant. Here, intervention variable is assigned to 1 after July 2007 corresponding the financial crisis. After calculation, the following statistics and AR(1) model are presented in Table. 1.

Based on the results, the proportional relationship between prices and financial crisis with relatively small S.E. of regression. This successfully predict the reality and is align with economic behaviors. Comparing this model with that without use of intervention, AIC, Schwarz Criterion and Hannan-Quinn Criterion are all significantly smaller, indicating a better model fit and prediction. This apparently small numbers are indeed significant enough to make a difference after considering the size variable D. Overall, this study has shown how ARIMA-intervention models solves the limitation of ARIMA models to predict effectively only in short time periods.

Table. 1 Summary of the regression.

Variable	Coefficient	Std	T-value	p-value
D	0.50	3.92e-5	12270.43	0.00
C	9.94	0.26	38.30	0.00
Volume	2.30e-9	3.07e-10	7.49	0.00
Inter	-2.32	0.08	-30.72	0.00
AR(1)	0.99	8.30e-5	11948.97	0.00
Adjusted R ²		0.99	AIC	5.08

3.2 LSTM

LSTM model is particularly suitable for making some long-term predictions, and it is suitable for building larger neural networks through its good performance as a nonlinear model. This paper by forecast stock price comparing the traditional RNN model demonstrate that LSTM model has more stability and accuracy. In the experiment, the daily data of SSE and Dow Jones indices from Oct. 31, 2007, to Oct. 31, 2017, are divided into input layer, hidden layer and output layer by collecting the data from Oct. 31, 2007 to Oct. 31, 2017. The nodes are set to 10. furthermore, the LSTM model four nodes to control the memory unit c, which is used to keep this information for a long time. The results are presented in Table. 2. After experimenting with the resulting images, the observation shows that the LSTM model is more stable, and the LSTM model shows less fluctuations in some turning points. In addition, by comparing the degree of bias of RNN and LSTM, LSTM has less bias. This also indicates that the prediction value of LSTM model is stronger than the traditional prediction model [14].

Table. 2 Reuslts of LSTM and RNN.

	LSTM	RNN
Shanghai Composite Index	7.55e-3	7.92e-3
Dow Jones Index	6.80e-3	8.02e-3

Likewise, in this research, it is still comparing RNN and ISTM forecasts in the stock market, except that the authors also count people with more outside influences, such as buyer and customer sentiment, weather, and policy. They construct multiple models for comparison and find the model that makes the lowest loss. After the experiment, the LSTM has a loss of 3.1464e-04, which has a lower loss compared to other models. They brought the real data from APPL, GOOG and TSLA into the model, removed the missing values and some other effects, and then evaluated the simulation. It was concluded that the LSTM model has more accurate results. They used Markowitz portfolio optimization to calculate the total amount of money they could invest in the market to get more benefit, which is productive for investors and reduces investment risk [15].

In addition, the LSTM model can be compared with many other models. In this study, they constructed 4 CNN and 5 ISTM models to perform simulation tests. First, the stock price records at five-minute intervals from December 31, 2012, to January 9, 2015, were used as the data in the model. The LSTM was then divided into four prediction models with univariate input data for the previous week, univariate input data for the previous two weeks, univariate input data for the previous three weeks, and univariate input data for the previous four weeks. the LSTM model will have 200 input nodes, 100 output nodes when it passes through the dense layer, and 5 nodes will be output when it

reaches the output layer. The test yielded an execution time of 316.74 seconds for LSTM model #1 and an average error value of 0.0059. depend on the test, although these models differ in execution speed, the data that the model generate are still very accurate [16].

4. Limitations & Prospects

In ARIMA model, it requires that the values satisfy a linear relationship to better predict future values, but most of the data are indeed nonlinear. This also affects the prediction of the ARIMA model. moreover, the ARIMA model is not very suitable for long term prediction, and it is difficult to predict the turning points, it is very subjective when one determines the order of the data used in the model. Similarly, in LSTM model, it requires more data to learn and simulate, and the prediction time is longer. This also limits the prediction of neural networks. During the testing phase, the LSTM model may be overfitted and performance degraded. The data is not easy to be modified. As more models and technologies become available, one can expect to get more accurate data by combining those models and taking the best of the best. For example, it can talk about the combination of ARIMA model and LSTM model. First, one simulates the prediction with ARIMA model to get the error value and the prediction result. Then, one can use these data to simulate in the LSTM model to correct the prediction results of the ARIMA model and reduce the error. When they are properly combined, the accuracy of their prediction results will be higher than the prediction results obtained from the individual models. Therefore, it can be expected that when these models are reintegrated and used appropriately, the stock market will be visualized in the future. One can accurately predict future stock prices and greatly reduce the risk of investment.

5. Conclusion

In summary, this study presents methods regarding the use of ARIMA model and LSTM model and the better prediction results when combined with other prediction models. In ARIMA model, it is more suitable for forecasting short time linear data and it can integrate these multivariate factors well. In LSTM, it is more suitable for long term forecasting as well as building larger neural networks and it performs well in nonlinear models. the ARIMA model is restricted to short term forecasting and the need for linear data, but most of the time, these data are nonlinear, which also leads to the limitations of the ARIMA model. In LSTM model, it needs strong data support and takes longer time to make predictions. During the prediction period, it may lead to over-fitting and its data cannot be easily modified. Therefore, these forecasting models can be mixed to complement each other's strengths and amplify their advantages, making the accuracy of forecasting improve and thus greatly reducing the risk. Overall, these results offer a guideline for price prediction for different underlying assets.

References

- [1] Demirgüç-Kunt A, Ross L. Stock Markets, Corporate Finance, and Economic Growth: An Overview. OUP Academic, Oxford University Press, 1 May 1996,
- [2] Niblock S J, Heng P, Sloan K. Regional stock markets and the economic development of Southeast Asia. Asian-Pacific Economic Literature, 2014, 28(1): 47-59.
- [3] Chen A S, Leung M T, Daouk H. Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. Computers & Operations Research, 2003, 30(6): 901-923.
- [4] Choi H K. Stock price correlation coefficient prediction with ARIMA-LSTM hybrid model. arXiv preprint arXiv:1808.01560, 2018.
- [5] Zeng Z, Khushi M. Wavelet denoising and attention-based RNN-ARIMA model to predict forex price. 2020 International joint conference on neural networks (IJCNN). IEEE, 2020: 1-7.

- [6] Kim H Y, Won C H. Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. *Expert Systems with Applications*, 2018, 103: 25-37.
- [7] García M V R, Krzemień A, del Campo M Á M, et al. Rare earth elements price forecasting by means of transgenic time series developed with ARIMA models. *Resources Policy*, 2018, 59: 95-102.
- [8] Ho S L, Xie M, Goh T N. A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. *Computers & Industrial Engineering*, 2002, 42(2-4): 371-375.
- [9] Livieris I E, Pintelas E, Pintelas P. A CNN–LSTM model for gold price time-series forecasting. *Neural computing and applications*, 2020, 32(23): 17351-17360.
- [10] Binhuraib T. An Introduction to Time Series Analysis with Arima. *Medium, Towards Data Science*, 16 Mar. 2021.
- [11] Arima Modeling. CORP-MIDS1 (MDS), 4 Aug. 2022, Retrieved from: <https://www.mastersindatascience.org/learning/statistics-data-science/what-is-arima-modeling/>.
- [12] Yang X. The prediction of gold price using ARIMA model. 2nd International Conference on Social Science, Public Health and Education (SSPHE 2018). Atlantis Press, 2019: 273-276.
- [13] Jarrett J E, Kyper E. ARIMA modeling with intervention to forecast and analyze Chinese stock prices. *International Journal of Engineering Business Management*, 2011, 3(3): 53-58.
- [14] Jiang Q, Tang C, Chen C, et al. Stock price forecast based on LSTM neural network. *International Conference on Management Science and Engineering Management*. Springer, Cham, 2018: 393-408.
- [15] Pawar K, Jalem R S, Tiwari V. Stock market price prediction using LSTM RNN. *Emerging trends in expert applications and security*. Springer, Singapore, 2019: 493-503.
- [16] Mehtab S, Sen J. Analysis and forecasting of financial time series using CNN and LSTM-based deep learning models. *Advances in Distributed Computing and Machine Learning*. Springer, Singapore, 2022: 405-423.