Prediction Model of Gold Yield Time Series Based on ARIMA and LSTM

Duo Wang *

Department of Political Economy, King's College London, London, United Kingdom

*Corresponding author: k20060315@kcl.ac.uk

Abstract. To improve the accuracy of stock forecast returns, this paper will select ARIMA and LSTM forecasting models from traditional forecasting models and nonlinear forecasting models, respectively, to establish a more suitable forecasting method for gold returns. As the research object, this paper will choose the gold return from January 1, 2000 to April 30, 2020, and take May 2020 as the forecast time range. First, this paper will introduce the background of gold yield. Next, this paper will explain the definition of the ARIMA model as well as the LSTM model and the research methodology. Then, this paper will compare the models set up for forecasting gold yields under different parameters. Finally, this paper chooses the model with the most suitable parameters for the normalized values and show by the results that both models are not suitable for analyzing the gold yield. The research of this paper has important theoretical and practical significance for the prediction in the financial field.

Keywords: Gold returns; ARIMA; LSTM; Financial.

1. Introduction

In the paper by Barro and Misra on gold returns the average real price change of gold in the US from 1836 to 2011 was calculated to be 1.1%, but since gold does not act as a hedge against macroeconomic recessions, its expected real return should be close to the risk-free rate of about 1% [1]. This is supported by Malliaris as well as Malliaris and they add that gold has responded very positively to fears of inflation, stock market correction, currency crises and financial instability in a relatively short period of time [2]. In conclusion, to precisely estimate yields and fluctuations, this article chose to use gold yields as a study. Due to its large range of trading services, quick money settlement, usability, two-way trading, favorable trends, and excellent value preservation, among many other advantages, it is a well-liked option for market participants when it comes to risk management and hedging. But gold has always increased the danger of investing. The model may be impacted by a wide range of factors since the causes of these changes are so complicated.

Real yields, for instance, have a direct impact on gold prices, whereas nominal interest rates and inflation have a secondary impact on real yields, which in turn have a secondary impact on gold prices. Keep in mind that the price of gold can fluctuate wildly, which can result in big gains or, on the other hand, in sizable investment and monetary losses. Gold has a substantial impact on gold mining firms as well since high-cost mining operations might become unprofitable if yields drop significantly in the future. Therefore, to minimize potential risks and aid them in making wise investment decisions, financial investors and central banks must build reliable and accurate forecasting models that depict and reveal the patterns in gold yields and the mechanisms underlying their evolution.

There are two types of forecasting models for gold yields: standard forecasting models and nonlinear forecasting models. On the one hand, research on forecasting gold yields has made great progress due to ARIMA models. For example, Sharma A. M. and Baby. S. used the ARIMA model to forecast and study the gold price in India [3]. In numerous studies, traditional forecasting methods have been shown to be successful in predicting the trend or dynamic characteristics of gold prices. However, due to the nonlinear and extremely noisy data characteristics of the gold futures price time series, it is still very difficult to predict accurately.

On the other hand, with the rapid improvement of computer hardware quality and the development of neural network research, nonlinear models began to be gradually used in the prediction problem of gold production time series. In recent years, many scholars have started to use LSTM models to simulate stock market forecasts, and some financial series forecasts using LSTM models have achieved promising forecasting results. However, LSTM models do have some drawbacks. When attempting to forecast time series, LSTM models rely heavily on step-by-step forecasting, which is a key drawback. As a result, long-term forecasting using LSTM is unlikely to be successful.

In the time series, this paper will investigate gold returns by comparing ARIMA models as well as LSTM models. On the one hand, the ARIMA model has made great progress in the study of gold return forecasting. However, the accuracy of the model's forecasts is still tricky. On the other hand, the LSTM model has had promising forecasting results for financial series forecasting. However, LSTM also has drawbacks, especially in long-term forecasting. Therefore, in this paper, ARIMA and LSTM forecasting models are selected from traditional forecasting models and nonlinear forecasting models, respectively, to construct a more suitable gold futures price forecasting method. As the research object, this paper selects the gold return from 2000 to 2020 and forecasts the gold return in May 2020.

2. Method

2.1 ARIMA Model

The ARIMA model, which combines the difference operation and the ARMA model, is an extension of the autoregressive moving average model. Using trend difference and seasonal difference combined with the fitting of the ARMA model, a series having strong seasonal effects (periodic effects) is turned into an astable one, which is modelled using the seasonal ARIMA model. The ARIMA (p, d, q)model is structured as follows:

$$\Phi(B)\nabla^d \widehat{Y_t}^1 = \Theta(B)\varepsilon_t \tag{1}$$

where t stands for time; The order of differences utilised to derive trend information is called d; B works the reverse shift,

$$(B)\widehat{Y_t}^1 = \widehat{Y_{t-1}}^1 \tag{2}$$

while p, d, and q stand for the autoregressive order, difference order, and moving average order, respectively. And ε_t is a white noise sequence.

Three steps are involved in ARIMA modelling and prediction: model identification, parameter estimation, and model diagnostic in addition to model prediction. It is first determined whether the sequence is stationary using the unit root test. In order to create a stable time series from nonstationary sequences, seasonal and trend influences are removed using logarithmic transformation and difference. The best ARIMA model is the one with the biggest log-likelihood function and the least Bayesian information criterion (BIC). The residual sequence should be white noise and have an autocorrelation coefficient (ACF) and partial autocorrelation coefficient (PACF) that are not substantially different from zero when the model is diagnosed.

2.2 LSTM Model

The long short-term memory network (LSTM) is a modified form of the conventional recurrent neural network (RNN). Hochreiter and Schmidhuber made the suggestion in 1997 [4]. The problem of gradient disappearance and gradient explosion during the training of long sequences is resolved by the addition of a memory cell in the LSTM as compared to the standard RNN. This enhancement makes it more effective over longer sequences.

The key of LSTM is the cell state. Three control gates—input gate, forget gate, and output gate—are positioned in a memory cell to safeguard and regulate the state of the cell. The figure below depicts the LSTM memory cell's structure. The memory cell's state is represented by a horizontal line that spans the diagram's top from the input on the left to the output on the right. The forgetting gate, input gate, and output gate of a neural network are all represented by layers, and their corresponding outputs are represented by and. The input and output of the memory cell are represented by the two levels in the picture, and the first layer creates a vector that is used to update the state of the cell.

According to the Figure 1, the LSTM is composed of three sections, each of which has a distinct role [5].

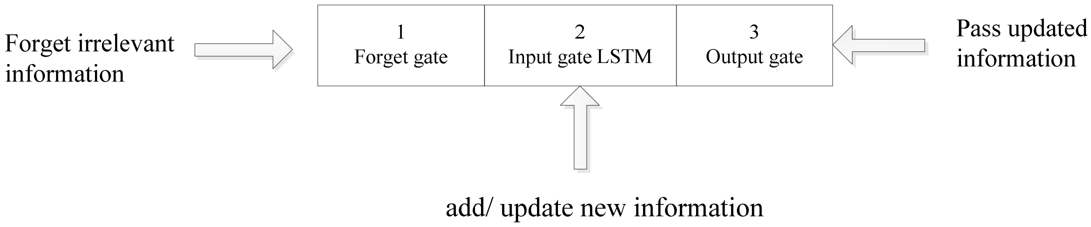

Fig. 1 The process of LSTM

The forgetting gate can initially choose which bits of the cell state are useful and which parts should be forgotten under the prior hidden state and the new input data in the first step of this process. The input gate is the second phase, and it can decide, depending on the prior concealed states and the fresh input data, which new information should be added to the network's long-term memory. It's important to note that this input is the same as the forgetting gate's input. A tanh-activated neural network that contains the new memory information tells us how much of each part of the network's long-term memory should be updated using the fresh information. Here, the tanh is applied because its value lies in [-1,1], so it can be negative. This also means that the possibility of negative values here is necessary if the study process wishes to reduce the influence of a factor in the cellular state. However, the forgetting gate does not check whether the new input data is worth remembering, so the input gate acts as a filter to identify which components are suitable for retention. Next, since the network's long-term memory has been updated, the role of the input gate as a final step is to decide on the new hidden state. The same as what the forgetting gate does. The input is the same and the activation is also sigmoid. it takes the cell state by tanh to get the numerical state between -1 and 1 and filters the cell state with the output of the sigmoid and the multiplication [6].

In summary, this paper uses the ARIMA model and LSTM model to forecast the time series of gold futures, respectively. However, these two models have their own advantages and disadvantages. On the one hand, the disadvantage of the ARIMA model is that it requires the time series data to be stable, or stable after differencing, and basically can only capture linear relationships, but not nonlinear relationships. Therefore, given that the gold futures price time series has nonlinear and highly noisy data characteristics, ARIMA is still difficult to predict accurately. LSTM models, on the other hand, excel at handling volatile time series because of their innate capacity to respond fast to abrupt changes in trends. For instance, an abrupt drop in the price of gold would not be easily corrected by an ARIMA model. LSTM models do, however, have some drawbacks. The LSTM's heavy reliance on stepwise forecasting is one of its key drawbacks for forecasting time series. As a result, using the LSTM for long-range forecasting may be useless. The magnitude of the data is another problem. The LSTM is a type of neural network, and like other neural networks, it needs a lot of data to be trained well. By analyzing the strengths and weaknesses of the models, the ARIMA model tends to predict more accurate results when there is a clear trend in the data, while the LSTM model tends to perform better in volatile time series with more stationary components, but neither model is suitable for forecasting gold returns.

To measure the effectiveness of the model in predicting gold yields. Mean absolute error(MAE)， Mean Absolute Scaled Error(MASE) and Root Mean Square Error (RMSE) are selected as the forecast error indicators, where the smaller the value of these three performance indicators, the closer the forecast value is to the actual value, and the smaller the forecast error is.

$$MSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)}\ ^2 \tag{3}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{4}$$

Here y_t is the actual observed value, $\hat{y}_t$ is the predicted value, and n is the number of prediction periods. The model with the lowest value obtained according to the above formula should be used as the most suitable model.

3. Data

3.1 The Source of Data

The closing gold returns in USD used in this study are information collected by Yahoo Finance [7]. It is reasonable to forecast gold price returns for only one day, a relatively short period, to ensure that the ARIMA and LSTM models' forecasts are valid, rather than next year or even longer, as gold prices and their returns are strongly influenced by the long-term global economic and political environment. Therefore, using the two models mentioned above, this study forecasts the gold price returns on May 1, 2020 and evaluates its forecasts in light of the actual situation.

3.2 Constructing ARIMA Model

3.2.1 Transforming the time series

In order to turn the gold price time series into the time series of the gold return rate, it is required to do a series of transformations for the data. In brief, after guaranteeing that the data is numeric (if it is not numeric, use as.numeric(unlist()) to turn it into a numeric form), a combined function diff(log()) is utilized to turn the gold price into the log return, which is similar to the rate of return. One thing about this process should be noted: it would not work if the function becomes log (diff ()).

3.2.2 Verifying the sequence stationarity

First, most time series would be stationary after the diff(log()) transformation. However, in this example, the authors go on to use some functions to prove the stationarity of the process to make the whole procedure more rigorous. The time series after the transformation is first plotted above. There is no trend, periodicity, or seasonality in the plot, indicating that the time series is stationary. In addition, since the ndiffs() function returns 0, the time series no longer needs to be computed differentially. Finally, this article strongly reject the null hypothesis in the Augmented Dickey-Fuller test that the time series is not stationary, as shown in Figure 2 and Table 1, since the p-value is 0.01. The resulting conclusion is that the series is perfectly stationary.

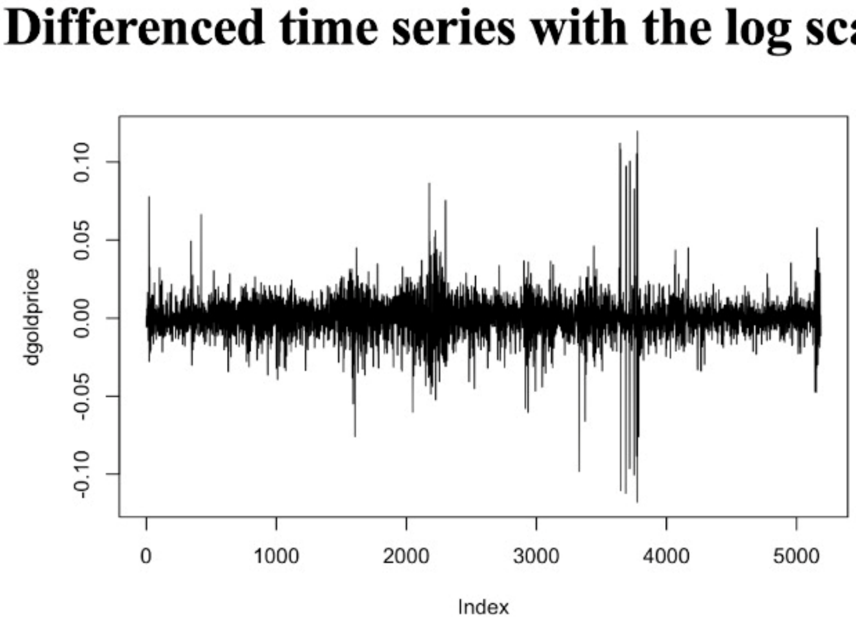

Fig.2 Differenced time series with the log scale

Table 1. Augmented Dickey-Fuller Test

Dickey-Fuller	-17.809
Lag order	17
p- value	0.01
Alternative hypothesis: stationary	

3.2.3 Identifying reasonable models

Next, this paper needs to come to the identification of reasonable models. auto.arima() function can be used to select parameters p and q in order to find an acceptable model. auto.arima() function shows that the best model in this case is ARIMA(2,0,2) with a non-zero mean (because the time series has been differentiated once, for the gold price, d value of 1), with the smallest value of AIC (Akaike Information Criterion, a measure of the degree of statistical model fit) compared to the other models. Moreover, according to the ACF and PACF plots, the p and q parameters are also non-zero, since their trajectories are biased towards zero. Therefore, it is reasonable to choose the ARIMA model provided by auto.arima(). These results are shown in Figure 3-4.

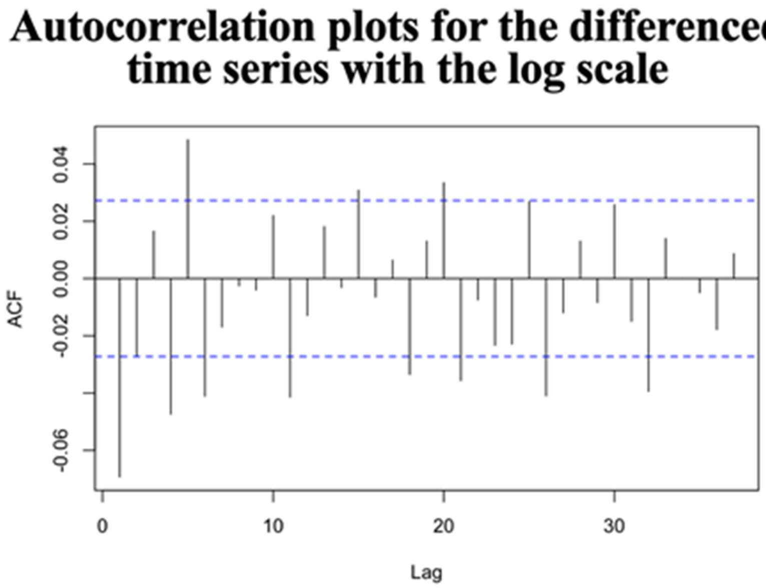

Fig.3 Autocorrelation plots for the differenced time series with the log scale

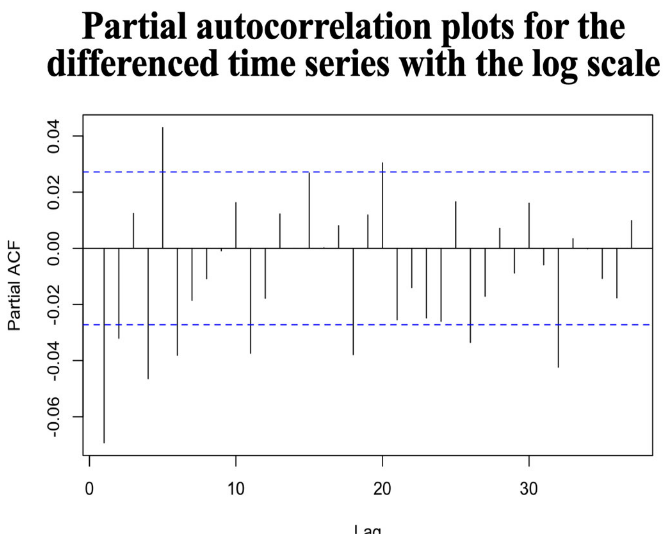

Fig.4 Partial autocorrelation plots for the differenced time series with the log scale

Next, this article needs to come to the identification of reasonable models. auto.arima() function can be used to select parameters p and q in order to find an acceptable model. auto.arima() function shows that the best model in this case is ARIMA(2,0,2) with a non-zero mean (because the time series has been differentiated once, for the gold price, d value of 1), with the smallest value of AIC (Akaike Information Criterion, a measure of the degree of statistical model fit) compared to the other models. Moreover, according to the ACF and PACF plots, the p and q parameters are also non-zero, since their trajectories are biased towards zero. Therefore, it is reasonable to choose the ARIMA model provided by auto.arima().The results are shown in Figure 5.

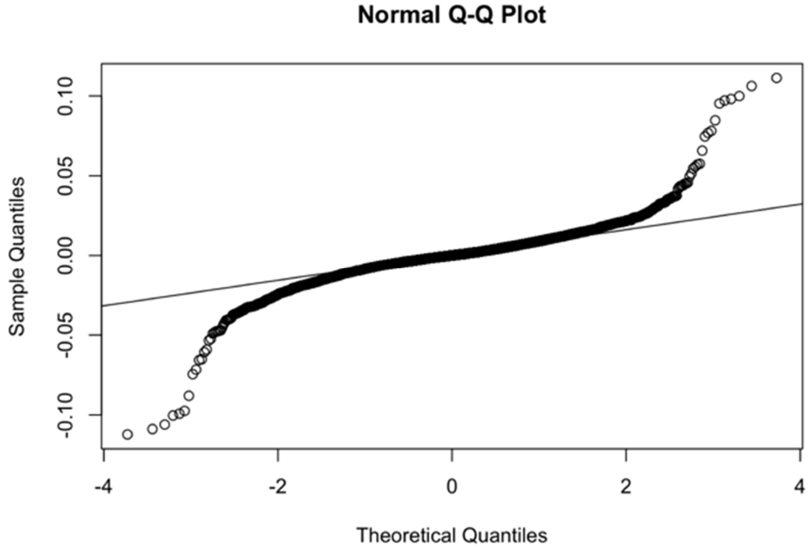

Fig.5 Normal Q-Q plot

3.2.4 Evaluating the model fit

After selecting a model, it is important to evaluate the fitness of the model. The normal Q-Q plot shown above was created in this example by applying the qqnorm() and qqline() functions. It shows that the residuals appear to have a coarse tail distribution, which indicates that extreme values are more likely to occur. Therefore, the point prediction should be accurate because the extreme values below and above the line are symmetric, and the confidence interval expected by the auto.arima() function - assuming the residuals have a normal distribution - would be wrong.

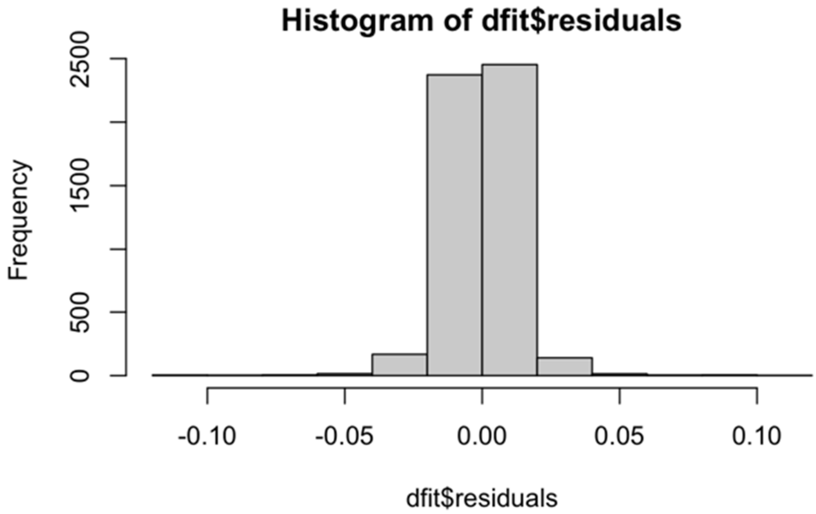

Fig.6 Histogram

In addition, the figure 6 shows that the residuals appear to have peaks that are substantially higher than the normal distribution. Therefore, the kurtosis () function was used to evaluate the kurtosis of the residual distribution to verify the previous statement. The output of this function is 15.35884, which indicates that it has fatter tails than the usual distribution and that it has considerably higher and sharper peaks. Therefore, in the next process, this paper must recreate confidence intervals that match the residual distribution.

Table 2. Box-L jung test

data: dfit$residuals	Box-L jung test
X-squared	2.0047
df	1
p-value	0.1568

Returning to the topic, the Box-Ljung test can be used to prove that all residual autocorrelations are zero. In this example, the result is not significant, with a p-value greater than 0.05, indicating that the autocorrelation is equal to zero. Because of these extremes, the ARIMA model looks good for data that are used to make point predictions, but it needs to be changed when determining confidence intervals.

3.2.5 Making Forecasts and analyses

Finally, this paper forecasts the gold return on May 1, 2020. The forecast() function can be used to obtain the estimated return for that date, which is 7.251787e-06. In addition, by using the autoplot() function, the results can be displayed as shown Fiugre 7 and table 3.

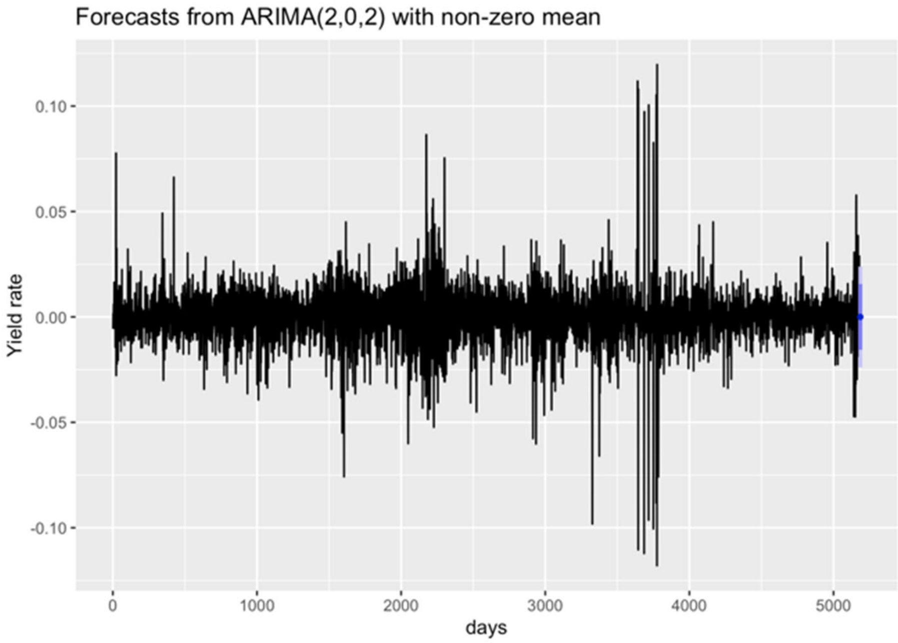

Fig.7 Forecasts from ARIMA with non-zero mean

Table3. Forecast in arima model

	Forecast
Point forecast	7.251787e-06
Low 80	-0.01559752
High 80	0.01561202
Low 95	-0.02385819
High 95	0.02387269

3.2.6 Limitation in Arima model

As the article mentioned above, since the residuals do not normally distribute, the predicted confidence interval below is invalid. Therefore, it is necessary to reconstruct the 90% confidence interval (the point forecast separately adds lower and upper 95% percentiles of the residuals). First, it is indispensable to locate the lower and upper 95% percentiles of the residuals. In detail, since the total amount of residuals is known, which is 5185, simple multiplication can be used to know which number of the sequence (a sequence arranging the residuals from small to large) is the lower and upper 95% percentiles, i.e., 259, 4927.

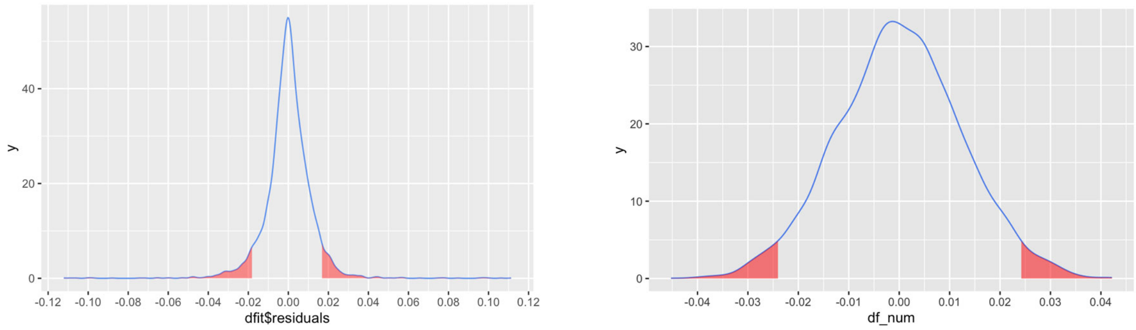

Fig.7 Histograms of the normal and fat-tailed distributions of the residuals, (left,right)

Therefore, the lower and upper 95% percentiles are separately the 259th number and the 4927th number of the sequence. Then the sort() function is used to arrange the residuals with their numerical value from large to small as shown below. In such a series, the authors locate the 259th number and the 4927th number with ease, that is -0.018062325 and 0.01652954, the value of lower and upper 95% quantiles of the residuals. Finally, since the point forecast (7.251787e-06) and certain percentiles are known, the confidence interval is (0.0165368, -0.01805507), which is narrower than the confidence interval predicted by the auto.arima() function. For explaining this phenomenon, the histogram of the normal and fat-tail distributions of the residuals are plotted and compared. The left graph above is the fat-tail distribution of the residuals, which is much more concentrated in the middle but more dispersive in the periphery compared with the normal distribution on the right side. To be specific, the frequency of 0 (almost the mean value which is the center of the whole graph) in the left graph is about 55, which is much greater than the frequency of 0 in the right graph which is about 33. What's more, between x=-0.25 to x=0.25, the left graph is much steeper than the right graph, which also indicates that the fat-tail distribution is much more concentrated in the middle. In other aspects, focusing on the periphery of the two graphs, the left graph has a greater range (from x=-0.11 to x=0.11) than the range in the right graph (from x=-0.05 to x=0.05). Hence, no doubt that the fat-tail distribution of the residuals is much more dispersive in the periphery than the normal one. In conclusion, since most of the residuals, in this case, are concentrated around the middle, the gap between the lower and upper 95% percentiles including those residuals would be narrow as the graph shows: the area between two red shadows has a high and sharp mountain which contains 90% percent of the residuals in a short range. Also, since the gap between two red shadows is the same as the difference between the upper and lower confidence interval, the range of the real confidence interval is much less than the one constructed by the normal distribution. The analysis above is the solution of resetting the confidence interval when the residuals are not normally distributed. After resetting, the prediction of the ARIMA model of the gold return is 7.251787e-06 with a 90% confidence interval (0.0165368, -0.01805507), which is more accurate than the result predicted with the model established by assuming that the residuals are normally distributed.

3.3 Constructing an LSTM Model

3.3.1 Gold yield volatility from 2000 to 2020

According to the above table for gold yields in 2000 as well as 2022, the yields of gold fluctuate a lot from 2000 to 2020. In fact, they do not show an increasing or decreasing trend in a consistent manner. On the contrary, they show increases and decreases. Thus, it allows us to guess the fluctuations of gold on the table 4.

Table 4. Change in gold returns from 2000 to 2022

Data	Open	High	Low	Close	Volume	Currency
2000-01-04	281.00	281.00	281.00	282.70	4	USD
2000-01-05	283.20	283.20	283.20	281.10	16	USD
2000-01-06	281.40	281.40	281.40	281.40	0	USD
2000-01-07	281.90	281.90	281.90	281.90	0	USD
2000-01-10	281.70	281.70	281.70	281.70	0	USD
…	…	…	…	…	…	…
2000-01-04	281.00	281.00	281.00	282.70	4	USD
2000-01-05	283.20	283.20	283.20	281.10	16	USD
2000-01-06	281.40	281.40	281.40	281.40	0	USD
2000-01-07	281.90	281.90	281.90	281.90	0	USD
2000-01-10	281.70	281.70	281.70	281.70	0	USD

3.3.2 Building the model and results

In the next step, this paper will use a three-layer LSTM model to be built with the purpose of predicting the gold yield. First, the actual collection yield is calculated using the function, and 10 epochs are trained by the data of gold yield. However, from the final model fitting test, the value of RMSE Rsquare MAE shows that the model does not fit well, and its prediction data is only 0.00019669 (Figure 8).

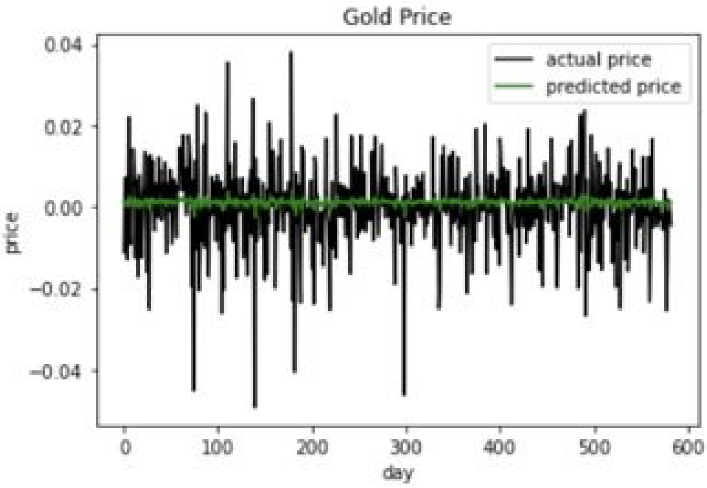

Fig.8 Gold price -Actual and Predicted

Table 5. LSTM model's layers

Layer (type)	Output Shape	Param
lstm	(None, 1, 50)	10400
dropout	(None, 1, 50)	0
Lstm_1	(None, 1, 50)	20200
Dropout_1	(None, 1, 50)	0
Lstm_2	(None, 50)	20200
Dropout_2	(None, 50)	0
dense	(None, 1)	51
Total params: 50,851		
Trainable params: 50,851		
Non-trainable params: 0		

Thus, the LSTM results also demonstrate a weak form of the efficient market hypothesis, which means that information about future returns cannot be obtained from past returns. Because the LSTM is a neural network that takes into account the dependence of the observations in the time series, ARIMA tends to predict more accurate results when there is a clear trend in the data, while the LSTM tends to perform better in volatile time series with more stationary components (Table 5).

Therefore, they are usually used for forecasting purposes. However, the main reason why LSTMs can handle volatile time series well is that they have an inherent ability to quickly adjust to sharp changes in trend. For example, an ARIMA model would not be able to quickly adapt to a sudden drop in the gold price (Table 6).

Table 6. Forecast results in LSTM model

Prediction	0.00019669
Print (mean_squared_error (actual_prices, predicted_prices))	0.00010343726205172879
Print (np.sqrt (mean_squared_error (actual_prices, predicted_prices)))	0.010170411105345192
Print (r2_score (actual_prices, predicted_prices)	-0.024415993400718428

3.3.3 Limitation

According to the flowchart in, the LSTM model, now considered as a rescuer, evolves in a similar way to the ResNet model to prevent cells and recall longer time steps [6]. This is because it avoids endless concatenated multiplication and instead multiplies as it adds, similar to ResNet itself. Thus, LSTM can solve various gradient vanishing problems.

However, this does not completely solve the problem, as shown in the previous graphic. The journey from the previous unit to this current unit is still sequential. This path now has add-ons and ignores the branches associated with it, thus making it more difficult. There is no doubt that LSTMs and their derivatives are capable of remembering large amounts of long-term data However, they are only capable of recalling sequences of the order of 100, not 1000 or more. Another problem is that training them puts a lot of strain on the hardware. These networks take a lot of resources to train, even if we don't need to. Running these models in the cloud requires just as many resources as running them locally. LSTM training is difficult because it necessitates storage bandwidth restricted processing, which is a hardware designer's worst nightmare and ultimately restricts the availability of neural network solutions. To execute in each successive time step, LSTM requires four linear layers (MLP layers) per unit. To the point where many computational units cannot be used, the linear layers require a large amount of storage bandwidth. Typically, this is because the system does not have enough storage bandwidth to support these computational units. It is also easy to add more compute units, but it is more difficult to increase the storage bandwidth. Therefore, hardware acceleration is not suitable for LSTM and its variants LSTM and its variants can solve the gradient problem to some extent, but it is still not enough. Although it can handle sequences of the order of 100, it is still challenging for sequences of the order of 1000 or longer. In addition, the LSTM takes a while to compute. If the network is very deep, the time span of LSTM is long, which will lead to a larger computation.

4. Discussion

4.1 ARIMA Model's Result

There is no doubt that the 90% confidence interval of ARIMA (0.0165368, -0.01805507) includes the actual gold return; however, there is a huge difference between the point forecast (7.251787e-06) and the gold return itself (3.94e-03). The golden return may have a huge residual because it is quite unstable, that's why. Its confidence interval can thus be quite large in order to take into account early data with significant variation that do not fit the model. In addition, since gold returns have a non-linear trend, which the ARIMA model cannot take into account, there will often be high values of residuals. As a result, the confidence intervals would be very large and would not provide accurate predictions for this paper.

4.2 LSTM Model

The results of the LSTM model support a weak version of the efficient market hypothesis, which holds that previous returns cannot be used to predict future returns. Eugene Fama extended the EMH and proposed it in 1970. According to Fama, in the case of a securities market, the market is considered efficient if the price accurately reflects all available information [8-10]. The first is whether the price is free to change in response to relevant information, and the second is whether the relevant information about the security is sufficiently disclosed and disseminated so that every investor has access to the same quantity and quality of information at the same time. These two factors together make up the extrinsic efficiency of a securities market. This presumption suggests that investors will use all available information swiftly and effectively when buying and selling stocks. According to this idea, all known factors that influence stock prices are already represented in stock prices, making a technical analysis of stocks worthless.

5. Conclusion

The forecasting profile of the ARIMA and LSTM models is generally unfavorable for the gold market because the price of gold is free to change based on relevant information that can also be fully published and disseminated so that every investor receives the same amount and quality of information at the same time. In this case, our technical analysis is not valid, because the stock price already takes into account all the factors known to influence its price. This is where our study must be strengthened, because it shows not only that gold yields are unstable, but also that neither model is useful for gold research.

References

[1] Barro, R. and Misra, S., Gold Returns. The Economic Journal, 2016, 126(594).1293-1317.

[2] Malliaris, A., & Malliaris, M. What drives gold returns? A decision tree analysis. Finance Research Letters,2015, 13, 45-53.

[3] Xu L. P.,Luo M. C.. Short-term analysis of gold price forecast based on ARIMA model. Finance and Economics Science, 2011 (01), pp.26-34.

[4] Hochreiter, S., & Schmidhuber, J. Long short-term memory. Neural computation, 1997, 9(8), 1735–1780.

[5] Saxena, S. LSTM | Introduction to LSTM | Long Short Term Memor. 2021.

[6] Dolphin, R. LSTM Networks | A Detailed Explanation. 2020.

[7] Finance.yahoo.com. https://finance.yahoo.com/, last accessed 2022/10/02.

[8] Song H， Du D， Li Y . Research on Precise Demand Forecast of Retail Commodities Based on LSTM-ARIMA Joint Model. Journal of Physics: Conference Series, 2021, 1865(4):042049-.

[9] Ding Yan, Xu Dehe, Cao Lianhai, etc Study on the Suitability of LSTM and ARIMA Models for Drought Prediction Based on CEEMD -- Taking Xinjiang as an Example Arid Zone Research, 2022, 39 (3): 11

[10] Bali V, Kumar A, Gangwar S . A Novel Approach for Wind Speed Forecasting Using LSTM-ARIMA Deep Learning Models. International Journal of Agricultural and Environmental Information Systems (IJAEIS), 2020, 11.