Gold Price Prediction Model Based on LSTM Neural Network and ARIMA

Yue Huang *, Mengyao Yang, Lingna Wang

College Of Mathematical Science, Yangzhou University, Yangzhou, China, 225009

* Corresponding author: yuehuang2024@163.com

Abstract. This study proposes a composite model based on LSTM neural network and ARIMA for predicting future international gold prices and providing investors with the best timing strategy for gold from the present moment to the coming year. We successfully predicted monthly gold prices using a simple seasonal model. However, due to the non-stationarity of the daily gold price, we were unable to apply the time series model directly. For this reason, we introduced IQR and linear interpolation to smooth the data by seasonal differencing, providing reliable inputs to ARIMA and LSTM neural networks. Combining these two methods, we derive daily gold price forecasts. After that, we introduce a double SMA timing strategy, which utilizes a 10-day SMA and a 30-day SMA to smooth the curve through a simple moving average. The curves are used to provide appropriate investment solutions. This comprehensive approach not only improves the accuracy of the daily gold price, but also provides investors with more reliable decision support.

Keywords: Gold Price Prediction, LSTM, ARIMA, Calculation of Assigned Values, Double MA Timing Strategy.

1. Introduction

1.1. Background and Significance

In recent years, the international situation has been grim and complex, with incessant wars and trade frictions, superimposed on the impact of the global epidemic of new coronavirus pneumonia, against which maintaining the stability of the economic market and rationally avoiding financial risks are of great practical significance. The gold market has always been a hot topic for investors, especially in the current economic environment of high uncertainty. In recent years, the holding value of gold has gradually come to the fore, becoming an ideal choice for hedging and preserving value. At the same time, the development of the gold market also provides investors with more trading opportunities and profit space [1]. Therefore, an in-depth study of the gold price can make us better predict the gold price, which not only helps the country to adjust the economic policy in time, but also helps individuals and organizations to control the financial risk, which has both theoretical and applied value.

1.2. Domestic and international research results

Liu et al. [2] presents the concept of time series and summarizes the relevant issues in the current research field of time series forecasting. Then, the time series forecasting methods are introduced by classification. On this basis, they summarized several potential research directions and unsolved problems. Their investigation informs research in the field of time series forecasting. Xie et al. [3] takes the constituent stocks of CSI 300 index as the research object and optimizes the disadvantages of the double averages trading strategy on the basis of the research period from January 1, 2010 to June 30, 2020. After fully considering the trend-following characteristics of the double SMA strategy, it is believed that the strategy has better results when the market is in the stage of unilateral trend (unilateral up and unilateral down). Yan et al. [4] investigated the impacts of climate change on future flood control risk of Yangfanggou Reservoir in Yalong River basin through combining future climate scenarios, SWAT model, ASD technique, ARIMA model and BMA method. Some main statistical characteristics of future floods may be included in these projections and can be extracted by the ARIMA model. Xu et al. [5] proposed a deep learning neural network model based on LSTM

networks and particle swarm optimization (PSO) in this paper. The results show that the PSO-LSTM model outperforms the M-EIES, ANN, PSO-ANN, and LSTM at all stations in the watersheds. Fan et al. [6] established a novel hybrid model that considers the advantages of linearity and nonlinearity, as well as the impact of manual operations. To compare the performance of the hybrid models ARIMA-LSTM and ARIMA-LSTM-DP with the ARIMA, LSTM, and LSTM-DP models, production time series of three actual wells are analyzed. The results of the experiments indicate that the single ARIMA model has a good performance in the steady production decline curves. Katharina Sachse et al. [7] presented to investigate the perceived investment risk of lay investors. Two surveys were conducted to examine the financial risk perception of German individual investors. Mohamed Bilel Triki et al. [8] employ an MV-GARCH model and dynamic copula to investigate the return link and volatility spillover between these two markets. This indicates that gold is a good diversifier and safe haven, particularly during great tension.

1.3. Innovation

The innovation of this paper lies in two aspects. On the one hand, this paper establishes a composite model combined with LSTM and ARIMA, and uses IQR and linear interpolation to improve the accuracy of the prediction model, which has reliable use value in short-term prediction, On the other hand, this paper adopts a relatively new double moving average timing strategy, establishes simple moving averages to find the optimal solution of the dual-means strategy, and visualizes the data, which provides investors with reliable decision-making support and also allows the double moving average strategy to be presented to the public as much as possible.

2. Gold price prediction

This paper starts by forecasting the international gold price from May 2023-August 2023, and first needs to visualize the data to determine what kind of forecasting is needed. In this paper, we consider forecasting the price of gold on a monthly basis to get a general trend, followed by forecasting the price of gold on a daily basis. The data comes from the World Gold Council (https://www.gold.org/goldhub/data).

2.1. Simple Seasonal Modeling and Solving

In this paper, we first use EXCEL's pivot table to process the daily closing price of gold, calculate the monthly average closing spot gold price.

Firstly, the monthly average spot gold closing price from January 2021-April 2023 is visualized in Figure 1 below.

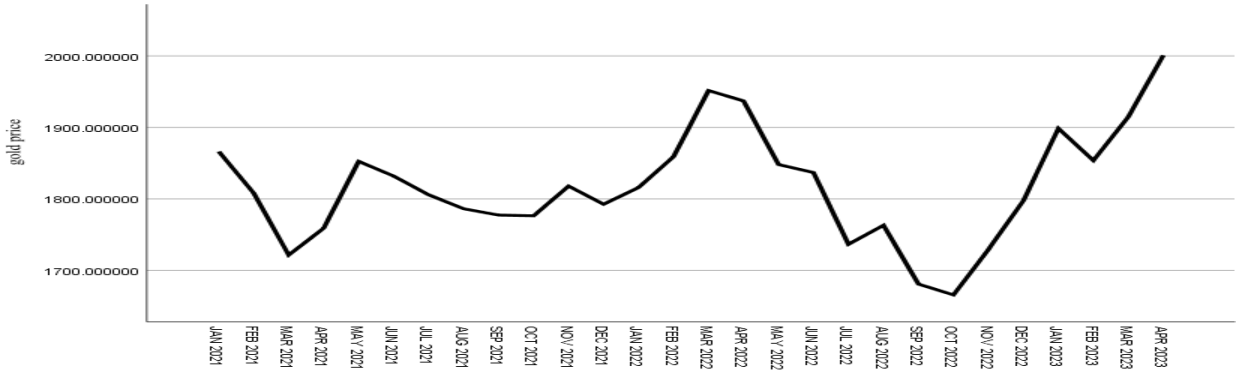

Figure 1. Monthly average spot gold closing price line chart

2.1.1. Data analysis

Since the monthly spot gold price is seasonal and has no obvious trend, a simple seasonal model is used to forecast the closing price of gold in the coming months. This paper uses the expert modeler for time series forecasting in SPSS software to solve the problem. The expert modeler for time series

forecasting in SPSS software is a tool specifically designed for predictive analysis of time series data. Expert Modeler is a trusted time series analysis tool that helps users identify and capture patterns and regularities behind complex time series data and perform forecasting and simulation.

2.1.2. Examination of residual white noise

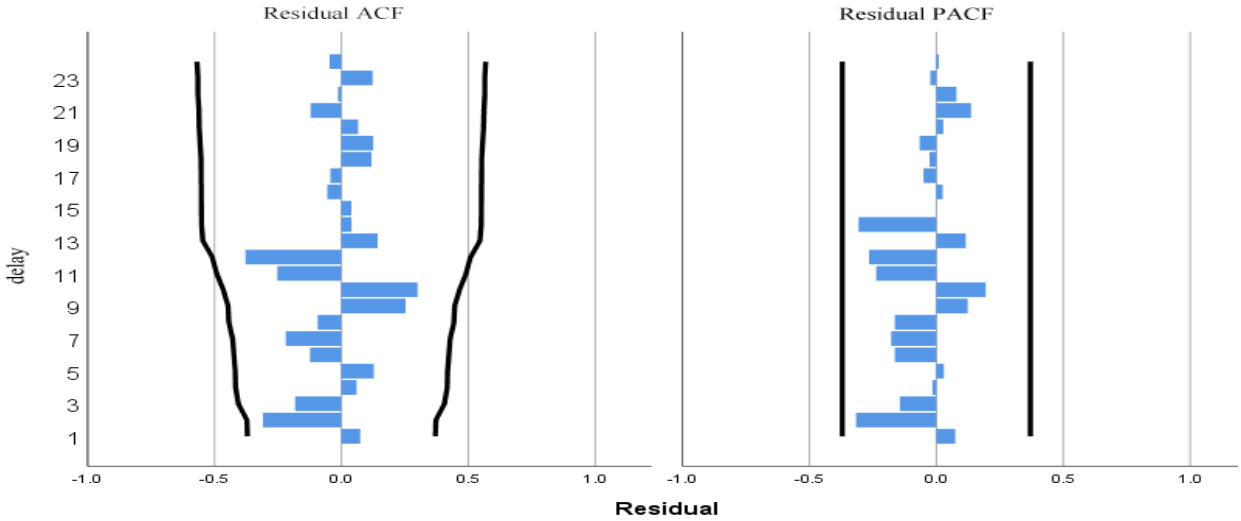

Figure 2. ACF and PACF plots of the residuals

Table 1. Simple Seasonal Model Statistical Tables

model	Number of predictive variables	Model-fit degree statistics	Young-Box Q (18)			outlier
		Smooth R-square	statistics	DF	p	
Simple seasonal model	0	0.803	28.513	16	0.027	0

From the Figure 2, it can be seen that the autocorrelation coefficients and partial autocorrelation coefficients of all lagged orders are not significantly different from 0. In addition, as can be seen in the Table 1, a Q-test of the residuals yields a significance p-value of 0.027, and thus we consider the residuals to be white noise series. Therefore, the simple seasonality model can well identify the monthly average spot gold closing price in this paper. Finally, we use the simple seasonality model to forecast the monthly average spot gold closing price from May 2023 to August 2023.

2.1.3 Modeling

The simple seasonal model is suitable for time series satisfying the following conditions:

The data should show a regular periodic change in the same month of each year or on the same day of each week, and there is no obvious trend in the time series data, so it is consistent here.

$$\begin{cases} l_t = \alpha(x_t - s_{t-m}) + (1-\alpha) l_{t-1} & (Horizontal\ Smoothing\ Equation) \\ s_t = \gamma(x_t - l_{t-1}) + (1-\gamma) s_{t-m} & (Seasonal\ Smoothing\ Equation) \\ \hat{x}_{t+h} = l_t + s_{t+h-m(k+1)}, k = \left[\frac{h-1}{m}\right] & (Prediction\ Equation) \end{cases} \quad (1)$$

m : Cycle length (take 12 for this article)

α : Horizontal smoothing parameters

γ : Smoothing parameters of the season

h : Number of forecast overrun periods

$\hat{x}_{t+h}$: Projections for the hth period

The parameters of the single seasonality model in SPSS software [9] were used to solve the solution, and the detailed results are shown in Table 2 below.

Table 2. Exponential Smoothing Method Model Parameters

Model		estimate	standard error	t
Simple seasonal	Alpha	0.999001	0.197717	5.052676
	Delta	0.000550	97.749117	0.000006

The future results predicted by the SPSS software are visualized in Figure 3 below:

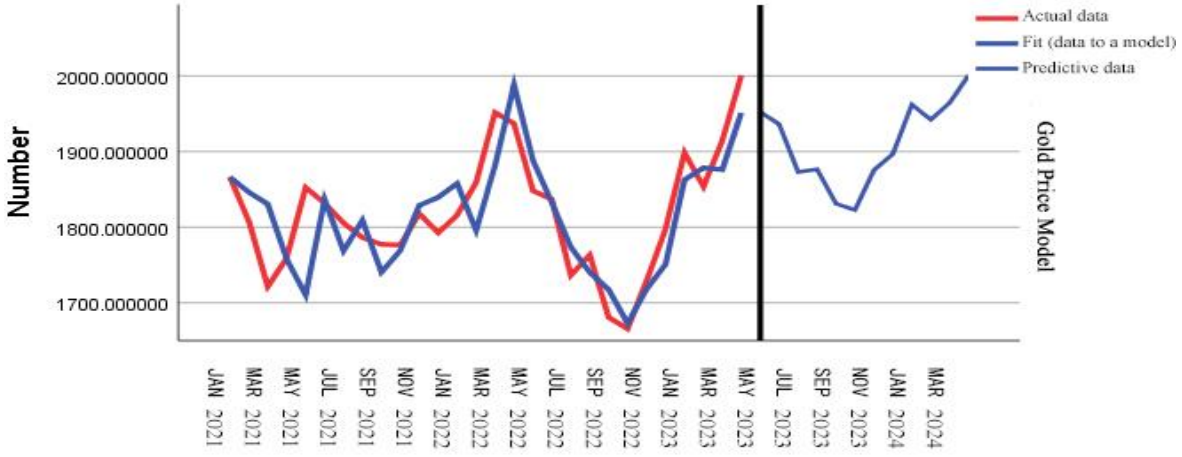

Figure 3. The SPSS prediction results

As can be seen from the figure, the difference between the real data and the fitted data is not big, which indicates that the simple seasonality model fits the original data quite well. In addition, the later months of the forecast retain the seasonal effect of the original series and are smoother, which indicates that the simple seasonality model can forecast the future monthly average spot gold closing price.

2.2. Establishment and solution of ARIMA and LSTM

2.2.1. Data analysis

The daily spot gold closing prices for January 2021-April 2023 are now visualized in Figure 4 below.

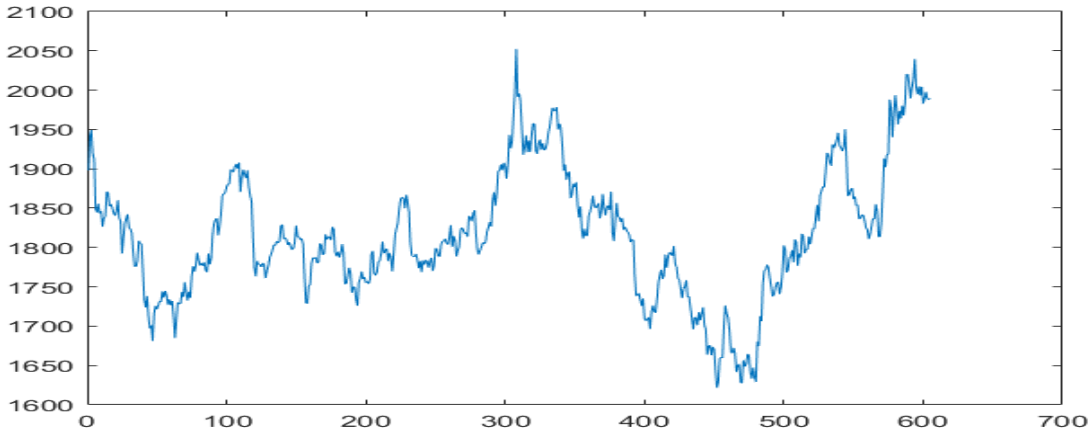

Figure 4. Daily spot gold prices

As can be seen from the time series plot above, the average spot gold closing price is not smooth and therefore not good for time series modeling.

2.2.2. Data preprocessing

We first process the raw data by filtering out outliers using IQR, and since the gold market opens Monday through Friday, we use linear interpolation to fill in the missing values. Next, we perform logarithmic transformation, first order differencing and seasonal differencing to make the data

smoother. Finally, we perform a smoothness test to ensure that the transformed data has been smoothed and plot the time series after seasonal differencing as follows in Figure 5

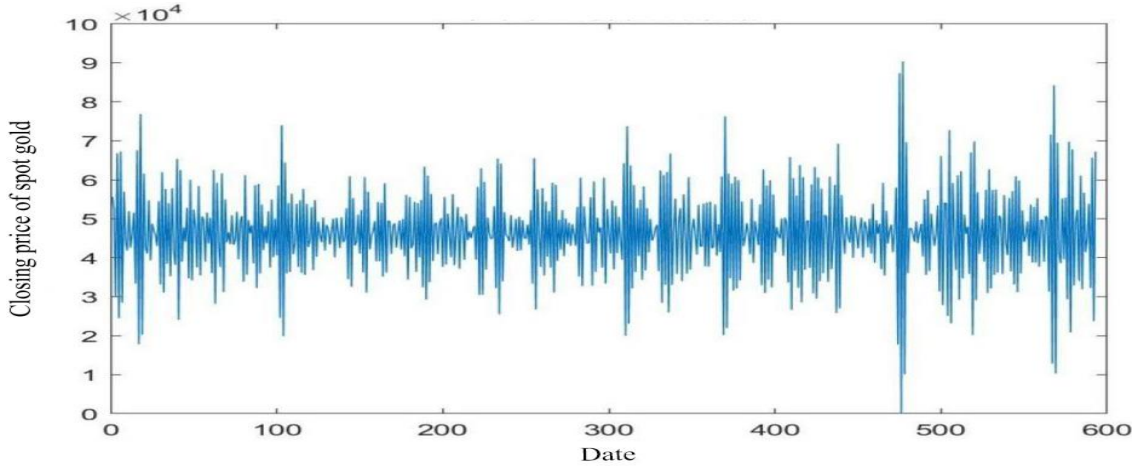

Figure 5. Seasonally differenced data

2.2.3. Solution of the ARIMA model

The processed data is smoother and more suitable for time series forecasting, so we imported the processed data into SPSS and used SPSS to test the residuals white noise. After observation, the model residuals are white noise series, which meets the requirements, and finally we use ARIMA (1,0,3) to forecast the price of gold in May 2023-August 2023 as shown in Figure 6 below

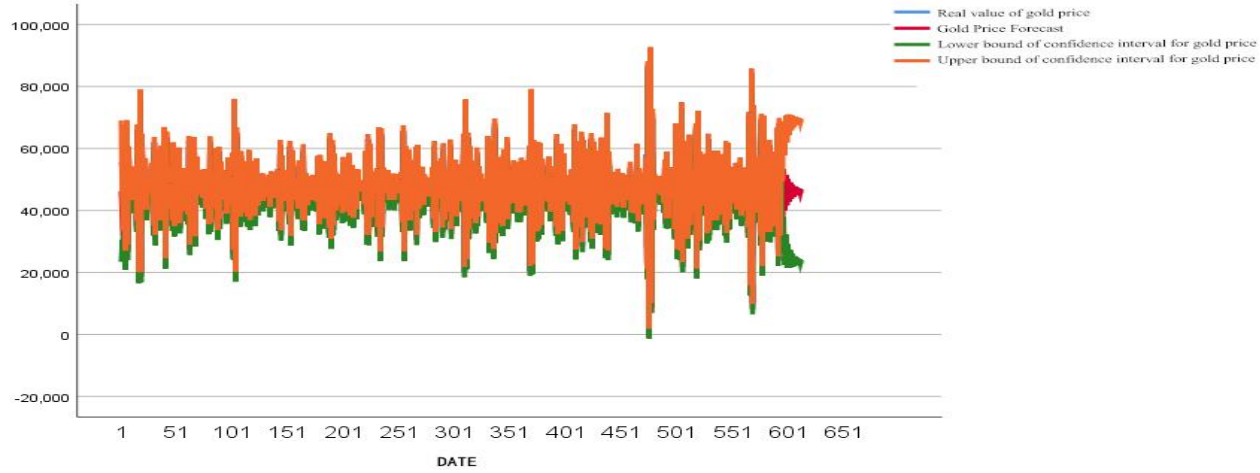

Figure 6. Comparison of predicted and real values

2.2.4. LSTM neural network prediction modeling

Since the error of using only ARIMA (1,0,3) is too large, we use a combination of ARIMA and LSTM neural network prediction to make the prediction, and assign the two to calculate the final prediction value. Table 3 and Table 6 show the information of LSTM related parameters, and the principle of LSTM will not be described in detail here.

Table 3. LSTM Grid parameters

parameter	describe	parameter	describe	parameter	describe
x_t	Input for the current moment	f_t	Oblivion gate control signal	C_{t-1}	The last moment memory unit
$\overset{\circ}{C}_t$	Information candidate status	i_t	Input gate control signal	C_t	Current moment memory unit
h_{t-1}	External state at the previous moment	o_t	Output gate control signal	h_t	Current moment external state

(1) Input gate i :

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\mathring{C}_t = \tanh(W_i \cdot [h_{t-1}, x_t] + b_c) \tag{3}$$

(2) Oblivion gate f :

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{4}$$

(3) Output gate o :

$$o_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t \times \tanh(C_t) \tag{6}$$

The LSTM neural network structure and hyperparameters are set as follows in Table 4.

Table 4. The LSTM neural network structure and hyperparameters

parameter	value
numFeatures	1
numResponses	1
numHiddenunite	150
MaxEpoch	300
GradientThreshold	1
InitialLearnRate	0.05
LearnRateDropperiod	120
LearnRateDropFactor	0.25

Using MATLAB to predict the closing price of spot gold for the days January 2021-April 2023, visualizing the predicted values gives the following Figure 7.

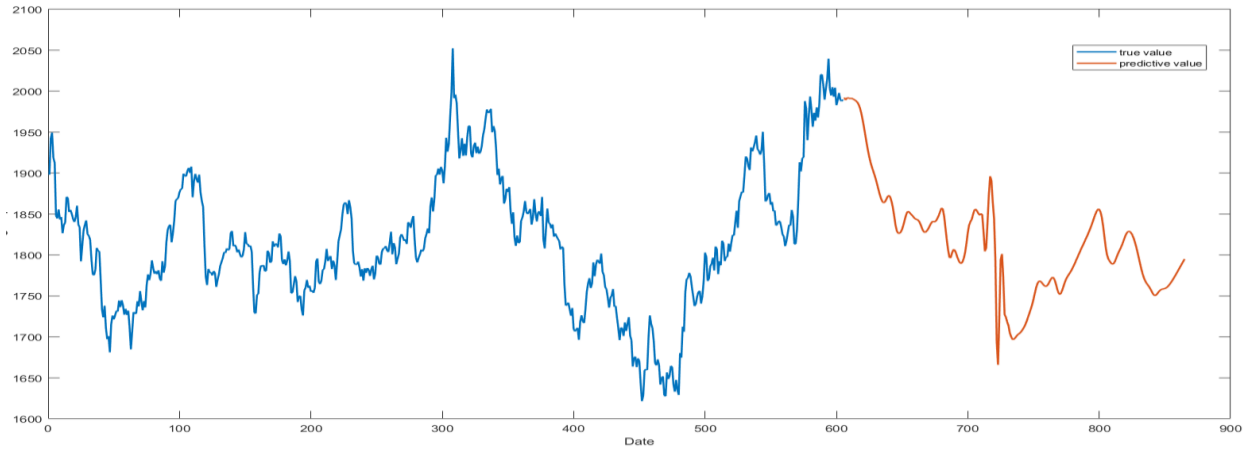

Figure 7. Spot gold closing price forecast chart

According to the figure, the prediction effect is still good, which can show the historical trend of the data. now we combine the two models and then use the error squared and derivative method to give the models different weights to make the predicted values more accurate.

The model is calculated as follows:

$$Combined\ Predicted\ Value = \frac{{\varepsilon_2}^2}{{\varepsilon_1}^2 + {\varepsilon_2}^2} \times ARIMA\ Projected\ Value + \frac{{\varepsilon_1}^2}{{\varepsilon_1}^2 + {\varepsilon_2}^2} \times LSTM\ Projected\ Value \tag{7}$$

${\varepsilon_1}^2$ is the sum of squared errors used by the ARIMA model predictions, ${\varepsilon_2}^2$ is the sum of squared errors for the LSTM neural network predictions. the ARIMA weights 0.31 and the LSTM neural

network weights 0.69. We calculated the two predicted values to obtain the final daily spot gold price, and the results are shown in Table 5 below.

Table 5. Daily spot gold price forecast value

date	ARIMA Forecast results	LSTM Forecast results	Final prediction results
2023-5-3	2,051.11	1988.4307	2007.861283
2023-5-4	2,051.11	1986.3107	2006.398483
2023-5-5	2,016.54	1983.3409	1993.632621
……	……	……	……
2023-8-29	1,816.82	1933.4895	1897.321955
2023-8-30	1,802.55	1942.8898	1899.384462
2023-8-31	1,780.70	1953.3367	1899.819323

3. Personal investment strategy model

Based on the results of the above forecast we can get the trend of international spot gold prices in the coming year. The main concern of most individual investors is to maximize their profits and minimize their risks. For this reason, we are going to use the Double SMA Timing Strategy to forecast the key moments in the future, give sound advice and state the reasons for it.

3.1. Double SMA Timing Strategy

The Double SMA Timing Strategy is a method of technical analysis based on moving averages that has been recognized for its effectiveness by market practice. The strategy uses two moving averages to determine buy and sell signals, usually a short-term and a long-term moving average. When a short-term moving average crosses a long-term moving average, it is called a "golden cross", which means that the price is trending upwards and is a buy signal; conversely, when a short-term moving average crosses a long-term moving average, it is called a "death cross", which means that the price is trending downwards and is a sell signal. This is a sell signal. By using the Double SMA timing strategy, you can avoid overreacting to sudden price fluctuations, and at the same time recognize trends and make adjustments when appropriate, thus allowing your capital to grow steadily over the long term!

3.2. Double SMA strategy correlation calculation

Considering that the price data of the international spot gold market is subject to large uncertainties and sharp fluctuations, the use of exponentially weighted moving averages may result in overly smooth curves, therefore, in this paper, we use the Simple Moving Average (SMA) to make a technical analysis of the international spot gold market in the coming year, so as to determine and capture buy and sell signals. The steps for the calculation of the SMA are as follows:

3.2.1. Determine the calculation time period

Determining the appropriate time period is very important, as either too long or too short a time period can have an impact on the results. In the field of financial analysis, one of the two combinations of 10 days-30 days or 20 days-60 days is generally taken as the short and long periods in the double average analysis. Considering that the forecasts in this paper are only for one year, which is a relatively short period of time, the 10-day moving average is used here for the slow line and the 30-day moving average for the fast line.

3.2.2. Calculate the double average

Since the two averages are calculated in an extremely similar manner, only the calculation of the 10-day moving average is presented here. Denote the spot gold price during 2023 as $X=(x_1,x_2,\cdots,x_{260})$.(Only the opening day is considered here, and Saturday and Sunday are not

counted) First the sum of the prices (S) over the time period is calculated, i.e. the closing price of the international spot gold price over the last 10 days is added together. This sum is then divided by the number of days in the selected time period to obtain the average price for the period, in this case divided by 10. As time passes and new trading information enters the market, the above two steps are repeated until the end of the observed time period. The formula is as follows

$$S_i = x_{i+1} + x_{i+2} + \cdots + x_{i+10}, (i = 0, 1, \cdots, 250) \tag{8}$$

$$MAs10_i = \frac{S_i}{10}, \tag{9}$$

This results in a moving average series $MAs10$

$$MAs10 = (1801.85, 1801.78, 1804.24, 1807.25, \cdots, 1799.36, 1799.73) \tag{10}$$

The 30-day moving average series is obtained by the same reasoning $MAs30$

$$MAs30 = (1816.09, 1817.78, 1819.53, 1819.45, \cdots, 1794.26, 1796.20) \tag{11}$$

3.2.3. Double SMA Visualization

Then analysis the two moving averages using MATLAB.

4. Development of specific investment strategies

Observe the intersection of the two averages in the chart, the 10-day moving average up through and down through the intersection of the 30-day moving average were expressed as recommended buy points and recommended sell points, and then take into account the risk factors, in the short-term averages up through, if the market trading price is still higher than the 10-day moving average, you can buy; in the short-term averages down through, if the market trading price is lower than the 10-day moving average, that is, it is recommended to sell operations Sell. After the judgment of the recommended buy and sell points in the chart marked out, as follows Figure 8, where the red points represent the recommended buy points, green points indicate the recommended sell points.

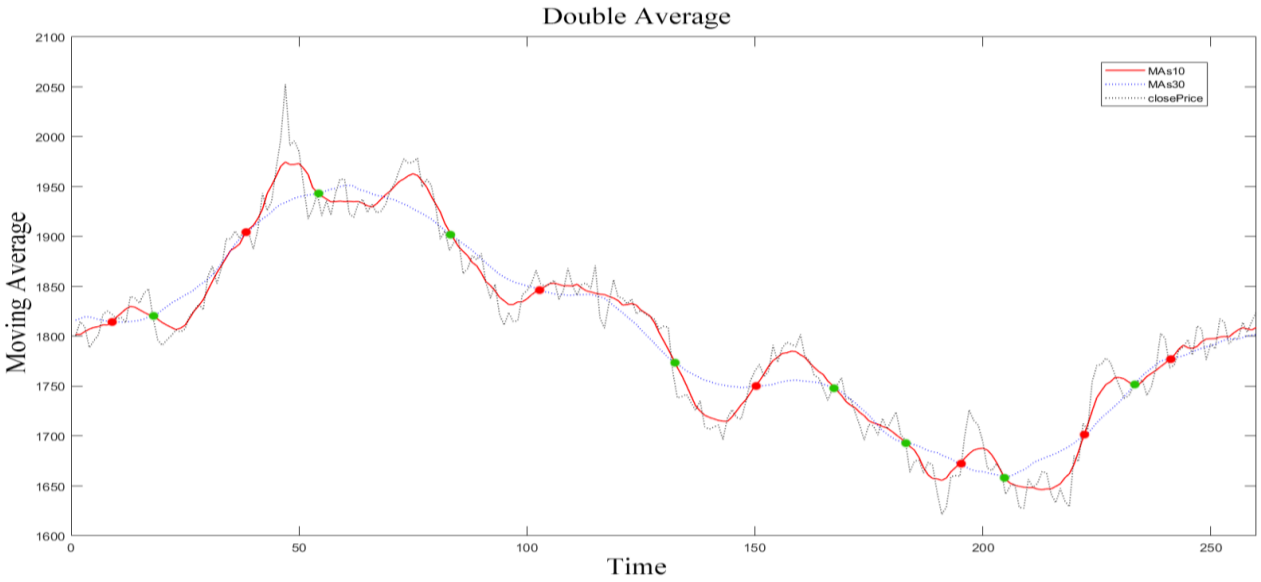

Figure 8. Double SMA Intersection

As can be seen from the chart, there are seven recommended buy points and eight recommended sell points, and the corresponding dates are included in the following Table 6.

Table 6. Recommended buy points and recommended sell points

date	buy/sell
2023/01/12	buy
2023/01/25	sell
2023/02/22	buy
2023/07/04	sell
2023/07/28	buy
2023/08/22	sell
2023/11/07	buy
2023/11/22	sell
2023/12/02	buy
...	....

Therefore, individual investors can make investment decisions in the coming year based on this table, taking into account the specific market situation. It is important to note that although this timing strategy is based on scientific advice given by the forecast results, it still requires investors to make timely decisions in light of specific circumstances in order to adapt to changes in the market. Although the recommendations in the table are specific to the date, the best date to buy or sell may not necessarily be at that point, but may also be at a date adjacent to it. Also, since the total amount of money in the hands of an individual investor varies, and the willingness to take risks varies, the decisions made by each individual will be different. If there is a large amount of money held in the hands, we recommend planning the funds [10]. For example, invest only a portion of the funds each time you buy and consider other investment directions at the same time, such as other stocks, futures, etc., in order to hedge the risk. It is recommended that before investing in the international spot gold market, do an in-depth investigation of the market situation and choose a professional investment platform when investing, monitor the market liquidity, and develop a clear stop-loss strategy in advance to control risk, avoid retracement, and protect their assets and profits.

5. Conclusion

In this study, we successfully constructed a reliable international gold price prediction model by combining the advantages of ARIMA and LSTM neural networks. By introducing IQR and linear interpolation in daily price forecasting, we effectively handle non-stationarity and improve the robustness of the model. The introduction of the double-mean timing strategy further strengthens the usefulness of the model and provides investors with clear timing signals. By comparing the difference between the model prediction results and the actual gold price, we validate the effectiveness of the model. While providing accurate forecasts, our study also provides investors with a comprehensive investment program, especially regarding when to buy and sell gold. Overall, this study provides a reliable reference for gold investment in the coming year, with practical implications for decision-making and risk management in the financial markets.

References

[1] Vieira D S, de Carvalho P V, Curto J D, et al. Gold's hedging and safe haven properties for European stock and bond markets [J]. Resources Policy, 2023, 85: 103817.

[2] Liu Z, Zhu Z, Gao J, et al. Forecast methods for time series data: a survey [J]. Ieee Access, 2021, 9: 91896 - 91912.

[3] Xie Junqi. Optimization and application of double averaging strategy based on CSI 300 index constituents[D]. Guangdong University of Foreign Studies, 2022.

[4] Yan B, Mu R, Guo J, et al. Flood risk analysis of reservoirs based on full-series ARIMA model under climate change [J]. Journal of Hydrology, 2022, 610: 127979.

[5] Xu Y, Hu C, Wu Q, et al. Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation [J]. Journal of hydrology, 2022, 608: 127553.

[6] Fan D, Sun H, Yao J, et al. Well production forecasting based on ARIMA-LSTM model considering manual operations[J]. Energy, 2021, 220: 119708.

[7] Sachse K, Jungermann H, Belting J M. Investment risk–The perspective of individual investors [J]. Journal of Economic Psychology, 2012, 33 (3): 437 - 447.

[8] Triki M B, Maatoug A B. The GOLD market as a safe haven against the stock market uncertainty: Evidence from geopolitical risk[J]. Resources Policy, 2021, 70: 101872.

[9] Rustamov S. Usage Method of SPSS Software in the Processing of Psychological Data [J]. ЦЕНТР НАУЧНЫХ ПУБЛИКАЦИЙ (buxdu. uz), 2021, 7 (7).

[10] Li Li. Risk aversion demand heating gold or meet the low do more window [N]. China Securities Journal, 2022 - 05 - 13 (A05).