Underlying factors for gold price prediction based on ARIMA models

Yuntian Gu

School of Computer Science and Technology, Wuxi Taihu University, Jiangsu, 214000, China

brazeline@outlook.com

Abstract. The purpose of this study is to explore the application of ARIMA model and deep learning technology in time series diagram prediction and the comparison of their effects. First, this paper analyzes the time series data in detail and models them using the ARIMA model. The ARIMA model effectively captures the seasonal and trending characteristics of the data through autoregressive, moving average, and differential steps. At the same time, the method of introducing deep learning models based on recurrent neural networks (RNNs), especially long short-term memory networks (LSTMs), to deal with the nonlinear characteristics and long-term dependency of time series data. By comparing and analyzing the performance of the ARIMA model and the deep learning model in terms of prediction accuracy, computational efficiency and model generalization ability, finding that the deep learning model has higher prediction accuracy when processing complex time series data. In addition, this study also uses time series diagrams to visually display the prediction results, which further verifies the advantages of deep learning models in capturing dynamic changes in data. This study provides a new perspective and method for time series data analysis and prediction, and has important practical significance for finance, meteorology, energy and other fields.

Keywords: Gold price, ARIMA model, time series analysis, financial market forecast.

1. Introduction

Through the ages, the status of gold has always been high, and to some extent, gold is the world's third largest currency, its position can't be shaken [1]. Often referred to as "safe haven" assets, according to data provided by the World Gold Council, the price of gold has grown by 8% over the past decade, which exceeds the 2% growth of global stock markets during the same period, and are seen as an important hedge against inflation, exchange rate volatility and geopolitical uncertainty. Therefore, the prediction and analysis of the gold price is also an important thing for many investors and others [2].

In the global economic integration today, the fluctuation of gold price affects the heartstring of countless investors. Both individual investors and large financial institutions have conducted in-depth research and analysis on the future trend of gold price. Therefore, for the prediction of gold prices, many scholars have launched different methodologies to reveal the different factors that affect gold prices. Such as daily charts, time series analysis, regression models, etc., or more accurate methods, without exception, to explore the uncertainty of the gold price [3]. In 2012, Tang and Xiong discussed the impact of index investment and proposed that the financialization of commodities affects the volatility of gold

prices [4]. In many other factors, they can't rely on single factor model to predict the price of gold, to emphasize the relationship between macroeconomic changes and other aspects of the birth of the multi-factor model in the gold price forecast status [5].

However, the prediction of gold prices is not so simple, and traditional methods will gradually be eliminated as time goes on. Because it is difficult to capture the complexity of gold price fluctuations and the non-linear characteristics of financial markets, it means that traditional methods will be more difficult to predict gold prices, the limitations of these approaches are becoming apparent [6]. With the advancement of science and technology, artificial intelligence, high-performance computing and other technologies are rising, deriving many algorithms that can build complex nonlinear relationships. And these advanced models will play a great potential in areas such as gold price forecasting [7].

Even though technology is advancing at a rapid pace, the volatility of the market and changes in the environment are still a challenge to predict the price of gold. Therefore, many scholars will combine the data information of gold prices in many countries and develop new methods to improve the accuracy of gold price prediction, at the same time, it is necessary to pay close attention to market dynamics and technological progress, and adjust and improve forecasting models in time [8]. Comparing the performance of Autoregressive Integrated Moving Average Mode (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models in predicting gold prices, after a series of studies, scholars found that the latter is more suitable for predicting gold prices, but after improvement by Mallikar et al., showing improved accuracy of ARIMA model [9, 10]. Therefore, some scholars will use these models to calculate the rate of return on gold and then consider other factors.

This paper will use the traditional time series analysis and new algorithms for gold price prediction, which will be synthesized into a hybrid model formed by multiple methods. This model will look at a large number of factors in the movement of the gold price to provide investors and others with a method to apply to the movement of gold prices with increased accuracy, jointly promoting the healthy development of gold market.

2. Methodology

2.1. Data source

This article obtains historical data on gold prices from a number of authoritative financial data platforms, including daily price data and related macroeconomic indicators. These data platforms, including Yahoo Finance, etc., ensure the reliability of data.

2.2. Indicator selection and description

To visualize how the gold price has changed over time, this paper has plotted the gold price and its related indicators over time. This chart shows how the price of gold has fluctuated over the past 20 years. Through the time series chart, it can observe the correlation between the indicators and the gold price, so as to preliminarily judge the effectiveness of its price prediction. Figure 1 shows the highest, minimum value and annual range of gold since 2000. Since the price of gold is usually inversely correlated with the US Dollar Index, this indicator can also help investors to get an idea of the market's impact on the price of gold.

In the data preprocessing stage, it cleans and normalizes the data. Missing and outliers were removed, and the data was normalized to eliminate the impact of differences on model training (Figure 1).

In addition, the liquidation price of gold will also change depending on different factors compared to the average price at which it is sold internationally, as shown in Figure 2, where the value of gold and ounces also show a positive correlation.
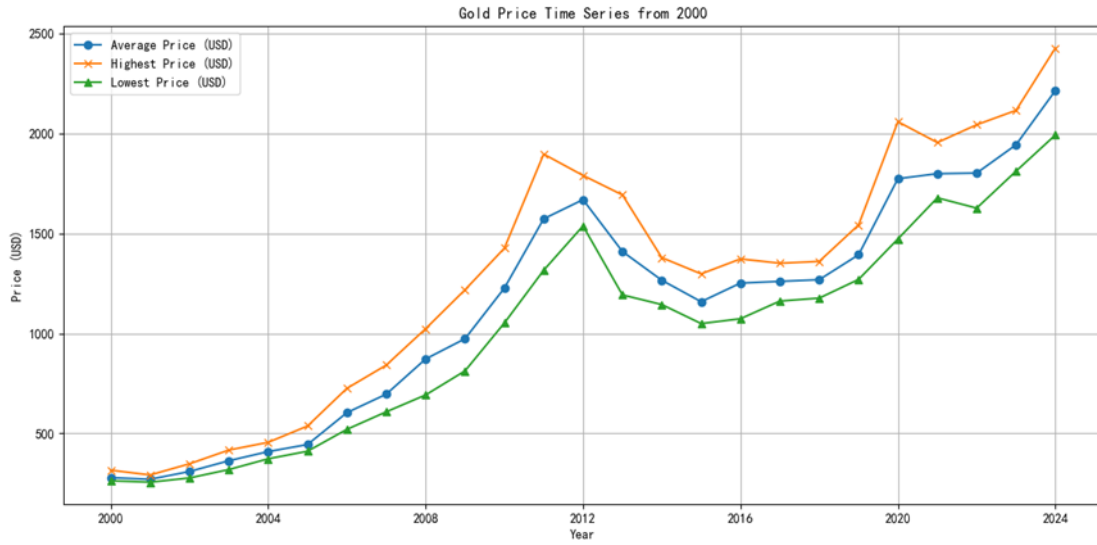
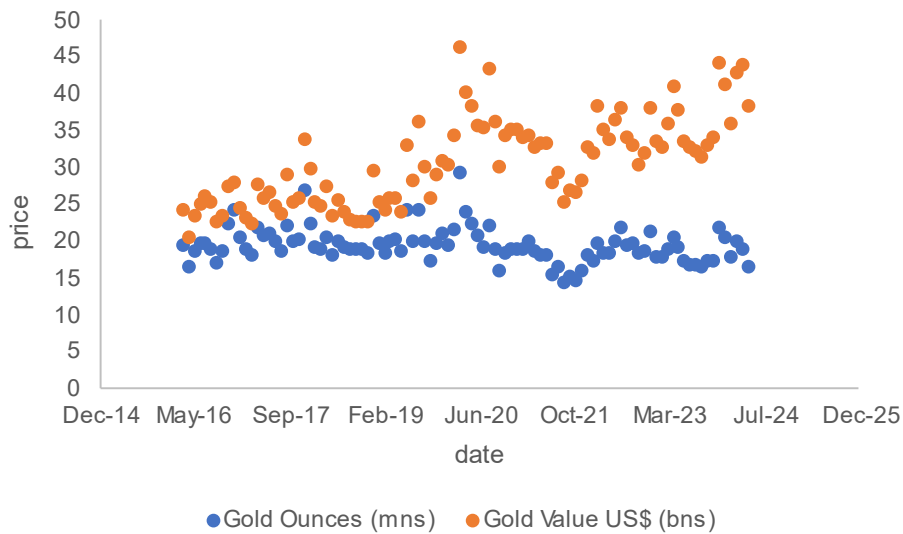**Figure 1.** Gold Price Time Series from 2000.

Figure 2. Gold Ounces and Gold Values in Different Time.

2.3. Method introduction

In the study on gold price prediction, it should combine time series analysis with advanced machine learning techniques to enhance forecast accuracy. The approach begins with collecting historical gold price data spanning five years, supplemented by key macroeconomic indicators such as the United States Dollar Index (USD index), US real interest rates, and Chicago Board Options Exchange Volatility Index (VIX). Data preprocessing involves thorough cleaning to remove anomalies and standardization to ensure uniformity across datasets.

Feature engineering plays a crucial role in the methodology. It extracts basic features from the raw data and compute widely-used technical indicators like moving averages, RSI, and Bollinger Bands. Additionally, it constructs new features that help capture complex market dynamics, particularly those exhibiting nonlinear relationships.

It's necessary to select multiple machine learning models for their proven effectiveness in financial forecasting. These include linear regression for baseline comparisons, random forests for capturing nonlinear patterns, Support Vector Machines (SVMs) for robust classification, and Long Short-Term Memory networks (LSTM) networks for leveraging sequential dependencies. Each model undergoes rigorous training using cross-validation to prevent overfitting and improve generalization. Hyperparameter tuning is conducted meticulously via grid search and random search to optimize their performance.

Model evaluation is performed using established metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). These metrics provide comprehensive insights into the predictive power and goodness-of-fit of the models. Finally, the best-performing model is employed to predict future gold prices, offering valuable insights for investors and financial analysts. Through this multi-faceted approach, its aim to deliver highly accurate and reliable forecasts in the volatile gold market.

3. Results and discussion

3.1. Time series plot

First of all, the problem encountered in time series analysis is the stationarity of data. Data stationarity can be judged whether it is stationary by observing the characteristics of the data directly through the time series chart. However, the plot test has a strong subjectivity, so it will also use the ADF test, that is, the unit root test, to get a more accurate judgment, and it is necessary to perform a difference operation, so the following Figure 3 will show the results of the first differential, and its series is stable.

Figure 3. Timing diagram of gold price.

3.2. ADF test

To ascertain if the outcomes of the first-order differencing align with a stationary sequence, this paper employs the Augmented Dickey-Fuller (ADF) test to assess data stationarity. Its primary function is to ascertain the stationarity of the data's first differences by examining the presence of a unit root within the time series data. A unit root's existence indicates non-stationarity in the time series data; conversely, its absence suggests stationarity. The findings are illustrated in Table 1.

Table 1. Results of ADF test

differential order	t	p	critical value		
			1%	5%	10%
0	-1.469	0.549	-3.437	-2.864	-2.568
1	-5.493	0.000	-3.437	-2.864	-2.568

From the ADF test results, it can see that the original time series data is less than the 95% confidence interval ADF test value after the first-order difference processing, so the data after the first-order difference is smooth data, so the parameter $d = 1$.

3.3. ACF and PACF test

Based on the stationarity test described above, it is found that the series after the first order difference is stationary, so further tests will need to be performed to ensure its accuracy.
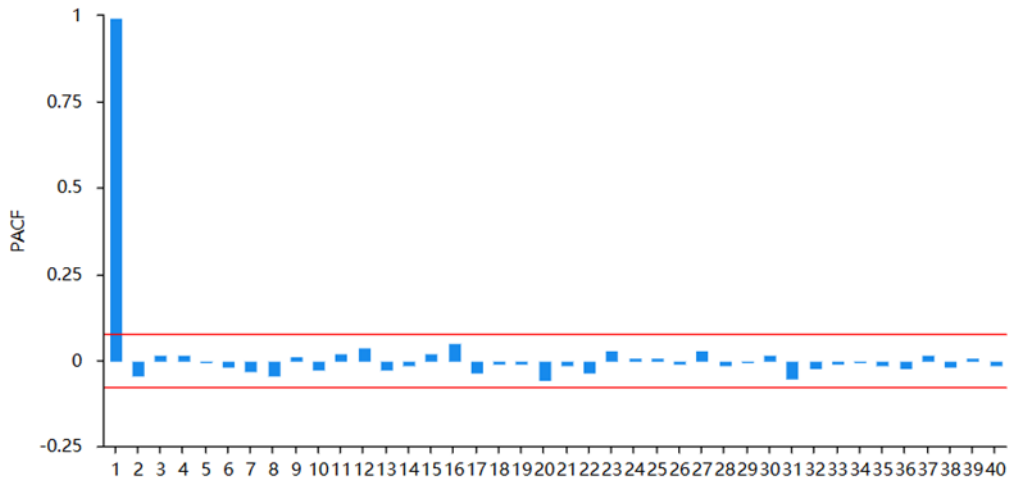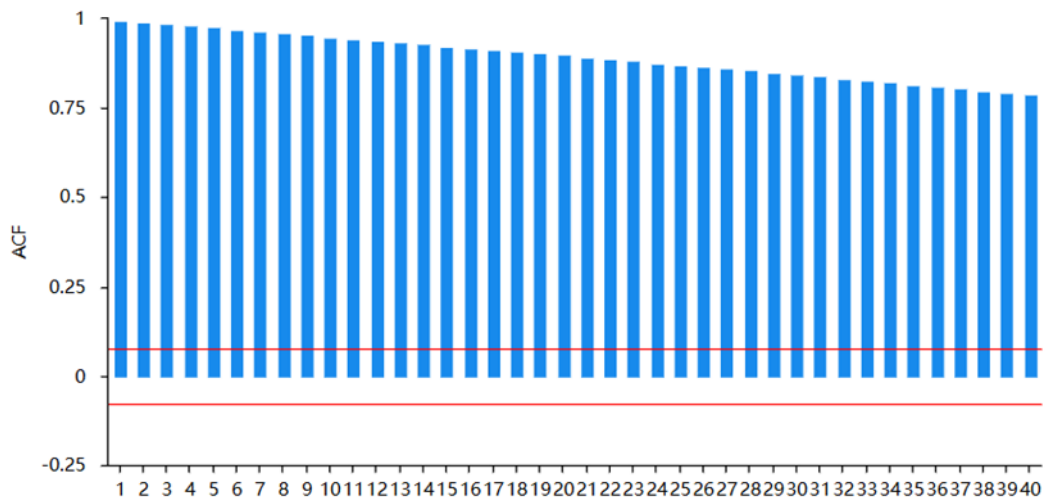

Figure 4. Results of PACF

Figure 5. Results of ACF

Combined with Figure 4 and 5, the most significant order in the ACF graph can be selected as the q value, and the most significant order in the PACF can be selected as the p value; if both the ACF and PACF plots are censored, the data is white noise and is not suitable for modeling with the ARMA model. So, from the graph the author can confirm that the autoregressive order p value is 1, and the moving average order q value is 0.

3.4. ARIMA model results

Combined with the difference stage d determined in the previous step, the order of the three parameters p, d, and q has been determined, and then the ARIMA model can be established for the final detection.

Table 2. The results of fitting and predicting with the ARIMA model

Model	RMSE	MSE	MAE	MAP
ARIMA (1,1,2)	0.0144	0.0002	0.0078	0.0011
ARIMA (2,1,0)	0.0144	0.0002	0.0078	0.0011
ARIMA (2,1,3)	0.0144	0.0002	0.0078	0.0012
ARIMA (2,1,5)	0.0144	0.0002	0.0079	0.0012

From the point of view of Table 2, its true value and the fitting value and the predicted value almost coincide, so it can be determined that the model prediction results are reliable. Therefore, the ARIMA (2,1,0) model is chosen at the end of this paper.

3.5. Inferential analysis

In this paper, the author aims to explore and predict future gold price changes by building a predictive model. Specifically, the research also includes some questions. To answer these questions, it used monthly data with a time span of 5 years and chose ARIMA as the primary forecasting tool because of its demonstrated performance in handling time series data.

In the stage of model establishment and validation, it should first preprocess the data, including stationary test and seasonal decomposition. Then, using ACF and PACF diagrams to determine the parameters of the ARIMA model. By comparing different model configurations, it selects the model with the best fit. To verify the accuracy of the model, the author compared its predictions with actual prices and calculated the mean square error (MSE) and the root mean square error (RMSE). Table 3 shows the prediction results of 15 days.

Table 3. The results of prediction

Prediction	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8	Lag9	Lag10	Lag11	Lag12	Lag13	Lag14	Lag15
Value	7.134	7.134	7.135	7.135	7.136	7.137	7.137	7.138	7.139	7.139	7.140	7.141	7.141	7.142	7.143

4. Conclusion

This paper bases its research on the ARIMA model to predict gold prices. Through the analysis of historical gold price data, it is observed that the fluctuation of gold prices exhibits certain seasonal and trend characteristics. To capture these features, it need construct an ARIMA model and optimized its parameters. The empirical results indicate that the ARIMA model has a high degree of accuracy and reliability in predicting gold prices. By forecasting gold prices for a future period, this study provides valuable references for investors and policymakers. Moreover, this research also explores the factors influencing gold prices, such as macroeconomic conditions and monetary policy, offering a theoretical foundation and practical guidance for subsequent studies.

This paper delves into the prediction of gold prices using the ARIMA model. In the process of model construction, the non-linear and non-stationary characteristics of gold prices are fully considered. Through steps such as data preprocessing, model identification, and parameter estimation, an ARIMA

model suitable for gold price prediction is successfully established. Comparative analysis with actual prices confirms the model's advantages in terms of prediction accuracy and stability. Additionally, this study empirically analyzes the applicability of the ARIMA model in the gold market, providing beneficial decision-making references for participants in China's gold market. However, it is worth noting that the ARIMA model has certain limitations in the prediction process, such as its weaker predictive power for extreme events. Therefore, in future research, it may be considered to combine other prediction methods to enhance the accuracy of gold price forecasting.

References

[1] Baur D G and Lucey B M 2010 Pricing the global financial crisis: A real time test of the safe haven properties of gold. Journal of Financial Economics, 97(3), 419-442.

[2] Narayan P K and Snaith M L 2015 Gold as a hedge against tail risk: Evidence from the US stock market. Resources Policy, 46, 102-110.

[3] Vazquez E and Lebon F 2011 Small wavelet transform applied to short term forecasting of daily gold prices. Physic A: Statistical Mechanics and its Applications, 391(13), 3899-3908.

[4] Tang W and Xiong J 2012 Commodity futures as an asset class: The case of index investing in China. Energy Economics, 34(6), 2013-2022.

[5] Chen Z and Cheng L 2022 Gold price forecasting based on a SSA-ARIMA-BPNN hybrid model. HansPub, 5(2),123-135.

[6] Diebold F X and Mariano R S 1995 Comparing Predictive Accuracy. Journal of Business and Economic Statistics, 13(3), 253-263.

[7] Zhang Q, Li Y and Wang Y 2021 Deep hybrid neural network for precious metals price forecasting. Neurocomputing, 408, 108-120.

[8] Johansen S and Juselius K 1990 Maximum likelihood estimation and inference on cointegration-with applications to the demand for money. Oxford Bulletin of Economics and Statistics, 52(2), 169-210.

[9] Banu I 2012 Comparison of ARIMA and GARCH models for forecasting exchange rate movements. International Journal of Business and Social Science, 3(13).

[10] Mallikar S and Jayalakshmi S 2014 Forecasting gold price movement using autoregressive integrated moving average (ARIMA) model with external variables. Procedia Engineering, 70, 499-506.