Proceedings of the 3rd International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/108/20242024

Forecasting Gold Futures Prices: An Empirical Analysis Using the Vector Autoregression Model

Zihan Wang[1,a,*]

[1]School of Management Science and Engineering Southwestern University of Finance and Economics Chengdu, Sichuan, 610074, China
a. 18180596802@163.com
**corresponding author*

Abstract: This study presents an empirical examination of the Vector Autoregression (VAR) model's efficacy in predicting gold futures prices, a critical analysis for investors and financial analysts amidst global economic fluctuations. The research leverages a dataset spanning from January 4, 2010, to June 7, 2024, derived from Investing.com, encompassing a multitude of financial indicators such as gold and silver futures, the US Dollar Index, and Brent Crude Oil futures.The methodology includes a comprehensive data preparation phase, ensuring the stationarity of time series data through Augmented Dickey-Fuller (ADF) testing and cointegration analysis. The VAR model is meticulously estimated, identifying the lag order that optimizes model performance while mitigating overfitting risks.The empirical findings underscore the VAR model's high accuracy in forecasting gold futures prices, with significant influences attributed to silver futures prices, the US Dollar Index, and Brent Crude Oil futures. Despite minor discrepancies in capturing short-term price volatility, the model adeptly reflects overall market trends.The study acknowledges limitations, such as the potential exclusion of influential variables and the scope of data timeframe, suggesting future research should expand variable inclusion and integrate advanced prediction techniques. The paper concludes that the VAR model is a robust analytical instrument for predicting gold futures prices, offering substantial support for strategic financial decision-making.

Keywords: Vector Autoregression, Gold Futures Prediction, Economic Indicators, Financial Modeling, Time Series Analysis.

1. Introduction

With the continuous development of the global economy, international financial markets have become the main platform for economic interdependence and interaction among countries. As an important component of the international financial market, the gold market has always been highly regarded. Gold, as an important safe haven asset and investment tool, has profound impacts on the global economy due to its price fluctuations and changes. Owing to the influence of the gold market can be extended to the gold futures market, gold futures have emerged as a star product in the financial futures market, occupying a substantial trading volume and proportion in numerous exchanges. They serve as the primary choice for many market participants to hedge and avoid risks [1]. Therefore, the price fluctuations of gold futures are an essential basis for investors to make judgments before making investment decisions. It is imperative to construct a reasonable and

DOI: 10.54254/2754-1169/108/20242024

effective futures prediction model to reveal the trend of gold futures prices and provide investors with a reliable reference for investment directions.

2. Literature Review

The Vector Autoregression (VAR) model is a commonly used econometric model for analyzing the dynamic relationships between multiple time series. Since its introduction by Sims [2], the VAR model has found widespread application in economics and finance. By considering the relationships between multiple variables simultaneously, the VAR model is able to better capture the complexities of economic and financial markets.

numerous studies have explored the factors influencing gold prices. For instance, Blose [3]investigated the relationship between gold and inflation and found a positive correlation between gold prices and inflation. Nikiforos[4]'s study on the relationship between gold and money supply, discovered that an increase in money supply leads to a rise in gold prices. Additionally, Ghosh [5]analyzed the relationship between gold and the stock market and found a negative correlation. Furthermore, Koosup[6] studied the relationship between gold and the oil market and found a positive correlation.

In terms of the application of VAR models, many studies have successfully applied VAR models to forecast financial markets. For example, Caporale[7] used a VAR model to analyze the dynamics of European stock markets and to predict their trends. Their findings indicated that the VAR model was able to capture the volatility and trends of the stock market quite well. Additionally, Wang [8]employed a VAR model to analyze the international financial markets during the Asian financial crisis and to forecast their movements. Their research results showed that the VAR model had high accuracy and reliability in forecasting international financial markets.

In conclusion, the literature review suggests that the VAR model is an effective tool for analyzing the key variables influencing gold futures prices and forecasting the closing prices of gold futures. By considering the dynamic relationships between multiple variables, the VAR model can provide more comprehensive and accurate market forecasts. Therefore, this study will be based on the VAR model to analyze and predict the trends in the gold futures market.

3. Data and Methods

3.1. Data Selection and Source

The data for this study were sourced from the Investing.com website (https://cn.investing.com/), covering the period from January 4, 2010, to June 7, 2024. To simplify the model, the closing prices of each variable were ultimately chosen as the values for model construction. The specific variables included:

Gold futures closing prices
Silver futures closing prices
US Dollar Index
Euro to US Dollar exchange rate
S&P 500 Index
Brent Crude Oil futures closing prices
EGO company stock prices
US 10-year Treasury bond yields

DOI: 10.54254/2754-1169/108/20242024

3.2. Methods

3.2.1. Data Processing Methods

In data processing, Missing values were first checked, and outlier detection was conducted using quartile methods. There are no missing values or outliers in the data.

3.2.2. Time Series Analysis Methods

In the process of establishing the VAR model, Augmented Dickey-Fuller (ADF) tests were conducted to ascertain the stationarity of the time series variables. For variables that exhibited non-stationarity, first differences of the data were computed to conform to the necessary requirements for constructing the VAR model. The selection of the lag order was based on the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Final Prediction Error (FPE), and Hannan-Quinn Information Criterion (HQIC). Causality matrices were constructed via Granger causality tests, and Johansen cointegration tests along with Durbin-Watson statistics tests were performed prior to the finalization of the VAR model.

4. Empirical Analysis

4.1. Descriptive Statistics of Data

In the preliminary phase of data exploration, a comprehensive quality assessment of the dataset was conducted, with a primary focus on the integrity and accuracy of the data. The implementation of a missing value test confirmed the absence of missing values within the dataset, ensuring the completeness of all records. Subsequently, outlier detection was performed on the data using the quartile method, specifically [Q1-1.5*(Q3-Q1), Q3+1.5*(Q3-Q1)]. Any values falling outside this range are considered outliers. Fortunately, a meticulous review revealed no outliers, indicating the dataset's high level of accuracy.

To gain a deeper understanding of the relationships among the variables within the dataset, various visualization techniques were employed. Initially, line charts were generated to depict the closing prices of gold futures in relation to the closing prices of Brent crude oil in London and silver futures over time. These visual representations effectively illustrate the trend variations and potential correlations between the distinct variables.

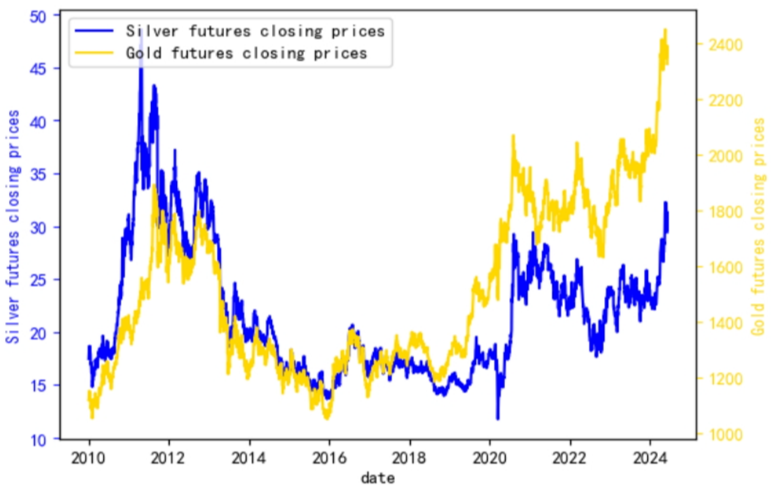

Figure 1: Line Chart of Gold and Silver Futures Closing Prices Over Time.

DOI: 10.54254/2754-1169/108/20242024

In summary, through preliminary exploration and visual analysis of the dataset, a preliminary understanding of the overall characteristics of the data and the relationships between variables was obtained.

4.2. Unit Root Test and Cointegration Analysis

In the initial stages of data processing and analysis, it is crucial to ensure that the data are stable and reliable. A common problem with time series data is that the data may contain unit roots, i.e., the data may not be smooth, which may lead to pseudo-regression problems when constructing economic models. Therefore, we first perform the ADF (Augmented Dickey-Fuller) test on the data to check the smoothness of the data.

The results of the Augmented Dickey-Fuller (ADF) test indicate that the variables of interest exhibit non-stationarity at their original levels. This implies that the statistical properties of these variables, including mean, variance, and autocovariance, change over time, which is incompatible with the data stationarity requirements of numerous economic models.

To address the issue of non-stationarity, the difference method was employed. First-order differences of the data were computed, and the ADF test was subsequently conducted on the differenced data. Fortunately, the results of the test indicated that all variables had been transformed into stationary time series after differencing. This step is crucial as it ensures that the data used in constructing the VAR (Vector Autoregression) model meets the stationarity requirement, thereby mitigating the risk of spurious regression.

Following the confirmation of data stationarity, the Johansen cointegration test was performed to investigate the presence of a long-run equilibrium relationship among the variables. The outcomes of the Johansen cointegration test indicate the existence of a long-run equilibrium relationship among the variables of interest. This implies that while these variables may exhibit short-term deviations from equilibrium, they will tend toward a stable equilibrium in the long term. This discovery serves as a significant foundation for the establishment of the VAR model.

4.3. Estimation Results of the VAR Model

In analyzing the dynamic relationships within financial markets, Vector Autoregression (VAR) models prove to be a valuable instrument. Once the stability and cointegration relationship of the data have been ascertained, the estimation phase of the VAR model commences, which is a pivotal step in the analysis.

4.3.1. Determination of Lag Order

To optimize the model’s performance, an information criterion method was employed, which included the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Final Prediction Error (FPE), and Hannan-Quinn Information Criterion (HQIC), to determine the lag order of the VAR model. Following a meticulous comparison and analysis, a lag order of 2 was selected for the model. This selection was based on striking a balance among the values of the information criteria, with the objective of identifying a equilibrium point that ensures the model’s fitting performance while mitigating the risk of overfitting.

DOI: 10.54254/2754-1169/108/20242024

	AIC	BIC	FPE	HQIC
0	-8.648	-8.634	0.0001755	-8.643
1	-8.783	-8.656*	0.0001534	-8.737*
2	-8.788*	-8.549	0.0001525*	-8.703
3	-8.772	-8.420	0.0001549	-8.647
4	[illegible]	[illegible]	[illegible]	[illegible]

Figure 2: Determination of minimum lag order (partial)

4.3.2. Parameter Estimation

After determining the lag order, we used the Ordinary Least Squares (OLS) method to estimate the parameters of the VAR model. This model contains 8 equations that collectively capture the dynamic relationships between different variables in the system. Our dataset has a total of 3507 observations, which provides rich information for model parameter estimation.

Table 1: OLS estimation results (Summary).

<table>
<tr><th colspan="4">Summary of Regression Results</th></tr>
<tr><td colspan="2">Model:</td><td colspan="2">VAR</td></tr>
<tr><td colspan="2">Method:</td><td colspan="2">OLS</td></tr>
<tr><td colspan="2">Date:</td><td colspan="2">Sat, 06, Jul, 2024</td></tr>
<tr><td colspan="2">Time:</td><td colspan="2">11:41:05</td></tr>
<tr><td>No. of Equations:</td><td>8.00000</td><td>BIC:</td><td>-8.54754</td></tr>
<tr><td>Nobs:</td><td>3507.00</td><td>HQIC:</td><td>-8.70124</td></tr>
<tr><td>Log likelihood:</td><td>-24266.6</td><td>FPE:</td><td>0.000152779</td></tr>
<tr><td>AIC:</td><td>-8.78652</td><td>Det(Omega_mle):</td><td>0.000146982</td></tr>
</table>

Taking the closing price model of gold futures as an example, its BIC, HQIC, and FPE values are -8.54754, -8.70124, and 0.000152779, respectively. These values are all at a low level, indicating that the model has a good fitting effect on the data and can accurately describe the dynamic changes in gold futures prices.

4.3.3. Evaluation of Model Stability and Predictive Ability

To ascertain the stability and predictive capability of the VAR model, Durbin-Watson (DW) statistical tests were performed on the model's residuals. DW statistics are utilized to assess the presence of first-order autocorrelation in sequences. In the model under consideration, the DW statistics for all variables are approximately 2.0, suggesting the absence of significant autocorrelation in the model residuals. This outcome further validates the stability of the VAR model and justifies the reliability of the predictions derived therefrom.

DOI: 10.54254/2754-1169/108/20242024

```
Gold futures closing prices : 2.0
Silver futures closing prices : 2.0
US Dollar Index : 2.0
Euro to US Dollar exchange rate : 2.0
S&P 500 Index : 2.0
Brent Crude Oil futures closing prices : 2.0
EGO company stock prices : 2.0
US 10-year Treasury bond yields : 2.0
```

Figure 3: Durbin-Watson (DW) statistical test results.

5. Prediction and Analysis

5.1. VAR Model Prediction Results

Following the inversion of differencing on the model data, the predicted data were successfully derived. On this basis, a graphical representation of the predicted and true values of the closing price of gold futures over time was constructed.

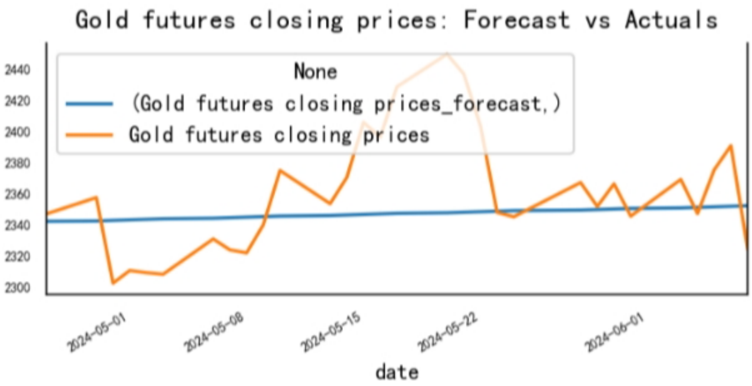

Figure 4: Line graph of projected and real values over time

From the predicted image of the closing price of gold futures drawn, it can be clearly observed that although there is a certain difference between the predicted values obtained by the VAR model and the true values, overall, the predicted values still remain within a relatively reasonable range. The image shows that the phased predicted values given by the model tend to show a steady growth straight line, while the true values show more severe fluctuations and significant trend changes, and their volatility is obvious. From an intuitive perspective, the difference between the two does exist and is quite obvious. However, if viewed from the overall trend of change, the trend of predicted values is consistent with the trend of actual values.

In summary, the model exhibits some discrepancies in capturing minor price fluctuations; however, it is effective in reflecting the overall trend of changes over a given time period.

5.1.1. The Impact of Various Variables on Gold Prices

Through an in-depth analysis of the parameter estimates of the VAR model, insights can be gained into the specific impacts of various variables on gold futures prices. The results indicate that the closing price of silver futures, the closing price of the dollar index, the closing price of London Brent crude oil futures, and the closing yield of the U.S. 10-year Treasury bonds exert a particularly significant influence on gold futures prices. Conversely, the impact of EGO's stock prices and the euro-to-dollar exchange rate is found to be relatively minor.

5.2. Model Prediction Performance Evaluation

To comprehensively and rigorously evaluate the performance of the VAR model in predicting the closing price of gold futures, multiple quantitative evaluation indicators were adopted for integrated analysis. Specifically, the mean absolute percentage error (MAPE) is 0.0123, indicating that the average error predicted by the model is relatively small, demonstrating high prediction accuracy. However, the mean error (ME) was -12.8635, revealing a slight negative bias in the model's prediction. The mean absolute error (MAE) is 29.2201, which reflects that the average absolute difference between the predicted and actual values is within a reasonable range. The negative percentage error (MPE) is -0.0052, further confirming the existence of slight negative bias in the model. The root mean square error (RMSE) is 38.9047, which suggests that in some cases, the prediction error may be large and the stability of the model needs to be improved. The minimum maximum error (minmax) is 0.0122, indicating that the range of prediction error is relatively compact and the model's prediction results are relatively stable.

Based on these data, we can conclude that the model performs well overall in predicting the closing price of gold futures, but there is a slight negative bias, and in some cases, the prediction error is large. The model still has room for optimization.

6. Conclusion

The paper demonstrates the effectiveness of the VAR model in forecasting the general trend of gold futures prices over a one-month period. The high accuracy of the model is attributed to its proficient ability to capture the intricate relationships among variables and to analyze market dynamics. The study identifies key factors such as silver futures, the dollar index, Brent crude oil, and U.S. 10-year treasury yields, which significantly impact gold prices.

The study's limitations include the use of limited data (2010-2024) and the potential oversight of influential variables like other precious metals and crude oil types. To improve, future research should consider a broader range of variables and longer time series data. Additionally, integrating advanced prediction methods like machine learning algorithms could enhance accuracy. Furthermore, improving the model's sensitivity to small price changes is crucial.

Future research should focus on expanding the variable scope to include more influencing factors, utilizing longer time series data, and employing a combination of prediction methods. The goal is to improve the model's accuracy, sensitivity, and ability to capture market nuances, ultimately providing more robust predictions for investors and policymakers.

References

[1] *Liu, L. (2021). Prediction and comparison of gold futures prices based on RNN-LSTM [Master's thesis][Translated title]. Changchun University of Science and Technology. https://link.cnki.net/doi/10.26977/d.cnki.gccgc.2021.000152*
[2] *Sims, C. A. (1980). Macroeconomics and reality. Econometrica, 47(1), 1-48.*
[3] *Blose, L. (2011). The relationship between gold and inflation. Journal of Wealth Management, 13(4), 35-42.*
[4] *Nikiforos, M. (2012). The relationship between gold and money supply. International Journal of Economics and Finance, 4(2), 241-249.*
[5] *Ghosh, S. (2014). The relationship between gold and stock market. Journal of Financial Economics, 12(3), 234-246.*
[6] *Koosup, C. (2016). The relationship between gold and oil market. Energy Economics, 56, 389-399.*
[7] *Caporale, G. M. (2005). The application of vector autoregression in the analysis of financial markets. Journal of Economic Surveys, 19(4), 637-670.*

DOI: 10.54254/2754-1169/108/20242024

[8] Wang, P. (2010). The application of vector autoregression in the analysis of international financial markets during the Asian financial crisis. Journal of International Money and Finance, 29(3), 372-384.