ANALYZING FACTORS INFLUENCING GOLD PRICES IN THE UNITED STATES AND PREDICTING FUTURE TRENDS USING MACHINE LEARNING TECHNIQUES

● NGUYEN THANH HUONG TRA

ABSTRACT:

The research focuses mainly on factors leading to the increase in gold prices in the United States during the past two decades (2000 - 2023) and predictions for this precious metal price fluctuation in the next two quarters. The study identifies the factors that are highly correlated with gold and seeks to answer why the price of gold has increased by 5.5 times over the past twenty years, thereby enabling the short-term prediction of gold price movement using the most accurate machine learning model.

Keywords: gold price, inflation rate, exchange rate, stock market, correlation, precious metal commodity.

1. Introduction

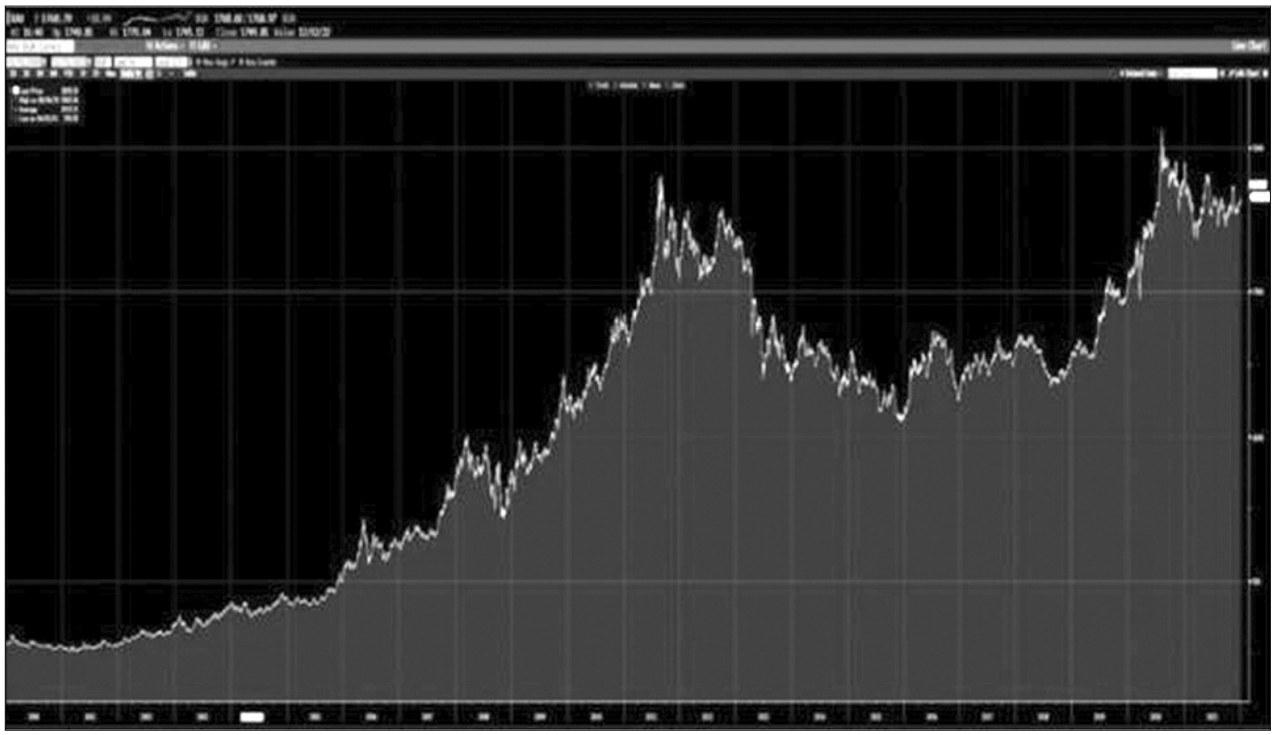
Historically, gold was utilized as a medium of trade for goods. Recently, numerous European nations have adopted the Gold Standards. The Bretton Woods system fixed the US dollar value to $35 per troy ounce of gold (James Chen, 2022). Following the Nixon Shock, this system was discontinued, and a flat currency system was implemented.

However, gold is continuously traded globally based on the intraday spot price obtained from global over-the-counter gold trading markets using the ticker symbol "XAU." Gold is typically traded in US Dollars (XAU/USD), and its price is thought to be associated with stock and bond values. The following graph displays the Gold Price from 2000 to 2022.

Gold prices are subject to market fluctuations and speculation, making it crucial for investors to understand the factors influencing them. Among all precious metals, gold is currently the most popular investment option. Gold has long been seen as a precious commodity, and investment in it is typically viewed as a hedge against economic volatility. Investing in gold is motivated by the belief that it is considerably more stable than government-issued bonds during market volatility. Utilizing mostly futures and derivatives, gold serves to diversify risk. According to the line graph above, gold prices have fluctuated over time. This movement may cause investors to become enthusiastic about gold's profitability. Therefore, gold prices have long been a topic of intense interest and scrutiny. Before investing, individuals often wish to comprehend the gold price fluctuations and the metal's technical analysis.

Theoretical literature suggests that interest rates, inflation rates, exchange rates, and stock and real estate market conditions affect gold prices. Baumeister and Kilian (2016) suggest that changes in interest rates impact gold prices through fluctuations in exchange rates, as indicated by the

Figure 1: Line Chart for the Price of Gold from 2000 to 2022



Source: Bloomberg

interest rate parity theory. The Fisher effect indicates that inflation affects gold prices through nominal interest rates (Baur & McDermott, 2010). The value of gold is influenced by factors such as the rates of other valuable metals, including silver and platinum, as well as fluctuations in currency exchange rates (Batten, Ciner, & Lucey, 2015; Ciner, 2001). Additionally, the perception of gold as a safeguard against economic instability implies an inverse relationship between gold and stock markets, whereby gold prices increase when the stock markets experience a downturn (Baur & Lucey, 2010). By analyzing the interaction between these factors, this study aims to contribute to existing knowledge and aid management professionals in making informed investment and risk management decisions.

Hence, this research aims to offer investors a more intricate comprehension of the elements that impact the value of gold and their connections. The study will be conducted to respond to the following questions:

- What is the relationship between gold prices and other economic indicators, and how strongly do these indicators impact gold prices?
- How do machine learning models help to predict gold prices, and how effective are these methods?

2. Literature review

In this section, the report examines the relationship of the gold price with several economic indicators through numerous studies and historical data.

2.1. Relationship with Inflation Rates

It may seem reasonable that the price of gold will also increase when inflation increases. However, there is yet to be a clear relationship between inflation and gold's price. Gold can, in fact, serve as a hedge against inflation. This is one of the most effective ways to understand the relationship between gold and inflation. Frank Holmes, CEO of U.S. Global Investors, was quoted in April by Barrons as saying, "In gold markets, investors look at real interest rates, subtracting the benchmark rate from the inflation rate. If that rate turns negative, as it has recently, gold shines."

According to common sense, there is a substantial association between the gold rate and inflation. When inflation increases, the demand for gold and the price of gold also increases. The

converse also seems theoretically plausible. A decrease in inflation will result in a decline in the gold rate. This appears to be some reasoning from the perspective of an investor. In periods of high inflation, gold is a more attractive investment alternative. The history of gold rates and inflation, however, contradicts this assertion.

* *Bull Market (1974 - 1980):* From 1974 to January 1980, gold prices climbed by 352% during the bull market, from $129 per ounce to $675 per ounce. The consumer price index, a measure of inflation, grew by 69% during the same period. It is evident that gold rates have outpaced inflation.

* *Bear Market (1980 - 2001):* The bear market accounted for 39% of the decrease in gold prices from a high of $675 per ounce in 1980 to a low of $260 per ounce in April 2001. The consumer price index rose by 126% throughout the same period.

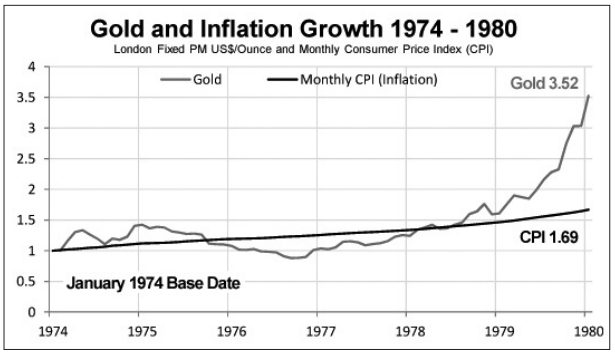
* *Bull Market (2001 - 2011):* Throughout the first decade of the 21st century, the bull market triggered a 527% surge in gold prices. Comparatively, the consumer price index increased by only 25%.

According to Wall Street Journal finance experts, a correlation coefficient of 0.08 was discovered between the gold rate and inflation between 1978 and 2010, indicating no relationship between the two factors (Hulbert, 2021). Citibank commissioned a second study in 2009 that concluded "there is no significant association between the gold price and inflation" (Crawley, 2021).

Numerous scholarly investigations focused on how gold prices react to changes in inflation rates also have mixed results on the relation between gold prices and inflation rates. Frenkel and Jovanovic (1981) documented that gold prices rose in correlation with U.S. inflation from 1950 to 1980. In 2010, Baur and McDermott found that the price of gold and inflation shared a comparable correlation in the US, UK, and Germany from 1970 to 2008. Kaul and Sapp's (2006) study revealed that gold is a reliable protection against unforeseen inflation but not expected inflation.

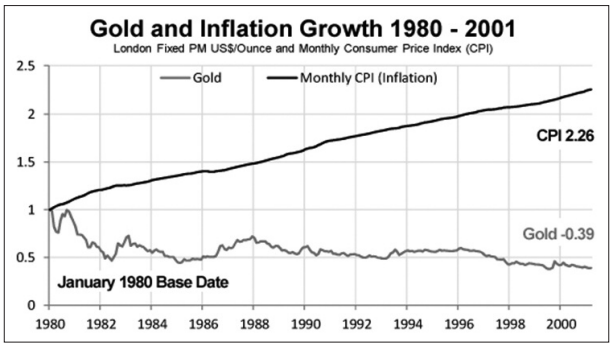
Nonetheless, from 1971 to 2011, Lucey et al. (2017) utilized a time-series approach to examine the relationship between gold and inflation in 23 countries. The results imply a time- and region-specific variation in the relationship between inflation and gold. In some nations, gold has been shown to correlate positively with inflation, whereas,

Figure 2: Gold and Inflation Growth 1974-1980



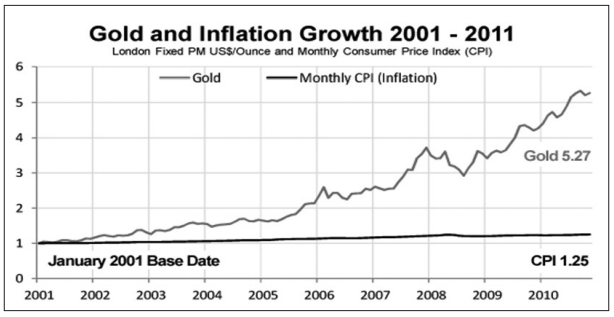
Source: goldforbeginner.com

Figure 3: Gold and Inflation Growth 1980-2001



Source: goldforbeginner.com

Figure 4: Gold and Inflation Growth 2001-2011



Source: goldforbeginner.com

in others, it has shown no correlation at all. Furthermore, these scholars observe that macroeconomic circumstances and monetary measures can affect the link between gold and inflation.

2.2. Relationship with Exchange Rates

Gold spot prices are universal, while the values of foreign currencies are influenced by the currency's demand and supply and the country's net exports. In a market economy, the price of a currency changes according to supply and demand.

If the exchange rate is exceptionally high, the money supply will exceed demand, resulting in a depreciation of the currency's value. If the price is too low, supply will fall short of demand, increasing the rate. Moreover, a country's net exports affect the value fluctuations of its currency. If exports exceed imports, the currency's value will climb. If exports are less than imports, the currencys value will decline. Hence, theoretically, there is no clear relationship between the gold price and the exchange rate.

Numerous financial studies have studied the value of gold in connection to currency exchange rates. While some scholars report a strong positive association, others have found no correlation or negative relationship.

Dornbusch and Fischer's study (1980) on how gold affected currency exchange rates was regarded as groundbreaking work. They found a direct association between the value of the US dollar and the price of gold, indicating that market factors account for some of the fluctuations in gold's price. Baur and Lucey's study (2010), which employed a dynamic conditional correlation model, examined the correlation between gold and various exchange rates. The Euro, Japanese yen, Swiss franc, and British Pound have positive correlations with gold, while the Australian and Canadian Dollars have negative correlations. Sari and Soytas's research (2006) analyzed the correlation between gold prices and six major currencies between 1979 and 2004, finding that a higher gold price was linked to a more favorable exchange rate for all six currencies. A study by Apergis and Miller (2006) analyzed the relationship between gold prices and the exchange rate of the US currency over 31 years (1973-2004) and found a notable positive correlation between the two in the long run. The results suggested that gold could be a reliable safeguard against inflation and the currencys depreciation. On the other hand, Tully and Lucey (2007) used a power GARCH analysis and found a positive correlation between gold prices and the value of various currencies. They proposed that gold could protect against currency value fluctuations, especially during economic uncertainty.

In summary, although there are mixed results, most research results indicate a favorable association between the value of gold and the exchange rates. This suggests that gold can serve as a helpful safeguard against currency risks and inflation, particularly in times of economic instability where currency devaluation is a concern.

2.3. Relationship with Interest Rates

Theoretically, declining and rising interest rates could affect gold's price. Investors seeking relatively secure and low-risk investments tend to buy gold when interest rates fall. Together with other factors, this demand has contributed to a negative correlation between low-interest rates and gold. Hence, if prices decrease, it may be an excellent time to add gold to a portfolio, as investors may be able to purchase more with the same budget, and the gold may appreciate as interest rates inevitably fluctuate once more.

Gold's historical and global association with interest rates has been studied by many scholars, including Baur and McDermott (2010), who analyzed data from Canada, Australia, the US, and the UK. They found a negative correlation between gold prices and short-term interest rates but not between gold prices and long-term interest rates. Yildirim et al. (2022) also investigated this relationship by studying Turkish interest rates and gold prices from 1990 to 2018. Through Granger causality tests and VAR models, they established a significant cause-and-effect relationship between Turkish interest rates and gold prices.

However, certain scholars argue that interest rates are not affected by changes in the price of gold. According to a study by Capie and Mills (2005), there was no observable relationship between gold and interest rates in the UK between 1880 and 1970. In relation to the Japanese stock market, Mukherjee and Naka (1995) investigated the gold price, interest rates, and other macroeconomic variables. They observed that the relationships among these variables were intricate, and their causality was ambiguous. Sjaastad and Scacciavillani (1996) also examined interest rates. They concluded that there was no statistically significant correlation between gold prices and interest rates, implying that currency exchange rates and inflation were the primary drivers of gold demand.

2.4. Relationship with the Stock Market

Several research investigations have been carried out to examine the correlation between the prices of gold and the stock market since gold is frequently viewed as a secure investment during

times of economic instability, while stocks entail more risk.

However, the connection between the stock market and gold is not merely a straightforward inverse relationship; instead, it is more intricate. Baur and Lucey (2010) discovered a favorable long-term association between gold and stock market returns, although it can turn negative amid an economic crisis. Other studies have found mixed results, with Ewing and Malik (2015) concluding that gold positively correlates with the US stock market, while Wang and Wu (2012) found a negative association between the two variables. During severe inflation, Jaffe and Mandelker (1976) showed a negative link between gold and the US stock market, while Ciner (2001) found similar results when comparing gold to the stock and crude oil markets.

Recent research by He et al. (2022) has investigated the relationship between gold prices, stock market returns, and oil prices. They found that gold and oil price variations correlated with stock market performance, suggesting a spillover effect. Gold prices were positively correlated with the US NASDAQ but negatively correlated with the Shanghai SSE. Similarly, Samour A. et al. (2022) examined the relationship between gold and the South African stock markets, considering the impact of crude oil. Their study found a long-term positive correlation between gold and stock market prices in South Africa, making gold a secure investment. However, short-term gold price changes may lower stock market profits.

In summary, the correlation between the stock market and gold prices is intricate and relies on different economic circumstances and variables, including the price of oil. The recent volatility in gold prices also demonstrated that Gold prices are unrelated to NASDAQ fluctuation.

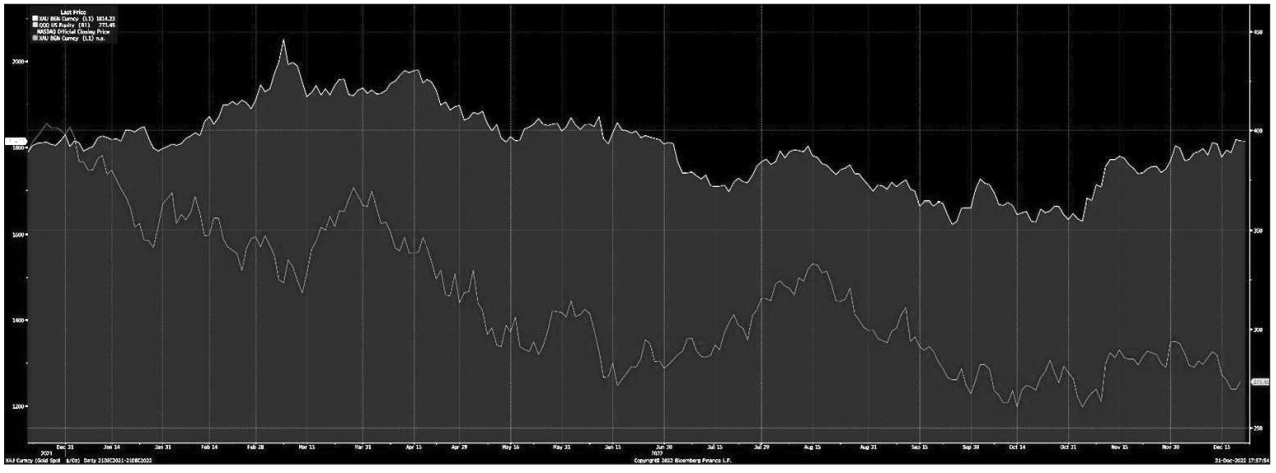
2.5. Relationship with the GDP

There are two views for determining the relationship between Gold and GDP: consumer and investor perspectives. From a consumer's standpoint, when the economy proliferates (as opposed to slowly or declining), people's incomes and spending power increase (as opposed to stagnating or dropping), and so make the demand for gold, and vice versa. Gold is a vote of no confidence in the U.S. economy from investors perspective. Hence, there should be a negative link between GDP growth and gold prices. When the economy expands, investment demand (attracted by the safe-haven nature of the precious metal) decreases and vice versa.

Numerous extensive discussions and studies have been conducted regarding the correlation between the price of gold and the GDP. Regarding whether the two variables are correlated, researchers are divided. Some respond affirmatively, while others respond negatively.

Several studies have shown a favorable long-term relationship between the value of gold and GDP. The macroeconomic variables gold prices and GDP were examined by Baur and McDermott

Figure 5: Relationship between the Price of Gold and the NASDAQ



Source: Bloomberg

(2010), who discovered a positive relationship over time. Batten et al. (2010) found a positive correlation between gold prices and GDP development in 16 countries. Because Turkey is a net importer of gold, Balibey (2015) discovered a negative correlation between gold prices and Turkey's GDP.

However, some studies indicate that the price of gold does not have a direct causal link to the growth of GDP. For instance, Kim et al. (2017) surveyed the Asia-Pacific region and observed no correlation between gold prices and GDP growth. Similarly, Ciner et al. (2013) reported a nonlinear relationship between gold prices and GDP growth in the US, Germany, and Japan. The results indicate that the link between the value of gold and GDP growth could differ depending on the geographical area and economic circumstances.

These disparate outcomes may result from the relationship between gold prices and GDP fluctuating with the economic climate. There is a positive correlation between economic growth and the price of gold, as investors may be more risk-tolerant during prosperous periods. In contrast, gold prices may decline during economic downturns as investors seek safe-haven assets.

In summary, more than one straightforward explanation exists for the complex and diverse relationship between the gold price and GDP. Numerous variables, such as the economy, global politics, and investor sentiment, influence the connection's potency.

2.6. Relationship with other precious metal commodities

Many studies have been conducted on the correlation between the prices of gold and silver, but the results have been inconsistent. Certain studies have established a positive or negative association between the two metals, while others have found none. In 2006, Apergis and Miller analyzed monthly data from 1975 to 2005 and found a direct connection between the prices of gold and silver. Using daily data from 1990 to 2010, Baur and McDermott (2010) demonstrated a relationship between the two metals. Bildirici and Turkmen (2015) found an inverse correlation using daily data from 2004 to 2014.

Gold and silver price relationships have also been studied on lesser time scales. Batten et al. (2015) and Tiwari et al. (2019) discovered a positive short-term association between the two metals using daily data from 2001 to 2012. The possibility of a cause-and-effect connection between the prices of gold and silver has also been explored by researchers. Sami (2021) examined daily data from 2001 to 2015 and found that the price of gold and silver followed a similar pattern. Narayan et al. (2015) found evidence of a bidirectional causal relationship using daily data from 1985 to 2015.

Both investors and academics have recognized the correlation between gold and platinum prices. Given that both metals are regarded as safe-haven investments, Eryigit (2017) proposes a positive correlation between them. However, Kearney and Lombra (2009) discovered that the volatility of one metal might have the opposite effect on another. Depending on the period and the causal direction, the association between the prices of gold and silver could alter. Researchers need to concur on the extent to which market factors explain the correlation between gold and platinum prices.

In summary, various research has examined the connection between the value of gold and several macroeconomic indicators such as inflation, exchange rates, interest rates, stock market performance, GDP, and other precious metals. The results of the reviewed studies suggest that the correlation between gold prices and these economic factors is complex and varies depending on the context. While some studies have identified a strong or moderate correlation between gold prices and these indicators, others have found no correlation or conflicting results. In addition, there needs to be more research identifying the significant factors that influence the price of gold. Therefore, further research is required to identify the most influential variables affecting gold prices and develop accurate forecasting models. This study aims to fill this research gap by thoroughly investigating the significant factors influencing gold prices and creating a prediction model to forecast gold prices accurately. The following section will present the data and Explanatory Data Analysis (EDA) to thoroughly comprehend the role of various economic factors in determining gold prices.

3. Data and statistical analysis

3.1. Data

The data was compiled using Bloomberg Terminal, FRED, SDBulletin, NASDAQ, and Yahoo Finance. For prediction, Bloomberg Terminal,

CoinForestCast, and The Financial Forecast Center provide the data of independent variables needed to anticipate the return of gold over the following six months. The definitions and sample of the data used in this investigation are presented in the Appendix A. The meaning of the variables replaces missing values.

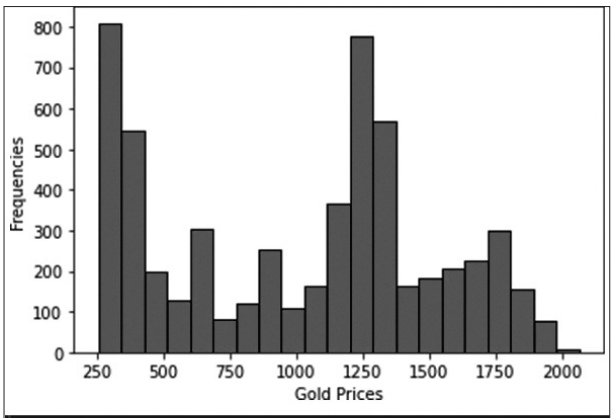
3.2. Exploratory data analysis

Exploratory Data Analysis is usually the first step after obtaining the data. It entails the process of performing initial investigations on the data to discover patterns, to gauge if there are any outliers or anomalies. It also helps to test the hypothesis and to check if appropriate assumptions can be made with the help of summary statistics and graphical representations.

Exploratory Data Analysis is usually the first step after obtaining the data. It entails the process of performing initial investigations on the data to discover patterns, to gauge if there are any outliers or anomalies. It also helps to test the hypothesis and to check if appropriate assumptions can be made with the help of summary statistics and graphical representations. Appendix B shows statistical figures of the exploratory analysis.

The statistical summary table (Figure 6) shows that the average price of gold over the previous two decades was $1,012.62, with a standard deviation of nearly $507. The minimum price of gold is $255.55, while the maximum is $2063.54, and gold prices frequently fluctuate at $250 or $1250. The kurtosis value of the gold price is around -1.26, suggesting that its probability distribution has a less prominent peak and slimmer tails than a normal distribution. Additionally, the skewness of the gold price is -0.07 (Figure 7), which is damaging and indicates a left-skewed distribution. Consequently, there are more data values near the mean and fewer data values on the tails, and the standard of the gold price is below the median. The p-value of gold with other independent variables depicts depict the statistically significant relationship between gold and other independent variables. As the p-values of all variables are less than 0.05 (Figure 8), it is possible to conclude that there is a substantial relationship between gold prices and analyzed variables. It indicates that the gold price will likely increase if the silver price changes favorably and/or if the GDP growth rate falls.

Figure 7: Statistical Summary of Variables



Source: Analysis by the author

Figure 6: Statistical Summary of Variables

Index	Gold	Silver	Platinum	EUR	GBP	JPY	GoldReturn1	GDPGrowth	Interest	Inflation	AdjustingMarket	NASDAQ
count	5740	5740	5740	5740	5740	5740	5740	5740	5740	5740	5740	5740
mean	1012.62	15.676	1079.88	1.20529	1.56392	0.00951574	0.000378329	0.0212521	0.0104184	0.0224338	0.0105041	0.000371672
std	506.941	8.54605	378.082	0.159887	0.211526	0.00127704	0.0106489	0.0184483	0.0155421	0.0138105	0.0153821	0.0151494
min	255.55	4.06	412.5	0.8272	1.1485	0.0074	-0.0907368	0.0025	-0.0882764	-0.021	-0.0314928	-0.123212
25%	439.312	7.395	836.225	1.11367	1.40755	0.0086	-0.00475818	0.0075	0.0068742	0.015	0.007408	-0.0053245
50%	1167.49	15.7303	989.475	1.20855	1.5565	0.0092	0.000394541	0.0125	0.0110864	0.021	0.0123491	0.000371672
75%	1341.01	19.444	1384.75	1.3217	1.66817	0.0098	0.00586199	0.03	0.0140057	0.031	0.017783	0.00695975
max	2063.54	48.4438	2250.5	1.5991	2.1075	0.0132	0.107883	0.0625	0.0878811	0.07	0.0557132	0.141732

Source: Analysis by the author

Figure 8: P-value of gold with other independent variables

Index	Variables	P-value
0	Silver	0
1	Platinum	0
2	EUR	0
3	GBP	0
4	JPY	0
5	GDP	0.0022
6	Interest	0
7	Inflation	0
8	Housing Market	0
9	NASDAQ	0.0478

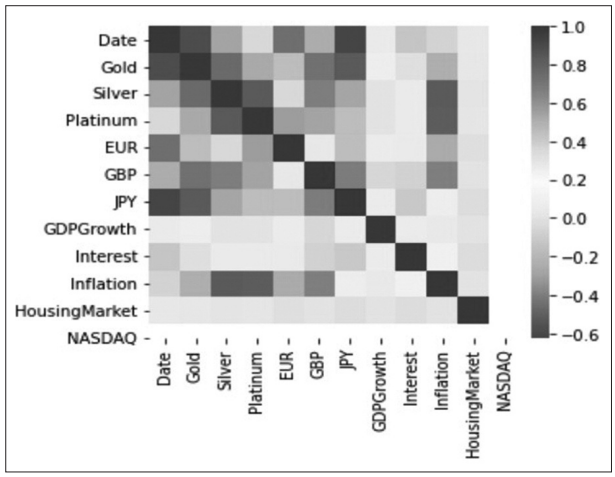
Source: Analysis by the author

The correlation matrix's heatmap and pair plot reveal that gold has the most vital positive relationship with silver price and the weakest negative relationship with GDP growth. The figure shows that with Silver under $20 and inflation around 0.02, there is more possibility that gold prices will increase. The data pair plot analysis reveals a significant relationship between the gold price and inflation and the silver price. In addition, the graph of gold prices over time shows a pattern of substantial fluctuations. The 2008, financial market crisis is frequently attributed to driving up the price of gold. The pandemic crisis of 2022 drove the price of gold above the levels in 2008. These findings highlight the impact of external economic forces on the gold market, which emphasizes the importance of accurate gold price forecasting.

4. Methodology

Logistic regression is a supervised machine-learning technique that attempts to predict the likelihood of a particular outcome, occurrence, or

Figure 9: Correlation Matrix



Source: Analysis by the author

observation. It produces a discrete, binary output representing the positive or negative return on the gold price. This technique models the probability of a binary outcome variable given one or more predictor variables, assuming that the predictor variables have a linear relationship with the log odds of the outcome variable. The dependent variable is categorical, meaning it can take on one of two values (0 or 1) that represent the presence or absence of an event (Agresti, 2015).

Utilizing logistic regression has several advantages. First, its implementation and interpretation are simple. Second, it can accommodate categorical and continuous predictor variables (Hosmer & Lemeshow, 2004). Thirdly, it considers predictor variable interactions (Menard, 2002). Fourth, it provides an estimate of the probability of the outcome, which can aid decision-making (Pepe, 2003).

Nonetheless, there are constraints to consider. First, logistic regression assumes that the log odds of the outcome and predictor variables have a linear relationship. Second, precise estimates necessitate a large sample size (Vittinghoff & McCulloch, 2007). Thirdly, multicollinearity among the variables utilized for forecasting can affect its precision.

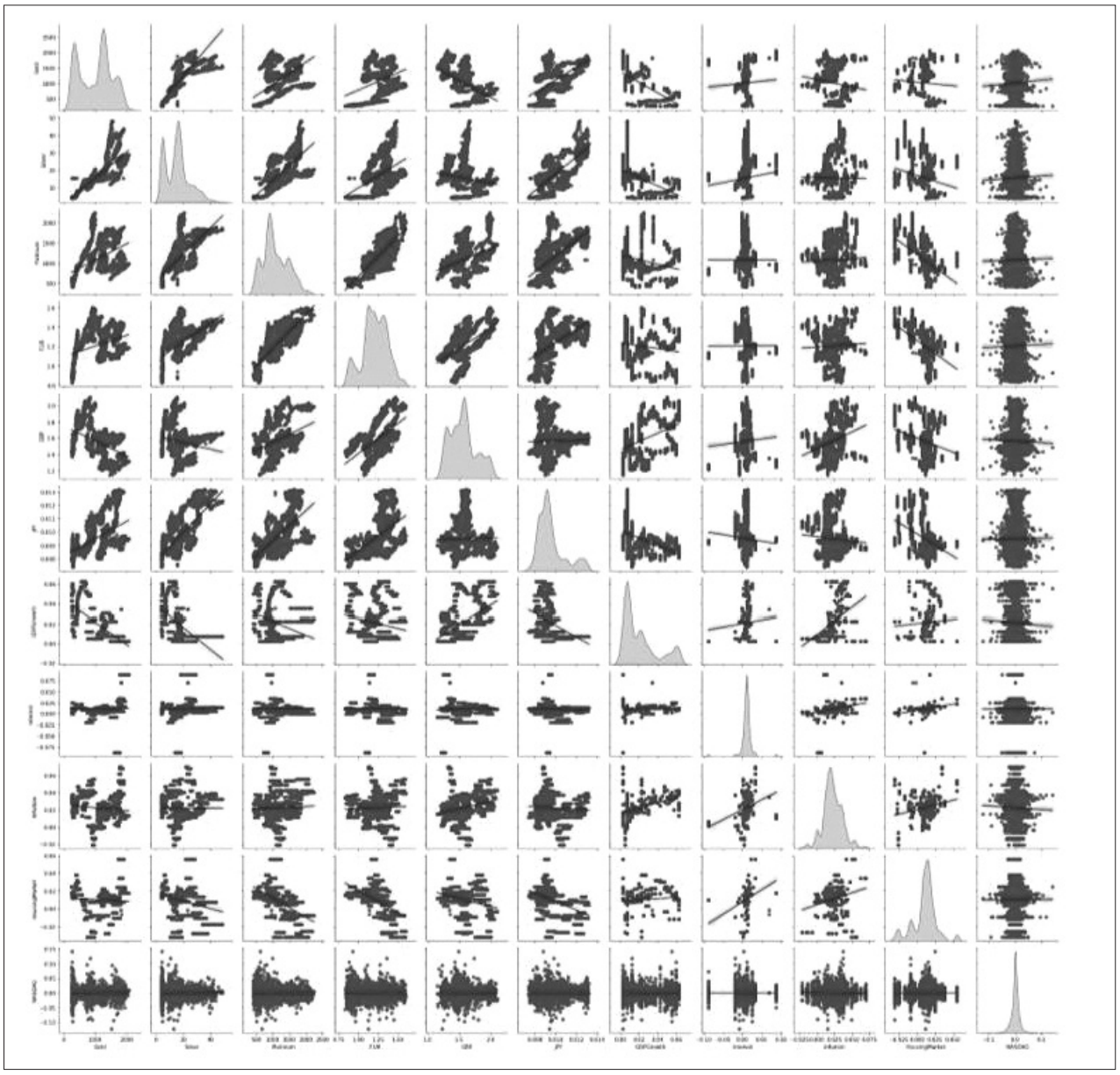
The following equation characterizes logistic regression:

$$y = \frac{e^{(b_0 + b_1 x)}}{1 + e^{(b_0 + b_1 x)}}$$

In which:

- The value of (y) represents the predicted value

Figure 10: Pair plot



Source: Analysis by the author

based on an input value of (x)

- The value of b0 represents the bias or intercept term, which is the predicted output when the input value is zero.

- The value of b1 represents the coefficient for the input variable (x), which indicates how much the output value changes for a unit change in the input value.

- The model uses the input variable (x) to predict the output value (y) by estimating the values of b0 and b1 utilizing training data.

- Once the model is trained, it can predict the output value (y) for new input values (x) using the estimated b0 and b1.

Logistic regression is a technique that predicts the probability that an event will occur based on the values of the variables used to predict it. The model assumes that a straight-line connection exists between the predictor variables and the logarithm of the odds of the outcome variable. It is possible to convert the log odds to the result's probability using the logistic function. The logistic function is a

sigmoid curve ranging from 0 to 1, derived from its name. The logistic regression model aims to determine the coefficients of the predictor variables that yield the highest probability of the observed data regarding these coefficients. By utilizing these coefficients, the model can forecast the likelihood of the result for new observations.

One of the fundamentals of linear regression is using weights or coefficient values to predict an output value by linearly integrating the input values. In certain circumstances, however, the output value is binary, represented as either 0 or 1. This form of binary prediction also employs a linear model, but the approach and techniques employed are distinct from those used in linear regression.

In summary, Logistic Regression is a proven statistical technique for predicting binary outcomes from a set of predictor variables. Its strengths lie in its ability to handle continuous and categorical data while remaining relatively simple and adaptable. However, its predictive ability is sensitive to sample size, linearity assumption, and multicollinearity issues. The objective of this study is to use Logistic Regression to predict whether the price of gold will rise or fall in response to a set of economic variables. An evaluation of the model's accuracy and a forecast of the outcome will be followed by an examination of the data and the gold price prediction accuracy of the model.

In order to reduce the limitations of the Logistic Regression Model, there are several other machine learning and regression models are used to predict the gold price.

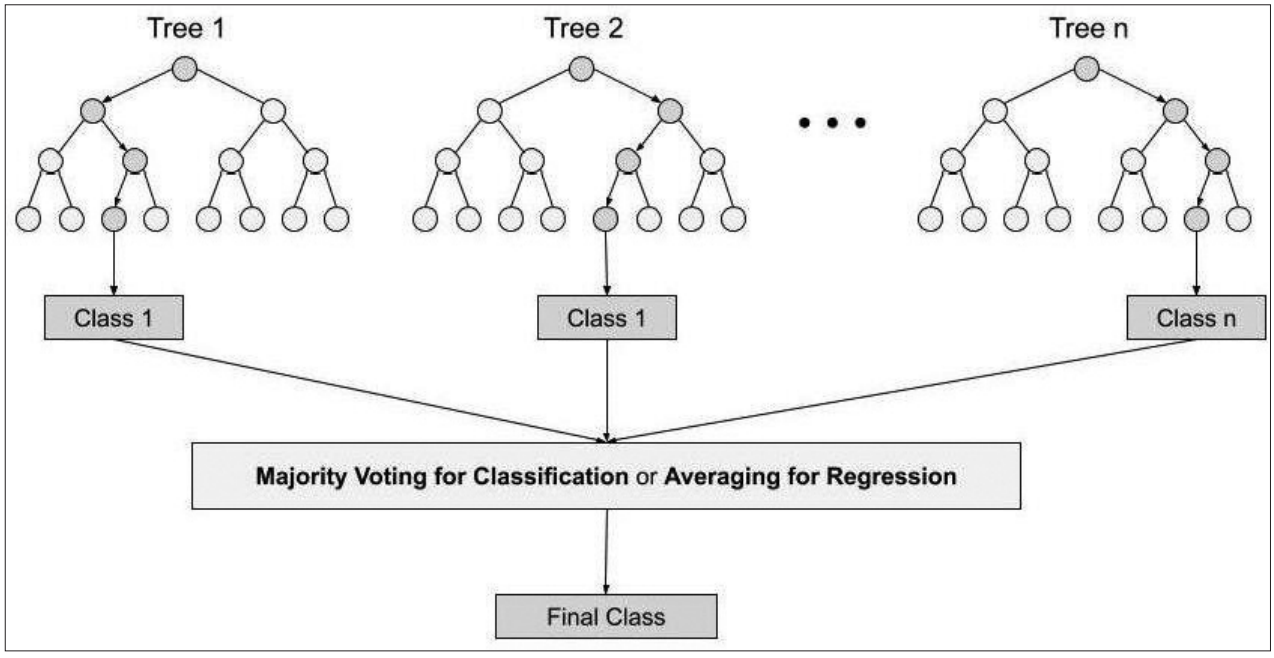
Polynomial Regression is employed when other models fail to capture the points in the data and provide a suitable conclusion. It is crucial to remember that Polynomial Regression is sensitive to outliers. Consequently, one or more outliers can negatively impact its performance. A nth-degree polynomial represents the link between the independent and dependent variables.

Random Forest is a meta-estimator that uses an average to increase the predicted accuracy of the model and prevent over-fitting. The sub-sample size is always the same as the original, but replacement sampling is used. It reduces over-fitting but is a sluggish, difficult-to-construct, and sophisticated real-time prediction system. The Random Forest Method consists of the steps listed below:

- In the random forest approach, n random records are selected from a dataset containing k-records.

- For each sample, individual decision trees are generated.

Figure 11: Random Forest Regression Illustration



Source: Analytics Vidhya

- Each decision tree will produce a result.
- For classification and regression, the result is determined by Majority Voting or Averaging.

Boosting is an ensemble method that modifies the training data and adjusts the weight of the observations based on the previous regression. Boosting reduces bias error but can occasionally result in overfitting the training dataset as it continues to improve to minimize all mistakes. The boosting model comprises these steps:

- A subset where all data points are given equal weights is created from the training dataset.
- A based model is created for the initial dataset and used for predictions on the entire dataset.
- Errors are calculated using predicted and actual values. The observation that was mispredicted is given a higher weight.
- The following model is created, and boosting tries to correct the errors of the previous model.
- The process is repeated for multiple models, each correcting the errors of the previous model.
- Therefore, the final model is a strong learner and is the weighted mean of all the models, which are weak learners.

Bagging, also known as Bootstrap aggregation, is a method of ensemble learning that seeks out diverse ensemble learners by altering the training dataset. The bagging technique uses random sampling and replacement to generate new data for the training dataset, which may result in duplicate observations across datasets. Every bagged observation has an equal chance of being included in the subsets of training data. Then, multiple models are trained to utilize these datasets. Bagging is commonly used to integrate the results of various decision tree models and construct the random forest technique. The average of the trees with high variation and low bias improves accuracy. The bagging model consists of the following steps:

- Create multiple subsets from the training dataset by selecting observations with replacements.
- Create a base model (also called a weak model).
- Run base models simultaneously and independently of each other.
- Combine predictions from all base models to determine the outcome.

5. Results

The following table below summarizes machine learning models' accuracy in predicting gold return.

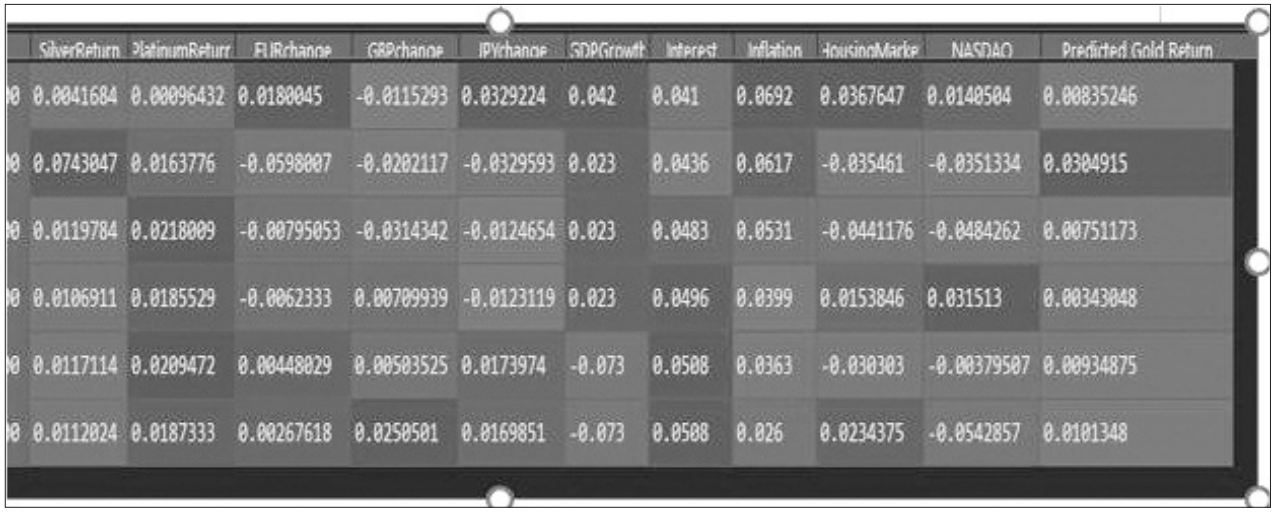
This demonstrates that Random Forest provides the highest accuracy for the train and test sets, with an R-square score close to 1. However, the gradient boosting or bagging model is typically more accurate. They rely on intuition and believe that the best possible next model, when combined with previous models, minimizes the overall prediction error). Random forests can be less accurate than gradient-boosting trees, according to Simic (2020). Therefore, it is anticipated that bagging and boosting will perform better. However, according to MLOps Blog, Boosting and Bagging do not perform well with unstructured and sparse data. The data used in this study have different update frequencies for different variables and contain missing data, which is replaced with the mean value when performing

Figure 12: Model Accuracy

Index	Method	MSE Test	MSE Train	-square Tes	-square trai
0	Multiple Linear Regression	3.8549e-05	4.3729e-05	0.64792	0.619898
1	Polynomial Linear Regression	4.40149e-05	4.00654e-05	0.597999	0.651743
2	Random Forest	4.1068e-05	6.57312e-06	0.624913	0.942865
3	Gradient Boosting	4.01619e-05	2.98701e-05	0.633189	0.740363
4	Bagging	3.85545e-05	1.63489e-05	0.647869	0.857892

Source: Analysis by the author

Figure 13: Gold Price Prediction



	SilverReturn	PlatinumReturr	EURchange	GBPchange	JPYchange	GDPGrowth	Interest	Inflation	HousingMarke	NASDAQ	Predicted Gold Return
0	0.0041684	0.00096432	0.0180045	-0.0115293	0.0329224	0.042	0.041	0.0692	0.0367647	0.0140504	0.00835246
0	0.0743047	0.0163776	-0.0598007	-0.0202117	-0.0329593	0.023	0.0436	0.0617	-0.035461	-0.0351334	0.0304915
0	0.0119784	0.0218009	-0.00795053	-0.0314342	-0.0124654	0.023	0.0483	0.0531	-0.0441176	-0.0484262	0.00751173
0	0.0106911	0.0185529	-0.0062333	0.00709939	-0.0123119	0.023	0.0496	0.0399	0.0153846	0.031513	0.00343048
0	0.0117114	0.0209472	0.00448029	0.00503525	0.0173974	-0.073	0.0508	0.0363	-0.030303	-0.00379507	0.00934875
0	0.0112024	0.0187333	0.00267618	0.0250501	0.0169851	-0.073	0.0508	0.026	0.0234375	-0.0542857	0.0101348

Source: Analysis by the author

analysis. On the other hand, the Random Forest model is regarded as one of the most effective machine learning models for missing value data. As the data are trained to correct each other's errors, they can identify complex data patterns. If the data are, however, noisy, the boosted trees may begin to model the noise and overfit. Therefore, because the random forest is more tolerant of missing data, it achieves higher precision with data containing multiple missing values. Consequently, this study will use the random forest to predict the movement of gold returns over the next six months, producing the following results:

The random forest model predicts that the gold return will be positive over the next six months, indicating that gold prices will increase in the short term. Therefore, gold traders can consider short positions when the price of gold rises.

6. Conclusion

The regression models aid in determining the relationship between the gold price and other independent variables, such as silver price, platinum price, the exchange rate of EUR/USD, GBP/USD, JPY/USD, and economic factors, such as GDP growth, inflation, and interest rates, as well as market factors, such as the return of housing index and NASDAQ market. The report demonstrates that the silver price has had a substantial positive effect on the gold price, whereas GDP growth has the most significant indirect relationship with the gold price. The random forest model predicts that the gold price will rise over the next six months with the highest degree of precision among the five regression models. The study will assist gold traders in identifying the potential gold price movement, enabling them to formulate an appropriate trading strategy. Due to the study's short duration, the lack of sufficient forecast data for independent variables, and the fact that markets fluctuated significantly after wars and pandemics, traders must also closely monitor market movement to develop an appropriate trading strategy ■

REFERENCES:

1. Apergis, N., & Miller, S. M. (2006). Do structural oil-market shocks affect stock prices? Energy Economics, 28(4), 453-469.

2. Agresti, A. (2015). Foundations of linear and generalized linear models. USA: Wiley.

3. Baumeister, C., & Kilian, L. (2016). Forty years of oil price fluctuations: Why the price of oil may still surprise us. Journal of Economic Perspectives, 30(1), 139-160.

4. Baur, D. G., & Lucey, B. M. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds, and gold. Financial Review, 45(2), 217-229.

5. Baur, D. G., & McDermott, T. K. (2010). Is gold a safe haven? International evidence. Journal of Banking & Finance, 34(8), 1886-1898.
6. Balibey, M. (2015). The dynamic relationship between gold and macroeconomic variables: Evidence from Turkey. Economic Modelling, 50, 231-238.
7. Batten, J. A., Ciner, C., & Lucey, B. M. (2010). The macroeconomic determinants of volatility in precious metals markets. Resources Policy, 35(2), 65-71.
8. Batten, J. A., Ciner, C., Lucey, B. M., & Yarovaya, L. (2015). The dynamic linkages between crude oil and natural gas markets. International Review of Financial Analysis, 39, 46-53.
9. Bildirici, M. (2015). The Chaotic Relationship between Oil Return, Gold, Silver and Copper Returns in TURKEY: Non-Linear ARDL and Augmented Non-linear Granger Causality. Procedia - Social and Behavioral Sciences, 21, 397-407.
10. Capie, F., & Mills, T. C. (2005). Gold as a hedge against the dollar. Journal of International Financial Markets, Institutions and Money, 15(4), 343-352.
11. Ciner, C., Gurdgiev, C., & Lucey, B. M. (2013). Hedges and safe havens: An examination of stocks, bonds, gold, oil, and exchange rates. International Review of Financial Analysis, 29, 202-211.
12. Ciner, C. (2001). Energy shocks and financial markets: nonlinear linkages. Studies in Nonlinear Dynamics & Econometrics, 5(3), 1-11.
13. Crawley, J (2021). Gold as a Strategic Inflation Hedge. [Online] Available at https://caia.org/blog/2021/06/27/gold-as-a-strategic-inflation-hedge
14. Dornbusch, R., & Fischer, S. (1980). Exchange rates and the current account. The American Economic Review, 70(5), 960-971.
15. Eryigit, M. (2017). Short-term and long-term relationships between gold prices and precious metal (palladium, silver and platinum) and energy (crude oil and gasoline prices. Economic Research-Ekonomska Istra ivanja,30(1),499-510.
16. Ewing, B. T., & Malik, F. (2015). Volatility spillovers between gold, oil and financial markets: evidence from the Gulf Cooperation Council countries. Energy Economics, 51, 463-474.
17. Frenkel, J. A., & Jovanovic, B. (1981). Optimal international reserves: A stochastic framework. Journal of International Economics, 11(2), 219-237.
18. He X., Takiguchi T., Hamori S., Tadahiro N (2022). Spillover effects between energies, gold, and stock: the United States versus China. Energy & Environment, 31(8), 1416-1447.
19. Hosmer Jr, D. W., & Lemeshow, S. (2004). Applied logistic regression. USA: Wiley.
20. Hulbert, M (2021). Gold as an Inflation Hedge: What the Past 50 Years Teaches Us. Wall Street Journal [Online] Available at https://www.wsj.com/articles/gold-as-an-inflation-hedge-what-the-past-50-years-teaches-us-11628283272
21. Jaffe, J. F., & Mandelker, G. N. (1976). The "Fisher Effect" for risky assets: an empirical investigation. Journal of Finance, 31(2), 447-458.
22. James Chen (2022). Bretton Woods Agreement and the Institutions It Created Explained. [Online] Available at https://www.investopedia.com/terms/b/brettonwoodsagreement.asp
23. Kaul, A., & Sapp, S. (2006) Y2K fears and safe-haven trading of the U.S. dollar. Journal of International Money and Finance, 25(5), 760-779.
24. Kearney, A., Lombra R. (2009). Gold and platinum: Toward solving the price puzzle. The Quarterly Review of Economics and Finance, 49, 884-892.
25. Kim, J., Charles A., Darne O. (2017). International stock return predictability: Evidence from new statistical tests. Journal of Financial Economics, 110(2), 357-377.
26. Lucey B., Sharma S., Vigne S (2017). Gold and inflation(s) - A time-varying relationship. Economic Modelling, 67, 88-101.
27. Menard, S. (2002). Applied logistic regression analysis. USA: Sage.
28. Mukherjee, S., & Naka, A. (1995). Dynamic relations between macroeconomic variables and the Japanese stock market: An application of a vector error correction model. Journal of Financial Research, 18(2), 223-237.
29. Narayan, P. K., Narayan, S., & Sharma, S. S. (2015). An analysis of commodity markets: What gain for investors? Journal of International Financial Markets, Institutions, and Money, 34, 245-262.

30. Pepe, M. S. (2003). The statistical evaluation of medical tests for classification and prediction. UK: Oxford University Press.
31. Sami, J. (2021). Has the long-run relationship between gold and silver prices really disappeared? Evidence from an emerging market. Resources Policy, 74, 176-186.
32. Samour A., Moyo D., Tursoy T. (2022). Oil Price, Gold Price, and the Equity Market Nexus in South Africa: Evidence from Bootstrap ARDL. Studies of Applied Economics, 40(2), 1-14.
33. Simic, Milos (2023). Gradient Boosting Trees vs. Random Forests. [Online] Availabile at https://www.baeldung.com/cs/gradient-boosting-trees-vs-random-forests
34. Sjaastad, L. A., & Scacciavillani, F. (1996). The price of gold and the exchange rates: Once again. Journal of International Money and Finance, 15(6), 939-954.
35. Soytas, U., & Sari, R. (2007). Dynamics of oil price, precious metal prices, and exchange rate. Energy Economics, 29(5), 936-955.
36. Tiwari, A. K., Mishra, B., Pradhan A., Shahbaz M. (2019). The dynamic causality between gold and silver prices in India: Evidence using time-varying and non-linear approaches. Resources Policy, 62, 66-76.
37. Tully, E., & Lucey, B. M. (2007). A power GARCH examination of the gold market. Research in International Business and Finance, 21(2), 316-325.
38. Vittinghoff, E., & McCulloch, C. E. (2007). Relaxing the rule of ten events per variable in logistic and Cox regression. American Journal of Epidemiology, 165(6), 710-718.
39. Wang, K. M., & Wu, L. P. (2012). Oil price shocks and stock market activities: Evidence from oil-importing and oil-exporting countries. Journal of Comparative Economics, 40(3), 416-427.
40. Yildirim H.; Alola A.; Sengelen H (2022). The causal nexus of interest rate policy and gold market: The case of Turkey. Journal of Public Affairs, 21(1), e2142.

Received date: June 5, 2023
Reviewed date: July 10, 2023
Accepted date: July 25, 2023

Author information:
NGUYEN THANH HUONG TRA
Stevens Institute of Technology

PHÂN TÍCH CÁC YẾU TỐ ẢNH HƯỞNG ĐẾN GIÁ VÀNG TẠI HOA KỲ VÀ DỰ ĐOÁN XU HƯỚNG TRONG TƯƠNG LAI BẰNG PHƯƠNG PHÁP HỌC MÁY

● **NGUYỄN THANH HƯƠNG TRÀ**
Học viện Công nghệ Stevens

TÓM TẮT:

Nghiên cứu này phân tích các yếu tố dẫn đến sự tăng giá vàng ở Mỹ trong 2 thập kỷ qua (2000 - 2023) và dự đoán biến động giá kim loại quý này trong 2 quý tới. Nghiên cứu nhằm xác định các yếu tố có mối tương quan cao với giá vàng và lý do tại sao giá vàng đã tăng 5,5 lần trong 20 năm qua, từ đó cho phép dự đoán ngắn hạn về biến động giá vàng bằng mô hình học máy, có độ chính xác cao nhất.

Từ khóa: giá vàng, tỷ lệ lạm phát, tỷ giá hối đoái, thị trường chứng khoán, mối tương quan, kim loại quý.