**GOLD PRICES**

STAT 486 – APPLIED STATISTICS AND ECONOMETRICS

FALL 2024-2025

DEPARTMENT OF STATISTICS OF

MIDDLE EAST TECHNICAL UNIVERSITY

BY

MEDİNE YAZICI

January 2025

**ABSTRACT**

This project aims to predict future price movements by examining and modeling the gold data set with time series analysis techniques. In this study, characteristic features such as seasonality, long-term trend, stationarity, and volatility in the data. Various methods were applied, such as ADF (Augmented Dickey-Fuller), ACF (Autocorrelation Function), PACF (Partial Autocorrelation Function) were used. Additionally, the ARIMA model was utilized for forecasting and the effectiveness of the model was evaluated with residual analysis and the Ljung-Box test. The findings show that gold prices have a significant long-term increasing trend, there is no seasonal pattern, and the model created is more effective in short-term forecasts.

**1. Introduction**

Gold is an important and safe investment tool in the financial markets and has an important place in the global economy. Affected by various factors such as inflation, economic recession, geopolitical tensions and fluctuations in global markets, gold has been considered a strategic asset not only for individual investors but also for central banks and international financial institutions. Therefore, analyzing gold prices is critical to better understanding the impact of economic trends and global events. This study aims to analyze and predict gold prices based on historical data.

This study examines gold prices between 2014 and 2024. The data set is recorded daily, providing sufficient detail to trace price trend and analyze price movements over time. Before starting the analysis process, the data set was reviewed, missing values ​​and outliers were identified, and these problems were addressed to ensure the data suitable. The aim of the study is to examine long-term trends and short-term fluctuations in gold prices and to make predictions about how price movements may take shape in the future. In this context, time series analysis methods were used.

This study is crucial for making investments decisions and helps minimize investment risk by predicting future movements in gold prices.

**1.1. Data Description**

The dataset is a time series dataset that includes daily gold prices from January 2014 to August 2024. The dataset contains the date, close, volume, open, high, and low columns. If we define them:

Date: Every trade day has a unique date noted.

Close: Gold's closing price on the relevant day.

Volume: Trade volume of gold on the relevant date.

Open: Gold’s opening price on relevant date

High: The highest price of gold recorded during the trading day.

Low: The lowest price of gold recorded during the trading day.

The dataset contains five numeric (close, volume, open, high, and low) and one time-index (date) and has 2529 observations. The time index allows the dataset to use time series analysis and models. Therefore, the dataset makes changes analyzable and future price changes predictable. The dataset contains some N/A inputs, and these should be cleaned before analysis.

**1.2. Exploratory Data Analysis**

There are six columns in the dataset. Summary statistics were extracted to examine these variables.

metin, ekran görüntüsü, yazı tipi, siyah içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 1- Summary Statistics***

After examining the R-generated summary, we discovered that the volume variable in our dataset had 18 missing values.

metin, yazı tipi, ekran görüntüsü, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 2- Standard Deviation and Range***

This R output includes the standard deviation and range of values ​​for each variable. The range is obtained by the difference between the max and min values ​​in *Figure 1* for each variable. In general, Figure 2 shows similar ranges for the closing, opening, high, and low prices. The Volume variable has a much higher standard deviation and range than the others, indicating that extreme situations in this variable should be considered for market analysis.

ekran görüntüsü, metin, dikdörtgen, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 3 - Frequency Histogram by Year***

Between 2015 and 2023, the dataset consistently contains between 250 and 261 entries each year. However, the years 2014 and 2024 have fewer records than this range. Since data entry for 2014 began in August, analyses for that year must be done with caution. Additionally, it is important to note that data entry for 2024 is still in progress and should be incorporated into the analysis.

metin, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 4 – Time Series Plot of Gold Closing Prices***

The graph reveals a clear upward trend. The increase relatively modest until 2020, but there has been a sharp and significant rise since then. There is also noticeable volatility in the prices after 2020. (In this case, our response variable is the Close/last variable.)

çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, metin, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 5 – Time Series Plot of Opening and Closing Prices of Gold***

When analyzing the opening and closing prices together, it is evident that both have increasing trend. These variables are closely following each other, with small differences between them. This shows that fluctuations in opening and closing prices are minimal and generally in balance end of the day.

öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi, ekran görüntüsü, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 6 – Time Series Plot of High and Low Prices of Gold***

In this graph, the daily high and low gold prices are compared. Both variables illustrate an upward trend, like the observed in the opening and closing prices. Although the difference between them is typically small, it has become more wider especially after 2020 period. This rise may be related to economic and global.

öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi, metin, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 7 – Time Series Plot with all Features***

The graph indicates that there was a slow increase in prices between 2014-2020 years. After 2020, a rapid increase is observed. If analyzing these two separate periods, the period after 2020 can be described as a time when price difference is larger, and prices fluctuate more sharply. Compared to the period after 2020, the years between 2014 and 2020 can be considered a more stable period.

öykü gelişim çizgisi; kumpas; grafiğini çıkarma, ekran görüntüsü, diyagram, mor içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 8 – Plot of Highlighting Volatility***

This graph illustrates the difference between the highest and lowest gold prices over time. Understanding how this difference fluctuates can provide important insights into market volatility. Since 2020, there have been significant instances where the daily price range occasionally exceed 100 units. After 2022, these fluctuations have become more balanced.

metin, çizgi, diyagram, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 9 – Decomposition of Gold Prices***

The observed panel represents the raw observations of the response variable (Closing Price). Overall, the graph shows an increase trend, but there are fluctuations from time to time (especially after 2020). From the trend panel, it can be interpreted that gold prices have risen steadily over the long term with a notable acceleration in this increase occurring after 2020. In the seasonality panel, it can be observed regular and recurring waves in gold prices throughout the year.

***metin, yazı tipi, ekran görüntüsü, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu***

***Figure 10 – Correlation Matrix***

There is a strong correlation between the opening, high, and lowt prices of the day and the closing prices of gold, as seen in the matrix. The market tends to move consistently among these criteria. On the other hand, the low correlation of the volume variable with other variables indicates that trading volume has almost no or limited effect on prices. Also, these impacts can be observed by correlation heatmap matrix:

ekran görüntüsü, metin, diyagram, dikdörtgen içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 11 – Correlation Matrix Heatmap***

Only missing values were observed in the dataset’s volume variable. These values are significant because they impact the accuracy and reliability of the analysis. During the detection phase, these values were identified using R, and the total number of missing entries was determined. To address the problem, the median value of volume variable was used to replace the N/A values. The version of final check:



***Figure 12 – Missing Values Checking***

Detecting outliers in the dataset is important, as they can cause significant deviations and misleading results in the analysis. To identify these extreme values, all variables must be numerical format. In this dataset, the volume variable was not numeric, so it was covered to the appropriate format. The Interquartile Range (IQR) method was used to detect outliers, allowing for the calculation of these values for each variable.

metin, ekran görüntüsü, yazı tipi, siyah içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 13 – Number of Outliers***

Although no extreme values were found in price variables, 79 extreme values in volume variable were identified.

**1.3 Research Questions**

**1)** Is there a significant long-term trend or consistent seasonal pattern in the gold prices?

**2)** How may the gold prices potentially rise in the future?

**3)** How effective is the created model in long-term predictions compared to short-term predictions?

**1.4 Aim of the Study**

The aim of the study to analyze past trends, seasonality and random components of data by focusing on response which is closing prices of gold. The goal is to predict future price movements. To achieve this, the ARIMA model, one of the methods of time series analysis, was used. This model generates both short-term and long-term price forecasts based on past observations. Additionally, the study aims to predict future trends in gold prices. The insights from this study are designed to guide investors for understanding future price movements and enhancing to financial data analysis research.

**2. Methodology / Analysis**

In this study, the ARIMA model, which uses part observations, was used to predict future gold prices. Initially, the trends of all variables, especially the response variable, were analyzed over time. Missing values in data set were determined, and they were covered with the median value of the respective variable. Also, outliers in data set were defined but were nor removed from the model to protect the integrity of the data.

To ensure the data was suitable for ARIMA model, the closing gold price was selected as the response variable. The Augmented Dickey-Fuller (ADF) test was used to check stationarity of the time series. Based on the result of the test, the first differencing method was applied to make the data stationary.

To create the ARIMA model, the model parameters were determined by plotting the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) graphs of the time series. These were generated by using historical gold prices in dataset. The auto.arima() function was used to confirm the accuracy of the created model. The purpose of the model is to predict the future values of the gold prices for the next 30 days. Prediction results are visualized and edited speeds are displayed.

In addition, a residual analysis was performed to test accuracy and suitability of the model, and Ljung-Box test was applied. To measure the model’s performance, RMSE (the Root Mean Square Error) and MAPE (the Mean Absolute Percentage Error) were examined. The R programming language was used for all methods and analyses in this study.

**3. Results and Findings**

çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, yazı tipi, el yazısı içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 14 – Time Series Plot of Closing Prices of Gold***

The same trend plot is found in the Exploratory Data Analysis section and has been concluded. As a clear conclusion, it can be observed that gold prices are in an increasing trend in the long term.

ADF test was applied to test whether the time series is stationary. (the null hypothesis: the time series is not stationary). Results:

metin, yazı tipi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 15 – ADF Test Result***

As can be seen in Figure 15, the p value is 0.6151 (p value > 0.05, then do not reject null hypothesis), which means that the time series is not stationary.

For ARIMA model, the time series must be stationary. The series can be made stationary by applying the differencing method. The ADF test should be applied again after differencing.

**metin, yazı tipi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu**

***Figure 16 – ADF Test Result of Differencing***

The ADF test results indicates that the series became stationary following the first differencing method.

To determine the parameters ( p (AR) and q (MA)) of ARIMA model, the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) graphs should be plotted by using differencing time series.

metin, çizgi, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 17 – ACF Plot***

ACF measures the correlation coefficient between the series and its past values. Also, it is used to determine the q (MA) parameter. When the graph is examined, except a few lags, most lags are within the blue line, meaning no autocorrelation. Based on this graph, the q value can be taken as 1.

**metin, diyagram, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu**

***Figure 18 – PACF Plot***

PACF offers a clearer view of the direct relation between variable and its previous values. Also, it is used to determine the p (AR) parameter. Based on this graph, the p value can be taken as 1.

The ARIMA model was created with parameters p=1, d=1,q=1, and the model’s output was generated by using R. The d value represents that first-order differencing to make the series stationary.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 19 – ARIMA (1,1,1)***

AR (1) = -0.0335 indicates a slight negative relationship between the current values and the previous value. MA (1) = -0.0067 suggests a slight negative relationship between the current residual and previous residual. Log Likelihood value is -10501.81, AIC value is 21009.62, and BIC value is 21027.13. Higher Log Likelihood and lower AIC, BIC values indicate better model performance compared to alternative models. RMSE is 15.4107 that reflects the magnitude of the model’s prediction errors. Lower RMSE values indicate that the model has better performance. MAPE is 0.6678 that represents percentage of errors and a value of less than 1% indicates that the model is highly successful.

The ARIMA model was established to predict the future values ​​of gold prices for 30 days. It was aimed to predict price movements based on past data. The prediction results were visualized to present the short-term projection of the results.

çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, metin, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 20 – Forecast Plot***

The Gold Price forecast chart shows the estimated movements of the closing price of gold over the next 30 days. The blue shaded area represents the 95% confidence interval of the predicted values, indicating the extent of uncertainty. The predictions appear to be in line with the current trend and the confidence interval remains narrow.

çizgi, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 21 – Forecast Plot***

This chart is a closer and more detailed view of the forecast plot. The predicted prices are expected to follow a horizontal trend, indicating that they will remain at their current levels without significant changes. The blue shaded are represents the potential range of changes in predicted prices.

Residual analysis and the Ljung-Box test were applied to test the performance of the model, evaluate the normal distribution assumption, assess its fit to the data, and check whether the residuals were random.

ekran görüntüsü, metin, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 22 – Residual Analysis***

In general, the random distribution of the residuals, the general absence of autocorrelation in the ACF plot, and the nearly normal distribution of the residuals, even though there are extreme values ​​in the histogram, suggest that the model created provides a good fit.

metin, yazı tipi, ekran görüntüsü, siyah içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 23 – Ljung-Box Test Result***

When the test results are examined, it is seen that the p-value is 0.2708, which is greater than 0.05. This indicates that the model's residuals do not exhibit autocorrelation.

As a conclusion, according to the results of residual analysis and Ljung-Box test, it can be said that the predictive performance of this model is reliable and suitable for the data.

**4. Discussion**

In this study, the ARIMA model was utilized to determine long-term trends and seasonality effects based on past data of gold prices and to predict future price movements. Residual analysis and the Ljung-Box test were conducted to evaluate the model’s performance. Using the established model and analyses conducted, it was determined that model offers higher accuracy in short term predictions, but uncertainty increases in long-term forecasts. The results, particularly for 30-day predictions, indicates constant upward trend in gold prices within a specified confidence interval. This finding suggests that the model is suitable tool for short-term investment decisions.

For long-term predictions, it may be beneficial to add economic indicators and macroeconomic factors into the model. Furthermore, deep learning models such as LSTM can be used in a more complex dataset that includes these elements.

In summary, this study employed time series analyses and the ARIMA model to predict future gold prices movements. These findings can also be used by investors and financial institutions in decision-making and investment processes.

**References**

GeeksforGeeks. (2023, November 22). *Autocorrelation and partial autocorrelation*. GeeksforGeeks. <https://www.geeksforgeeks.org/autocorrelation-and-partial-autocorrelation/>

Ng, T. (2023, November 1). Gold Price Forecasting with ARIMA - Timmy Ng - Medium. *Medium*. <https://medium.com/@timmyng83/gold-price-forecasting-with-arima-0d06aa99edd1>

Tombare, P. (2024). ANALYSIS OF RELATIONSHIP BETWEEN GOLD PRICE AND STOCK MARKET PRICE IN INDIA ANALYSIS OF RELATIONSHIP BETWEEN GOLD PRICE AND STOCK MARKET PR. In [*https://www.researchgate.net/publication/381739095\_ANALYSIS\_OF\_RELATIONSHIP\_BETWEEN\_GOLD\_PRICE\_AND\_STOCK\_MARKET\_PRICE\_IN\_INDIA\_ANALYSIS\_OF\_RELATIONSHIP\_BETWEEN\_GOLD\_PRICE\_AND\_STOCK\_MARKET\_PRICE\_IN\_INDIA*](https://www.researchgate.net/publication/381739095_ANALYSIS_OF_RELATIONSHIP_BETWEEN_GOLD_PRICE_AND_STOCK_MARKET_PRICE_IN_INDIA_ANALYSIS_OF_RELATIONSHIP_BETWEEN_GOLD_PRICE_AND_STOCK_MARKET_PRICE_IN_INDIA).

**Appendices**

**Appendix A: Data Preprocessing Code**

**metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü, yazılım, multimedya yazılımı içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü, yazılım, web sitesi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**Appendix B: ARIMA Model Code**

**metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü, yazılım içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu**