Forecasting Time Series of Precious Metals

Tatev Baghdasaryan

ASDS, YSU

02/06/2024

The dataset offers detailed, up-to-date information on precious metals futures. Futures are financial contracts obligating the buyer to purchase, and the seller to sell a particular precious metal (such as gold, silver, platinum, etc.) at a predetermined future date and price.

Here' s the link to the dataset:

Gold, Silver and Precious Metals Futures Daily Data

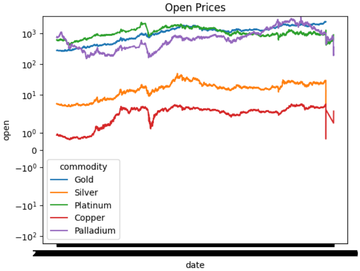
Precious metals have always played a pivotal role in economies worldwide, serving not only as currency, but also as a protection against economic ups and downs. This is the reason I found this topic quite interesting and chose it for my final project.

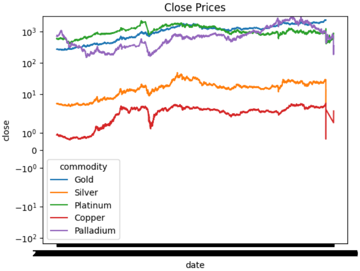
One of the things I've researched is this article: The Investopedia article on the history of gold prices details the significant highs and lows of gold prices over the years. Gold has historically been a safe-haven asset, seeing price increases during economic uncertainty and inflation. The article highlights key periods, such as the 1980 peak during high inflation and the 2011 peak amid global economic instability. It also covers the more recent trends, including the impact of the COVID-19 pandemic on gold prices. You can find more here: Gold Price History: Highs and Lows

Stock Open/Close Prices

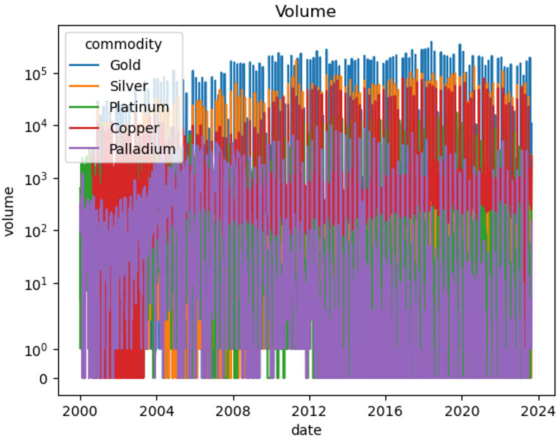
We can first investigate the trend of the stock open and close prices over time.

Observations: With the exception of Gold, the open/close prices for all the stocks have dips in their stock prices around the year 2008. We can speculate this is due to the 2008 financial crisis that occurred around that time.

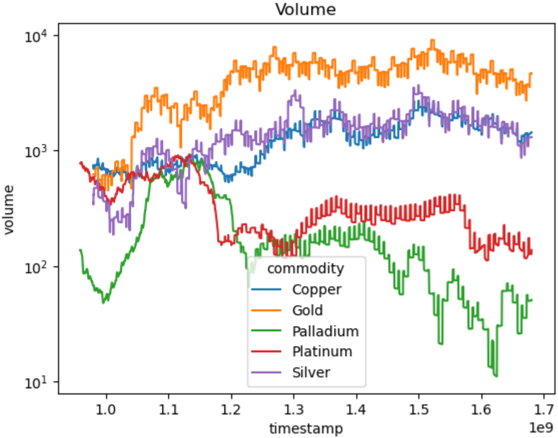




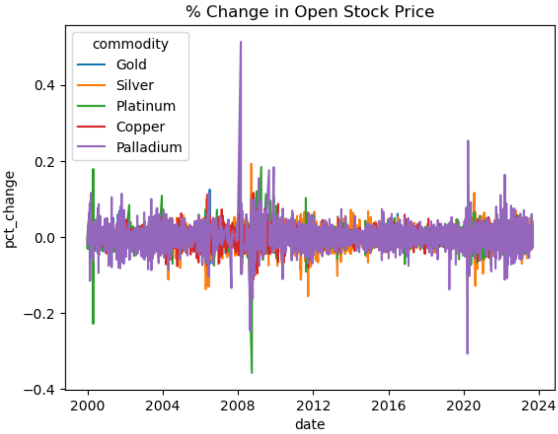
Stock Volume
We can also investigate the trends of the stock volumes using a similar analysis.



We see that there is a lot of variability in the data. So, instead, something we could do is plot the moving average of the data.



Now we can investigate the trends of the percentage change over time:



Stationary test is a vital step in analyzing time series, so let's go on and analyze the results obtained from the Stationary Test.
We observe that the p-values of all the tests are greater than 0.05, which signifies that the time series data have unit roots, and so are non-stationary.

Our goal will be to predict the (open) stock prices of the different commodities in 2023 using the previous stock data. To do this, we will need to add some features to our dataset before passing it into modeling:

- We should add extra features that incorporate more information about the date variate; for example, the day, week, month, day of week, etc.
- We should add columns for the open, close, low and high prices for each commodity in our dataset.

Listing 1: Python code for adding datetime specific features

```python
df["timestamp"] = df["date"].apply(lambda d: d.
timestamp())
df["day"] = df["date"].dt.day
df["week"] = df["date"].dt.isocalendar().week
df["month"] = df["date"].dt.month
df["year"] = df["date"].dt.year
df["day_of_week"] = df["date"].dt.dayofweek

df.sample(5)
```

To improve modeling, we should make the series stationary. To do this, we will:

- Apply a log transformation to the open stock price
- Compute the differences between the log open stock prices

Listing 2: Python code for a log transformation

```
df["log_open"] = np.log(df["open"])
for comm in commodities:
    data = df.loc[df["commodity"] == comm, "log_open"]
    df.loc[df["commodity"] == comm,
                                "diff_log_open"] = data.diff()
    df.loc[(df["commodity"] == comm).idxmax(),
                               "diff_log_open"] = 0
df.head()
```

Now the data is stationary. The p-values are now extremely small (basically zero), which means this transformed variate is now stationary.

Optimal Lag Selection, Handling Insignificant Variables, Model Stability Check, Testing Residual Autocorrelation

```python
# Fit the VAR model
model = VAR(commodity_close_prices_diff)
lag_selection = model.select_order(maxlags=15)

# Get the optimal lag length based on AIC
optimal_lag_aic = lag_selection.aic

# Fit the VAR model using the optimal lag length
var_model = model.fit(optimal_lag_aic)

# Display the summary of the VAR model
var_model_summary = var_model.summary()
print(var_model_summary)

# Check model stability
is_stable = var_model.is_stable()

# Calculate Durbin-Watson statistics for each equation
dw_stats = durbin_watson(var_model.resid)

# Perform Ljung-Box test for residual autocorrelation for each time series residuals
lb_test_results = {}
for col in var_model.resid.columns:
    lb_test_results[col] = acorr_ljungbox(var_model.resid[col], lags=[optimal_lag_aic], return_df=True)
```

Summary of Regression Results

```
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
Date:           Sat, 01, Jun, 2024
Time:                     14:34:47
--------------------------------------------------------------------
No. of Equations:         2.00000    BIC:                    2.54489
Nobs:                     5930.00    HQIC:                   2.49926
Log likelihood:          -24104.9    FPE:                    11.8813
AIC:                      2.47496    Det(Omega_mle):         11.7580
```

Results for equation Gold

Results for equation Gold

	coefficient	std. error	t-stat	prob
const	0.372667	0.172884	2.156	0.031
L1.Gold	-0.021983	0.020291	-1.083	0.279
L1.Silver	0.204514	0.662132	0.309	0.757
L2.Gold	0.008988	0.020290	0.443	0.658
L2.Silver	-0.617417	0.661550	-0.933	0.351
L3.Gold	0.046106	0.020321	2.269	0.023
L3.Silver	-1.172886	0.661832	-1.772	0.076
L4.Gold	-0.028468	0.020321	-1.401	0.161
L4.Silver	0.347683	0.661633	0.525	0.599
L5.Gold	0.010653	0.020315	0.524	0.600
L5.Silver	-0.009064	0.661555	-0.014	0.989
L6.Gold	-0.018422	0.020302	-0.907	0.364
L6.Silver	-0.845509	0.660999	-1.279	0.201
L7.Gold	-0.066422	0.020279	-3.275	0.001
L7.Silver	1.699251	0.661007	2.571	0.010

Results for equation Silver

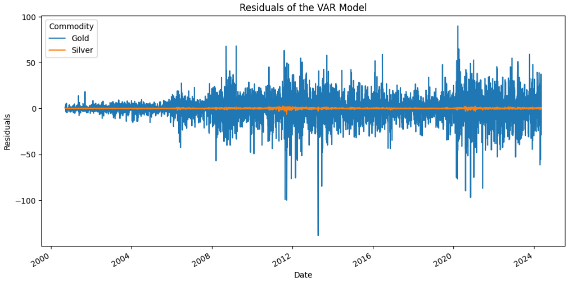
Results for equation Silver

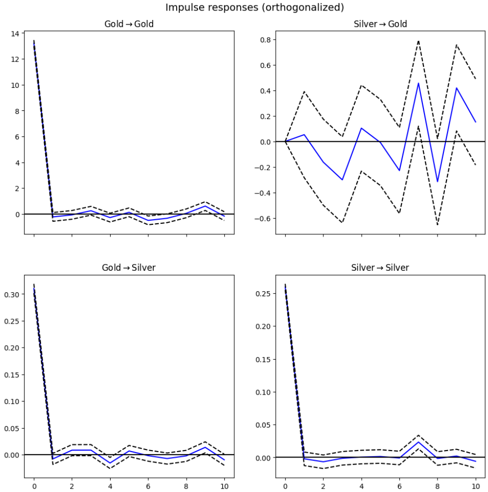
	coefficient	std. error	t-stat	prob
const	0.004591	0.005293	0.867	0.386
L1.Gold	-0.000412	0.000621	-0.663	0.507
L1.Silver	-0.008151	0.020272	-0.402	0.688
L2.Gold	0.001218	0.000621	1.960	0.050
L2.Silver	-0.025528	0.020254	-1.260	0.208
L3.Gold	0.000788	0.000622	1.267	0.205
L3.Silver	-0.006102	0.020263	-0.301	0.763
L4.Gold	-0.001171	0.000622	-1.882	0.060
L4.Silver	0.001467	0.020257	0.072	0.942
L5.Gold	0.000305	0.000622	0.490	0.624
L5.Silver	0.007445	0.020254	0.368	0.713
L6.Gold	-0.000060	0.000622	-0.097	0.923
L6.Silver	-0.003743	0.020237	-0.185	0.853
L7.Gold	-0.002623	0.000621	-4.225	0.000
L7.Silver	0.088567	0.020237	4.376	0.000

Correlation Matrix of Residuals and Stability Results

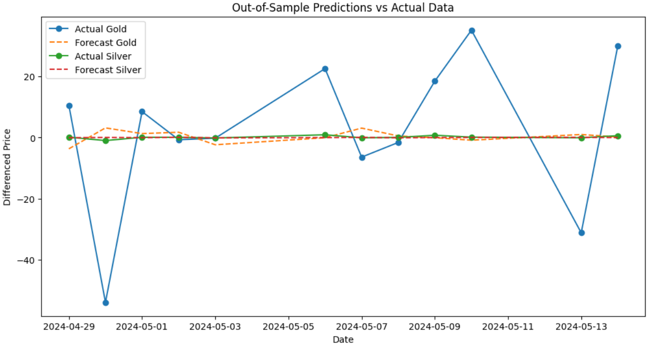
```
Correlation matrix of residuals
              Gold    Silver
Gold      1.000000  0.766905
Silver    0.766905  1.000000



Stability Results:
{'Model Stability': True, 'Durbin-Watson': array([1.99815208, 2.00005332]), 'Ljung-Box Test': {'Gold':      lb_stat  lb_pvalue
15  0.223032        1.0, 'Silver':      lb_stat  lb_pvalue
15  0.113422        1.0}}
```

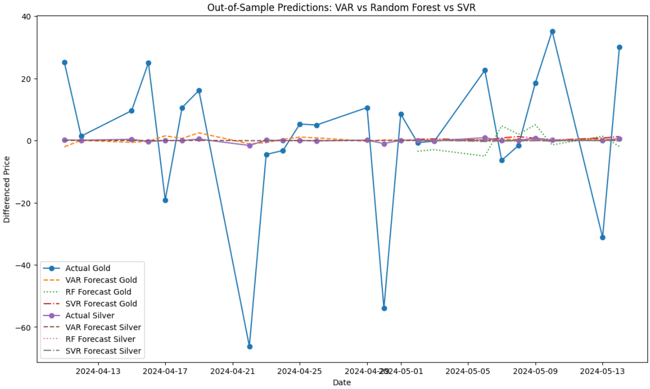
Residuals of VAR

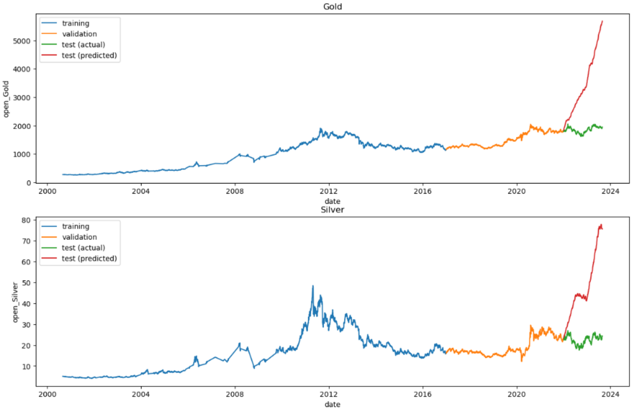


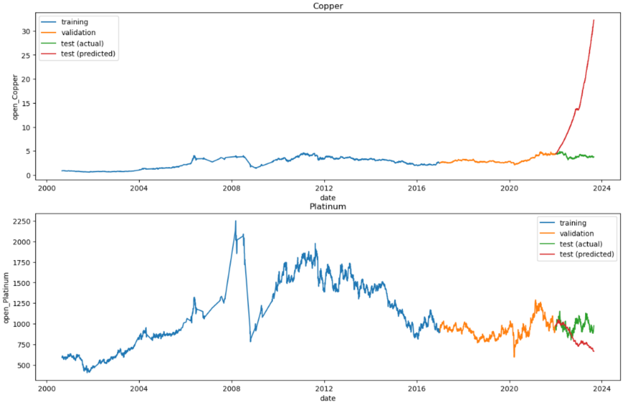


Impulse responses (orthogonalized)
Gold → Gold
Silver → Gold
Gold → Silver
Silver → Silver


Out-of-Sample Predictions vs Actual Data
Actual Gold
Forecast Gold
Actual Silver
Forecast Silver
20
0
−20
−40
Differenced Price
2024-04-29
2024-05-01
2024-05-03
2024-05-05
2024-05-07
2024-05-09
2024-05-11
2024-05-13
Date


Out-of-Sample Predictions: VAR vs Random Forest vs SVR
40
20
0
−20
−40
−60
Differenced Price
2024-04-13
2024-04-17
2024-04-21
2024-04-25
2024-04-29
2024-05-01
2024-05-05
2024-05-09
2024-05-13
Date
Actual Gold
VAR Forecast Gold
RF Forecast Gold
SVR Forecast Gold
Actual Silver
VAR Forecast Silver
RF Forecast Silver
SVR Forecast Silver


Gold
training
validation
test (actual)
test (predicted)
open_Gold
0
1000
2000
3000
4000
5000
2000
2004
2008
2012
2016
2020
2024
date
Silver
training
validation
test (actual)
test (predicted)
open_Silver
10
20
30
40
50
60
70
80
2000
2004
2008
2012
2016
2020
2024
date


Copper
training
validation
test (actual)
test (predicted)
open_Copper
0
5
10
15
20
25
30
2000
2004
2008
2012
2016
2020
2024
date
Platinum
training
validation
test (actual)
test (predicted)
open_Platinum
500
750
1000
1250
1500
1750
2000
2250
2000
2004
2008
2012
2016
2020
2024
date

Appendix: As we observed, gold prices are really sensitive to some unknown factors. The standard view is that S&P 500 and gold markets are negatively linked: when the stocks go up, gold prices decrease, and vice versa. This might be because gold is a safe haven, so traders sometimes may prefer gold to relatively risky stocks (S&P 500 - green line, Gold - yellow line).