Gold and Silver Price Analysis and Prediction

By Syed Mehran and Aidan Olson

1. Introduction

Gold and silver have been important commodities throughout history. These commodities have been used as currencies for thousands of years. Egyptians began producing gold-bearing shekels around 1500 B.C. and silver coins started appearing in the same areas several hundred years later.

The Founding Fathers of the United States also used gold to stabilize the young USD. The Coinage Act of 1792 was the first law in writing to set an exchange rate for the USD. The act specifically declared that 15 pounds of silver was equal to 1 pound of gold. Additionally, it established that each dollar was equivalent to 0.057 ounces of fine gold or 0.85 ounces of fine silver[1].

Given the historical significance of gold, the researchers have decided to develop a predictive model to forecast the price of gold.

Datasets

The Gold & Silver datasets provide 10 years (2013-2023) of information about the commodity price and history such as:

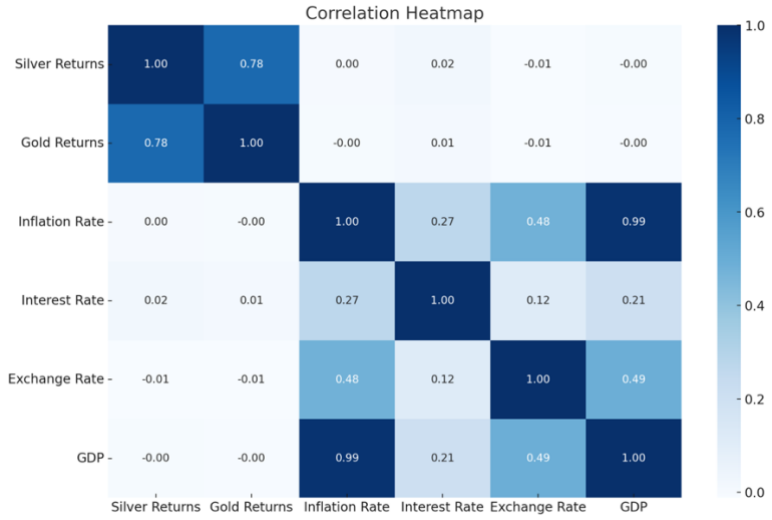
1) Date – The date of data collected
2) Close/Last ($) – The price the commodity closed at during market hours
3) Volume – Total number of trades with commodity
4) High ($) – The highest price the commodity achieved in each day
5) Low ($) – The lowest price the commodity experienced in each day

However, due to the COVID-19 pandemic, which had unpredictable and massive implications on the markets, we will snip the data to end a few months prior to the pandemic. This 'cut-off point' will help to develop more accurate models since they may not encounter as much noise within the datasets.

Other datasets we are bringing in include IEF (a bond) data, S&P 500 Index data, Consumer Price Index (CPI), GDP, inflation rates, USD-EUR conversion rates and 10 Year Treasury Note prices.

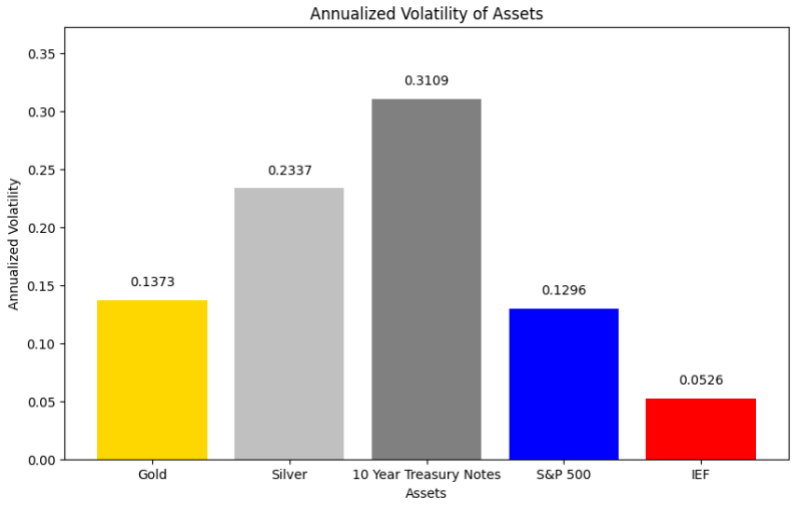
2. Exploratory Data Analysis

Correlation Analysis



After successfully completing the Extract, Transform, Load process of the data science pipeline, researchers delved into Exploratory Data Analysis. To begin, correlations between different economic indicators and commodity returns were tested. While gold and silver did not share much correlation with these economic indicators, it is notable that they had relatively high correlations with one another. Therefore, it's important to note that the price of gold and silver are correlated when one is observing their daily returns. This will be important when we are customizing our model.

Risk Analysis



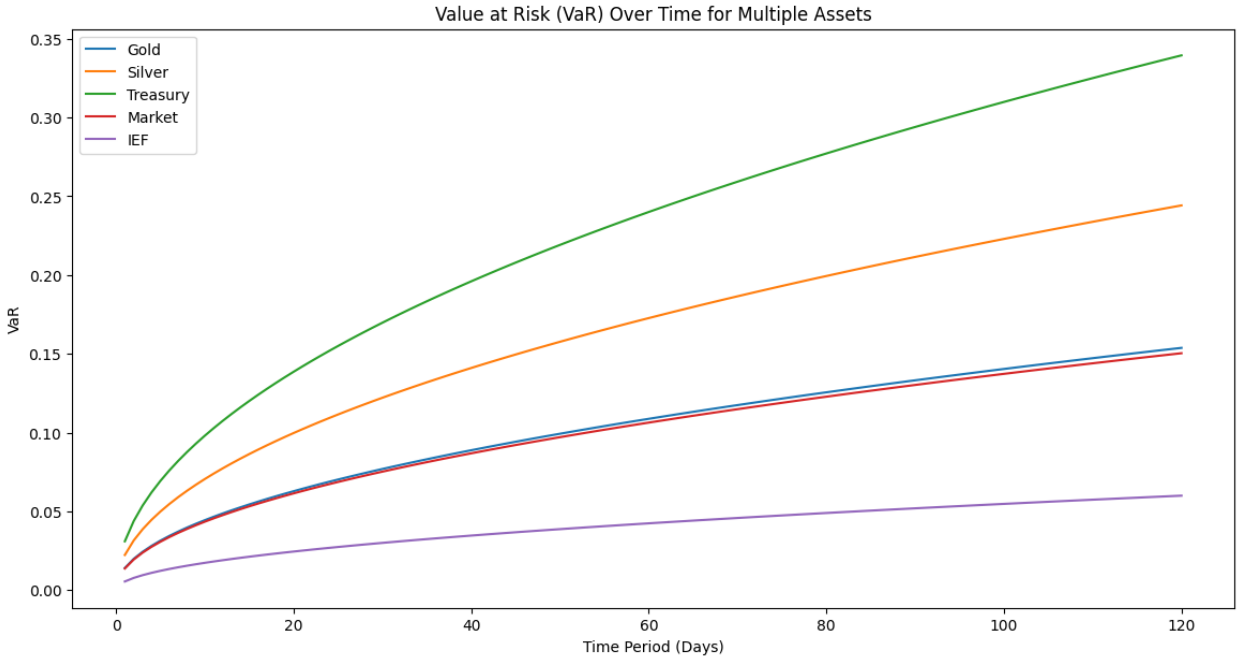
When making informed investment decisions, considering risk is crucial. The above plot shows the annualized volatility of our selected assets. Annualized volatility is calculated as follows:

$$\sigma_{\text{annualized}} = \sigma_{\text{daily}} \times \sqrt{252}$$

Where σ_{daily} is the standard deviation of daily returns.

Utilizing annualized volatility, we see that gold is a relatively safe investment when compared to silver and the 10Y Treasury note. Gold has a similar annualized volatility to the market. Silver is the 2nd highest volatile asset in comparison to the others. We see that IEF is the safest investment in terms of annualized volatility.

Value at Risk (VaR)

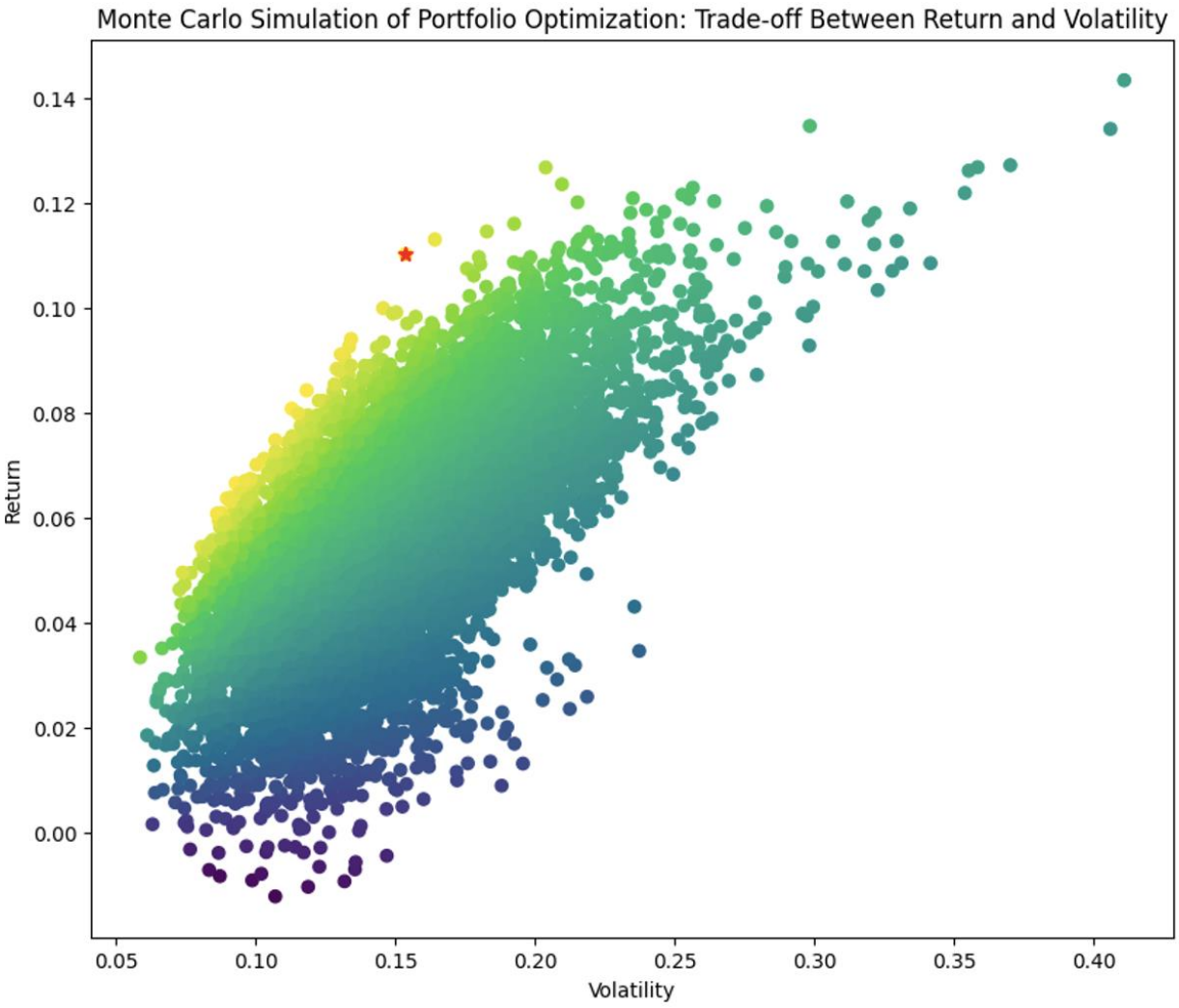


Value at Risk (VaR) is a widely used risk management tool that measures the potential loss in value of a risky asset or portfolio over a defined period for a given confidence interval. The VaR for each asset increases over time, which is expected because the uncertainty and potential loss in value typically increase with the length of the time horizon. The Market and Silver show a higher level of risk, as indicated by their higher VaR values. Silver shows a more pronounced increase in VaR over time, suggesting that it may be more volatile or have a higher risk of loss over a

longer time horizon. Gold's VaR is near that of the market's. It starts very close to the lower-risk assets but diverges as the time period increases, indicating its risk increases at a faster rate than IEF and the market, but not as fast as Silver.

3. Sophisticated Data Analysis

Monte Carlo



We did a Monte Carlo simulation to help determine the best portfolio composition when spreading wealth between the 10-year treasury, the S&P 500, bonds, silver, and gold. It determines the highest ratio of return to volatility, thus guaranteeing the highest chance that your investments return the best yield per unit. This is the best wealth distribution within a portfolio:

```
Weights of the Best Portfolio:
Silver: 6.57%
Treasury: 1.49%
S&P 500: 44.32%
IEF: 46.77%
Gold: 0.85%
```

These results indicate that investing in gold and silver is not worth the risk and that we're better off investing in the market and bonds.

4. Model Selection

Complications and Mistakes

In an act of poor foresight, we trained our models on pre-pandemic data (2013-2020) and then tested them on post-pandemic data (2020-2023). Having COVID-19 as the breaking point between our training and testing data messed up our model significantly. Since our model can't predict for global events, the test data was awkward. To fix this, we changed the date domain to 2013-2020 and adjusted our training and test data accordingly. After this change, we saw a 30-40% decrease in training RMSE and 60-80% decrease in test RMSE.

Ordinary Least Squares (OLS)

The beginning of our model exploration was the standard linear regression model (OLS in the sklearn python library). For our two models, we had four independent variables and one response variable. Since gold and silver correlated so significantly in our correlation analysis, we made sure to include them in each other's OLS model. The other three independent variables were bond prices from the IEF data set, the S&P 500 market value, and 10-year treasury bond data. This model proved to be quite accurate. We saw a 1.07 training RMSE and 4.68 test RMSE on silver predictions and a 29.12 training RMSE and 86 test RMSE on gold. Considering that gold is worth far more than silver, these values make sense and represent great margins of error for our OLS model.

ARIMA

This model stands for Auto Regressive Integrated Moving Average. This model is univariate, so it only considers prior gold closing prices. When we included data that was affected by the COVID-19 pandemic, we saw a gold test RMSE of 628**.** Excluding data affected by COVID lowered our gold test RMSE to 88. Silver saw similar drops in test RMSE as well. Excluding data due to COVID-19 greatly increased our accuracy.

5. Conclusion

In this project, we successfully used datasets of various economic measurements to create a model that predicts the price of gold and silver. This project helped us to understand how precious metals are affected by the economy. Additionally, we learned about the volatility of these metals and that, while they may be lucrative, they are not very reliable investments.