Predicting Gold from the S&P 500 Index

Shane McCallum

Data Science Career Track Capstone Project,
February 25, 2021

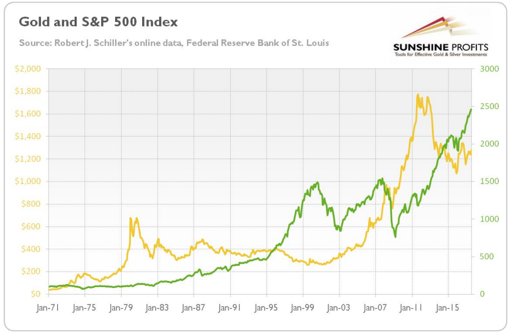
Problem and Task

An age-old debate:

- Can Gold (GLD) be accurately predicted using the S&P 500 Index (SPX)?
- Is this relationship unique to GLD & SPX? What about Silver(SLV) & SPX?

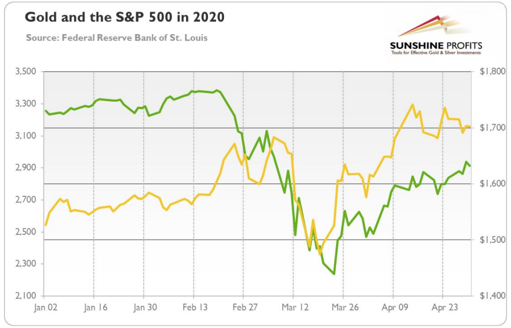
Solution:

- Test for cointegration & causality; if significant, then:
- Develop accurate Time Series model for predictions.



Gold and S&P 500 Index

Source: Robert J. Schiller's online data, Federal Reserve Bank of St. Louis



Gold and the S&P 500 in 2020

Source: Federal Reserve Bank of St. Louis

"S&P 500." *Precious Metals Investment Terms A to Z*, Sunshine Profits, www.sunshineprofits.com/gold-silver/dictionary/gold-sp. Accessed 25 Feb. 2021.

Who cares?

- Mutual Funds
- Retail Investors
- Market Analysts
- Financial Institutions

Northwestern Mutual®

Robinhood

JPMorgan Chase & Co.

M1 Finance

Ipsos

Webull

PROACTIVE WORLDWIDE®

Vanguard®

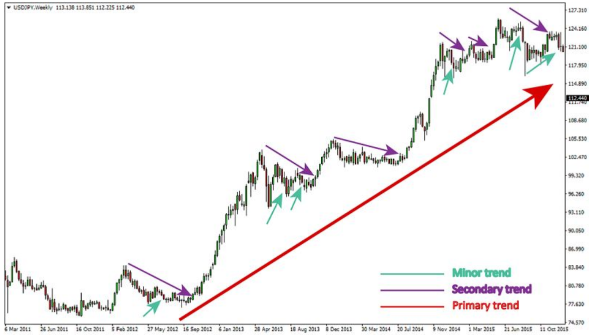
Forethought

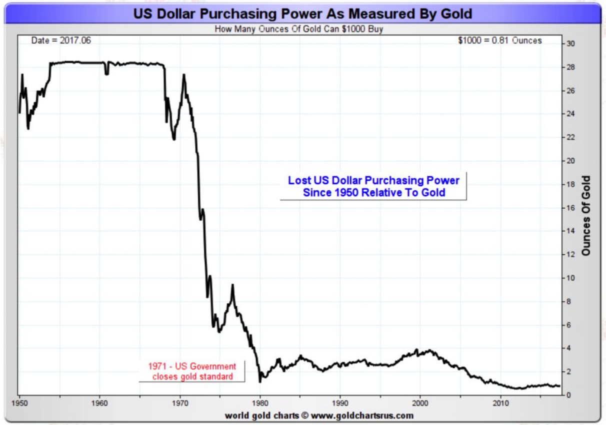
Trends:

- Always changing
- At least 18 weeks long

Gold history:

- Gold Standard no longer used
- USD is now a fiat currency





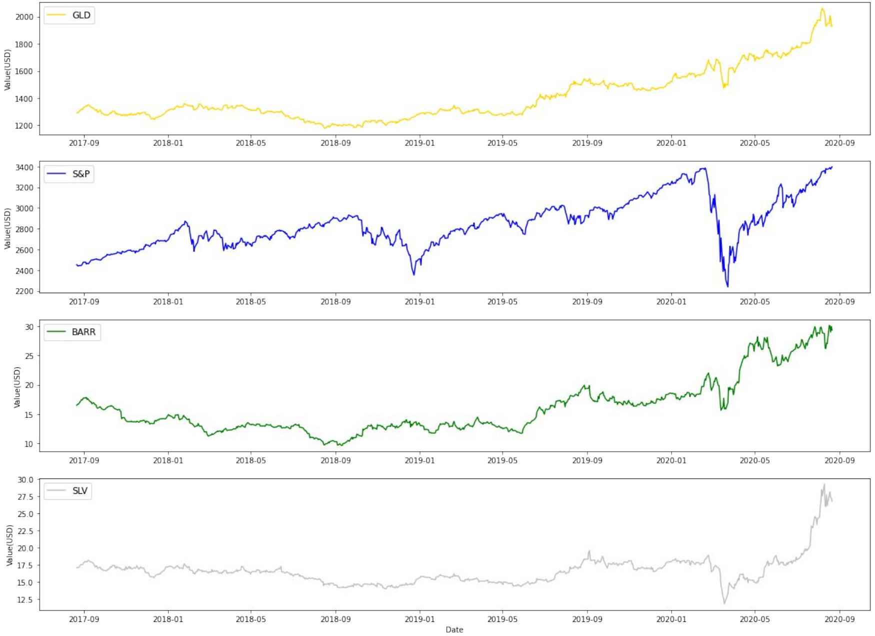
Data Wrangling

Sources:

- Yahoo! Finance
- MacroTrends
- MarketWatch

Time Frame:

- August 21, 2017 - August 21, 2020
- Test data will be > 18 weeks

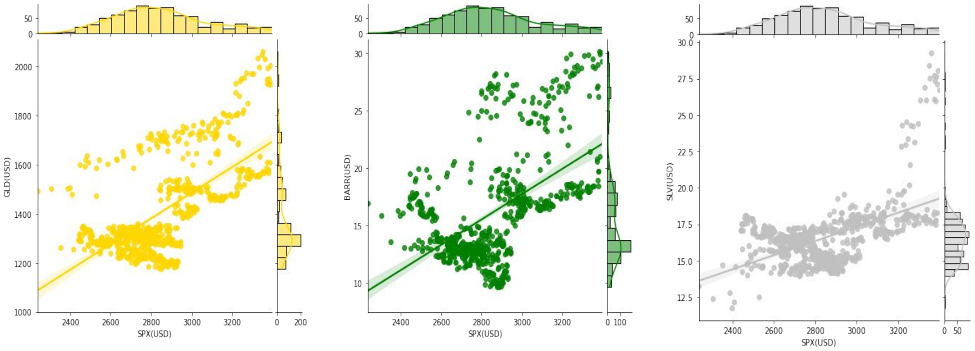


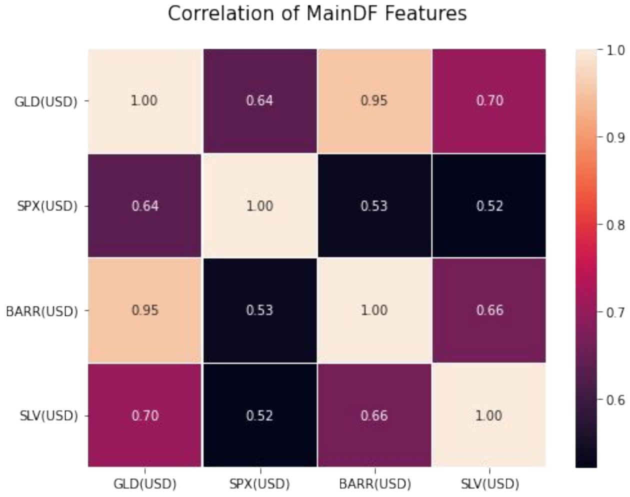
Data Exploration

Correlation:

- Seaborn Jointplot
  - Shows us positive correlation between GLD & SPX

- Seaborn Heatmap
  - GLD & SPX lacking significant correlation





Data Exploration (cont.)

Cointegration:

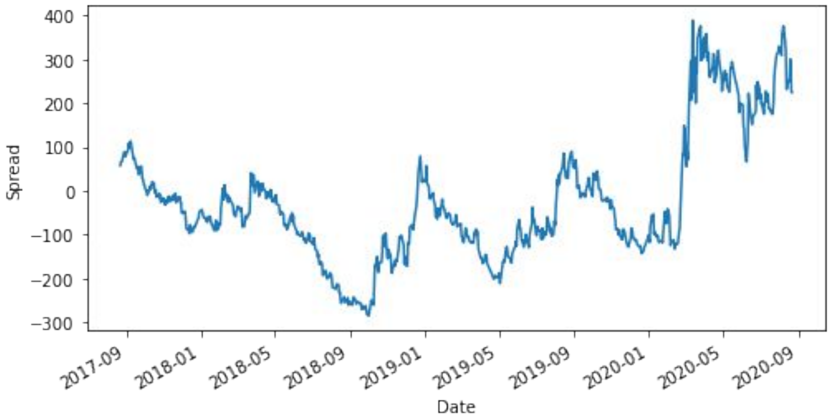
- Identifies if two series are sensitive to the same average price over a specific period of time.
- Used for hedging.

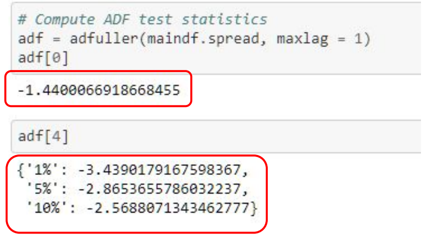
CADF Test:

- -1.44 > -2.865, not significant.

Johansen Test:

- 7.61 < 14.26, not significant.



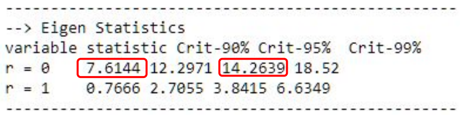


```
# Compute ADF test statistics
adf = adfuller(maindf.spread, maxlag = 1)
adf[0]
```

```
-1.440066918668455
```

```
adf[4]
```

```
{'1%': -3.4390179167598367,
 '5%': -2.8653655786032237,
 '10%': -2.5688071343462777}
```

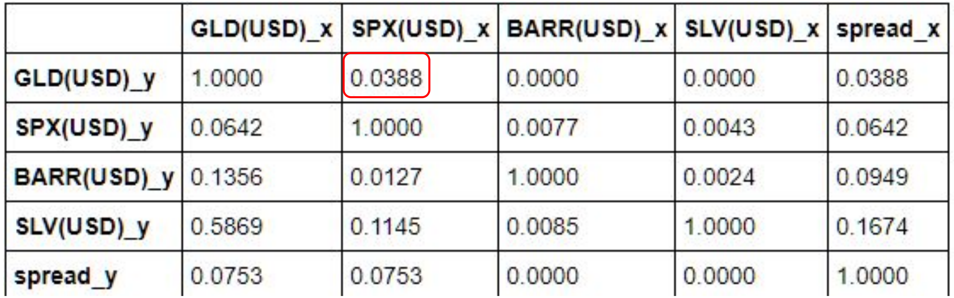


```
----------------------------------------------------------
--> Eigen Statistics
variable statistic Crit-90% Crit-95%  Crit-99%
r = 0     7.6144 12.2971 14.2639 18.52
r = 1     0.7666 2.7055 3.8415 6.6349
----------------------------------------------------------
```

Data Exploration (cont.)

Granger Causality Test:

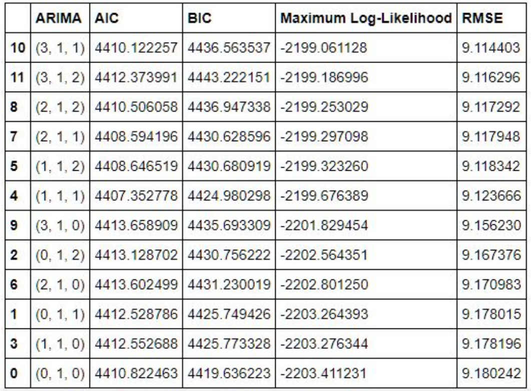
- Tests time series to see if they are useful in forecasting (causal relationship) other time series
- SPX has a significant causal relationship to GLD, 3.88% < 5.00%



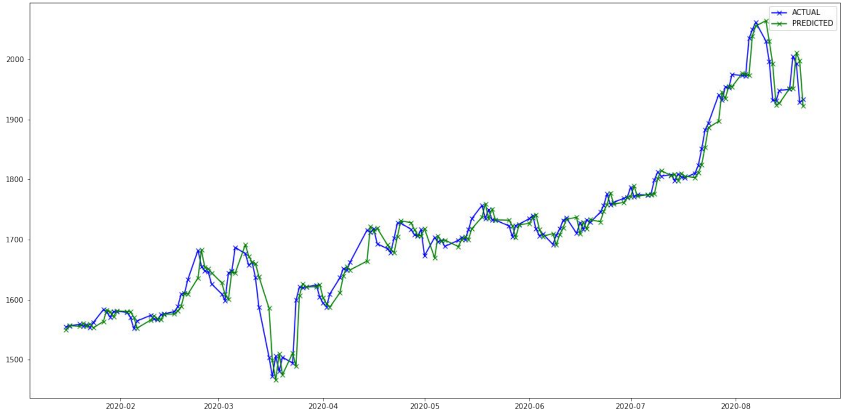
	GLD(USD)_x	SPX(USD)_x	BARR(USD)_x	SLV(USD)_x	spread_x
GLD(USD)_y	1.0000	0.0388	0.0000	0.0000	0.0388
SPX(USD)_y	0.0642	1.0000	0.0077	0.0043	0.0642
BARR(USD)_y	0.1356	0.0127	1.0000	0.0024	0.0949
SLV(USD)_y	0.5869	0.1145	0.0085	1.0000	0.1674
spread_y	0.0753	0.0753	0.0000	0.0000	1.0000

ARIMA Model

- Predicts future GLD value from past GLD value.
- Uses ‘p, d, q’ values for predicting
- Cross-validation model
  - p, d, q =2, 1, 1
- Measures of Accuracy:
  - RMSE: $22.84
  - MAPE: 8.374%, ~92% accuracy

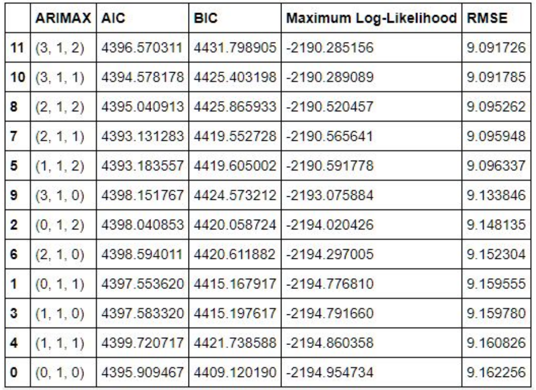


	ARIMA	AIC	BIC	Maximum Log-Likelihood	RMSE
10	(3, 1, 1)	4410.122257	4436.563537	-2199.061128	9.114403
11	(3, 1, 2)	4412.373991	4443.222151	-2199.186996	9.116296
8	(2, 1, 2)	4410.506058	4436.947338	-2199.253029	9.117292
7	(2, 1, 1)	4408.594196	4430.628596	-2199.297098	9.117948
5	(1, 1, 2)	4408.646519	4430.680919	-2199.323260	9.118342
4	(1, 1, 1)	4407.352778	4424.980298	-2199.676389	9.123666
9	(3, 1, 0)	4413.658909	4435.693309	-2201.829454	9.156230
2	(0, 1, 2)	4413.128702	4430.756222	-2202.564351	9.167376
6	(2, 1, 0)	4413.602499	4431.230019	-2202.801250	9.170983
1	(0, 1, 1)	4412.528786	4425.749426	-2203.264393	9.178015
3	(1, 1, 0)	4412.552688	4425.773328	-2203.276344	9.178196
0	(0, 1, 0)	4410.822463	4419.636223	-2203.411231	9.180242

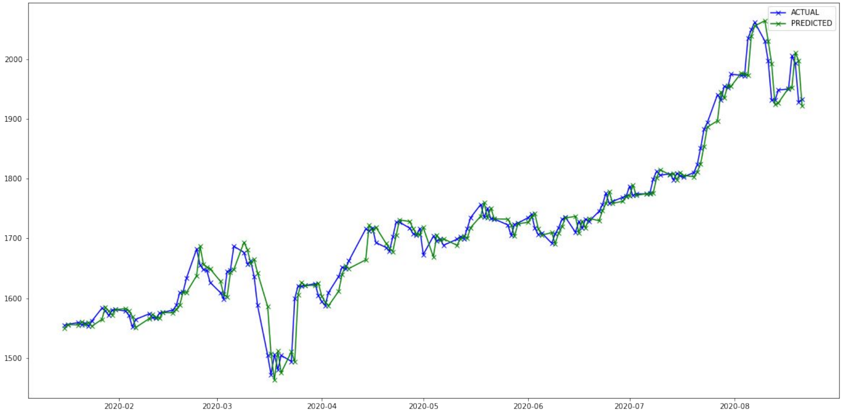


ARIMAX Model

- Predicts future GLD value from past GLD value and SPX value.
- Cross-validation model
    - p, d, q =2, 1, 1, nice!
- Measures of Accuracy:
    - RMSE: $22.95, higher
    - MAPE: 8.369%, ~92% accuracy, 0.5% better.



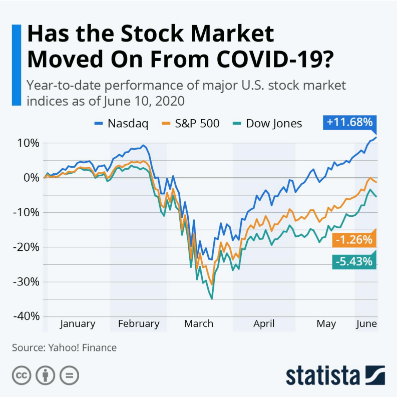
	ARIMAX	AIC	BIC	Maximum Log-Likelihood	RMSE
11	(3, 1, 2)	4396.570311	4431.798905	-2190.285156	9.091726
10	(3, 1, 1)	4394.578178	4425.403198	-2190.289089	9.091785
8	(2, 1, 2)	4395.040913	4425.865933	-2190.520457	9.095262
7	(2, 1, 1)	4393.131283	4419.552728	-2190.565641	9.095948
5	(1, 1, 2)	4393.183557	4419.605002	-2190.591778	9.096337
9	(3, 1, 0)	4398.151767	4424.573212	-2193.075884	9.133846
2	(0, 1, 2)	4398.040853	4420.058724	-2194.020426	9.148135
6	(2, 1, 0)	4398.594011	4420.611882	-2194.297005	9.152304
1	(0, 1, 1)	4397.553620	4415.167917	-2194.776810	9.159555
3	(1, 1, 0)	4397.583320	4415.197617	-2194.791660	9.159780
4	(1, 1, 1)	4399.720717	4421.738588	-2194.860358	9.160826
0	(0, 1, 0)	4395.909467	4409.120190	-2194.954734	9.162256



Summary

Concluding points:

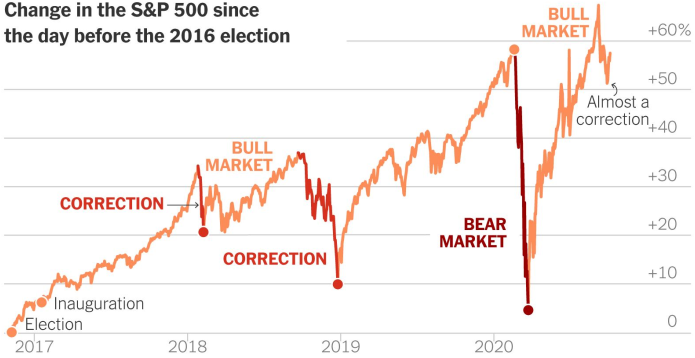
- Hypothesis is true, however:
  - Not much better than a standard & simpler ARIMA model.
- Reasons:
  - Gold Standard gone
  - Current Market Trends
  - Current USD value



Has the Stock Market Moved On From COVID-19?

Year-to-date performance of major U.S. stock market indices as of June 10, 2020

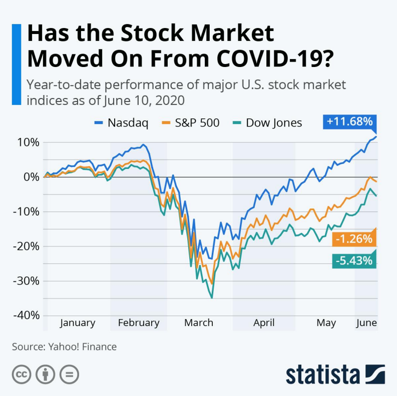
Source: Yahoo! Finance



Change in the S&P 500 since the day before the 2016 election

Future Test

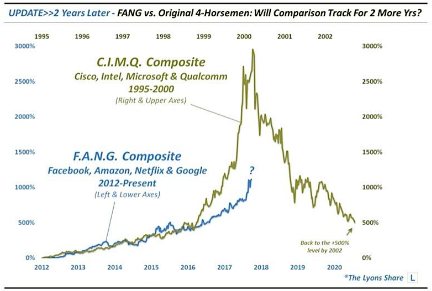
- Attempt again after market becomes more stable.
- Current Market is a, ‘Tech Bubble.’
- Consider using Barrick Gold Mining, Corp.
    - Causal relationship with GLD and SPX



Has the Stock Market Moved On From COVID-19?

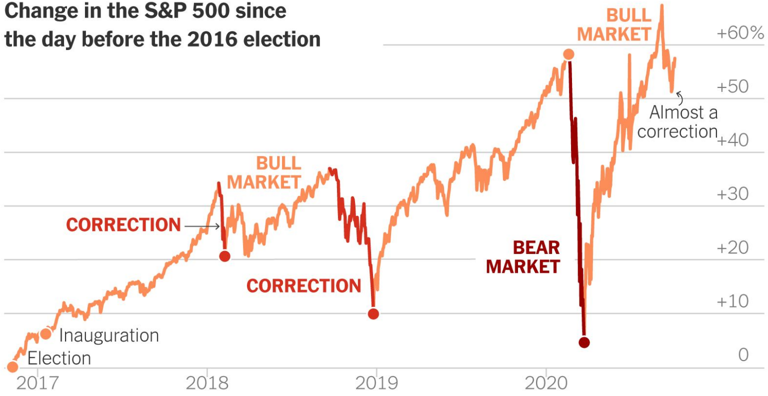
Year-to-date performance of major U.S. stock market indices as of June 10, 2020

Source: Yahoo! Finance



UPDATE>>2 Years Later - FANG vs. Original 4-Horsemen: Will Comparison Track For 2 More Yrs?

©The Lyons Share



Change in the S&P 500 since the day before the 2016 election

Thank you!

Shane McCallum,
Data Scientist and Sociologist

Contact Information:
Email: McCallum.D.Shane@gmail.com
LinkedIn: https://www.linkedin.com/in/shane-mccallum/
GitHub: https://github.com/Shane-McCallum

Project Report: https://github.com/Shane-McCallum/ARIMAX-Gold-and-S-P500-Time-Series/blob/master/README.md