ECA5315 Research Proposal

1. Literature Review

The purpose of this research proposal is to expand upon the extensively studied topic of the co-movements of prices of **Gold, Crude Oil** and **US Dollar Index**. The ability to correctly identify their relationships would be of immense interest to financial markets participants, who could exploit potential price deviations for arbitrage purposes.

Denominated in USD, Gold and Crude Oil prices have been commonly thought to be inversely related to US Dollar Index prices. Similarly, being a commodity, Crude Oil prices have also been thought to be inversely related to Gold prices, i.e. viewed as a traditional safe haven investment. Several past studies have sought to elicit their price transmission relationships as follows:

a) Study by Wang et al. for **January 1989** to **December 2007** (1):
   - Gold and Oil prices have positive influence on each other.
   - US Dollar negatively affects Gold prices and has no effect on Oil prices during the following period.

b) Study by Arfaoui et al. for **January 1995** to **October 2015** (2):
   - Oil prices are positively affected by prior Gold and US Dollar prices.
   - Gold prices are positively affected by prior Oil and US Dollar prices.
   - US Dollar prices are negatively affected by prior Oil and Gold prices.

c) Study by Chan et al. for **September 2007** to **December 2011** (3):
   - Oil prices are negatively affected by prior Gold and US Dollar prices.
   - US Dollar prices are negatively affected by prior Gold and Oil prices.
   - Gold prices are positively affected by prior US Dollar and Oil prices.

Given some time since the publishing of studies above, significant changes in their global supply and/or demand structure in these variables could have happened then and re-validation of their dynamic relationships would be necessary. Here, the research proposal would seek to elicit their relationship via Vector Autoregression (VAR) for the period of **January 2015** to **April 2020.** Ancillary analyses such as Granger Causality Tests and Impulse Response Tests would be performed too. For forecast evaluation purposes, typical forecast metrics such as Mean Absolute Error (MAE) would be computed for the period of 1st March to 18th April 2020, i.e. the particularly volatile COVID-19 period to assess its predictive efficacies.

2. Description of Data

Daily closing-price data from **January 2015** to **April 2020** would be extracted from Yahoo via Python's pandas-datareader library and utilised here for regression. Any variables with data-points falling on dates not available on other variables would be dropped to ensure consistency between all three variables. The tickers of interest are **GC=F**, **CL=F** and **DX-Y.NYB** for **Gold, Crude Oil** and **US Dollar Index** respectively.

2.1. Summary Statistics

The summary statistics of the 3 time series have been presented in Fig 1 below:

Fig 1: Summary Statistics:

	DXY	Gold	CrudeOil
count	1322.000000	1322.000000	1322.000000
mean	96.282980	1283.960515	52.402292
std	2.737386	125.050276	9.987789
min	88.589996	1050.800049	18.120001
25%	94.620003	1207.900024	46.432500
50%	96.514999	1267.950012	52.139999
75%	97.959999	1322.775024	59.027499
max	103.290001	1769.400024	76.410004

For better visualisation of the above summary statistics, the plots of individual time series have been presented in **Fig 2 to 4** below:

Fig 2: Spot Gold Prices over Time

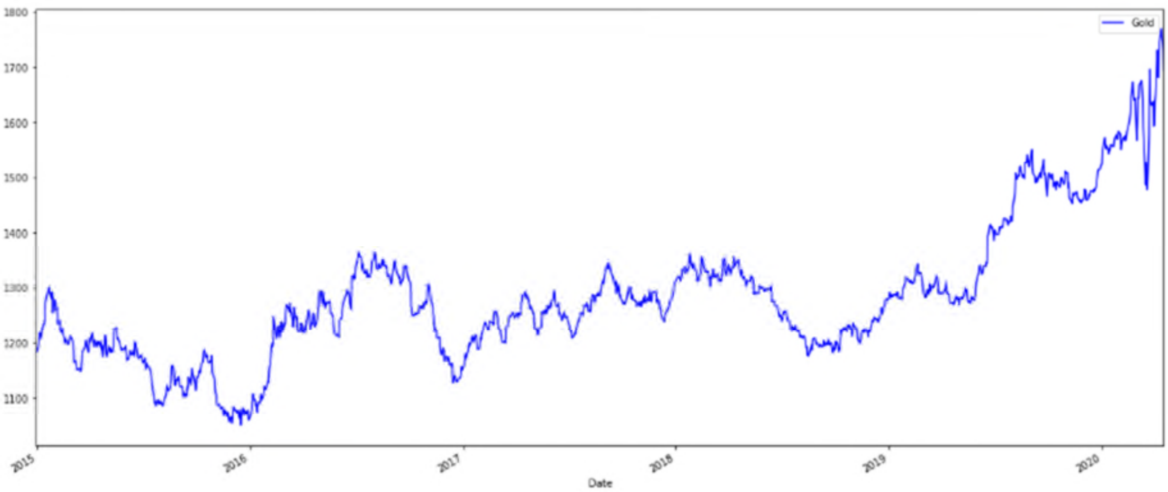


Fig 3: Spot Crude Oil Prices over Time

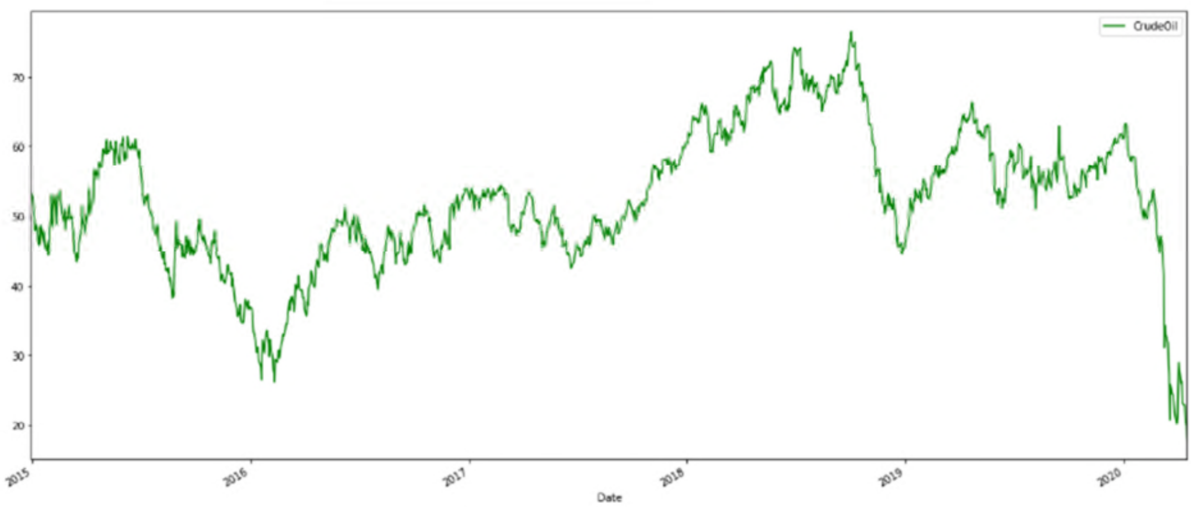
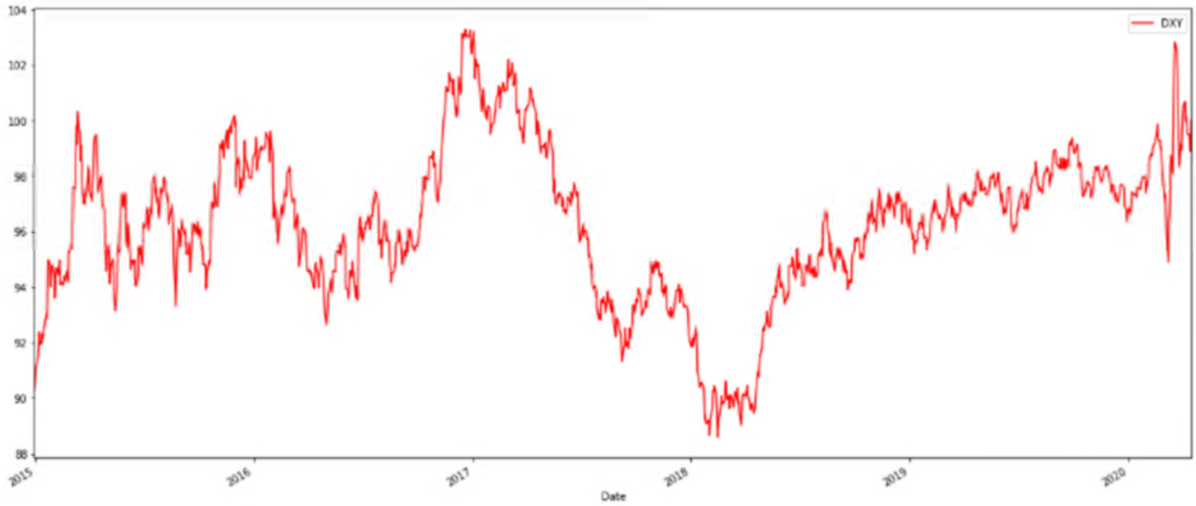


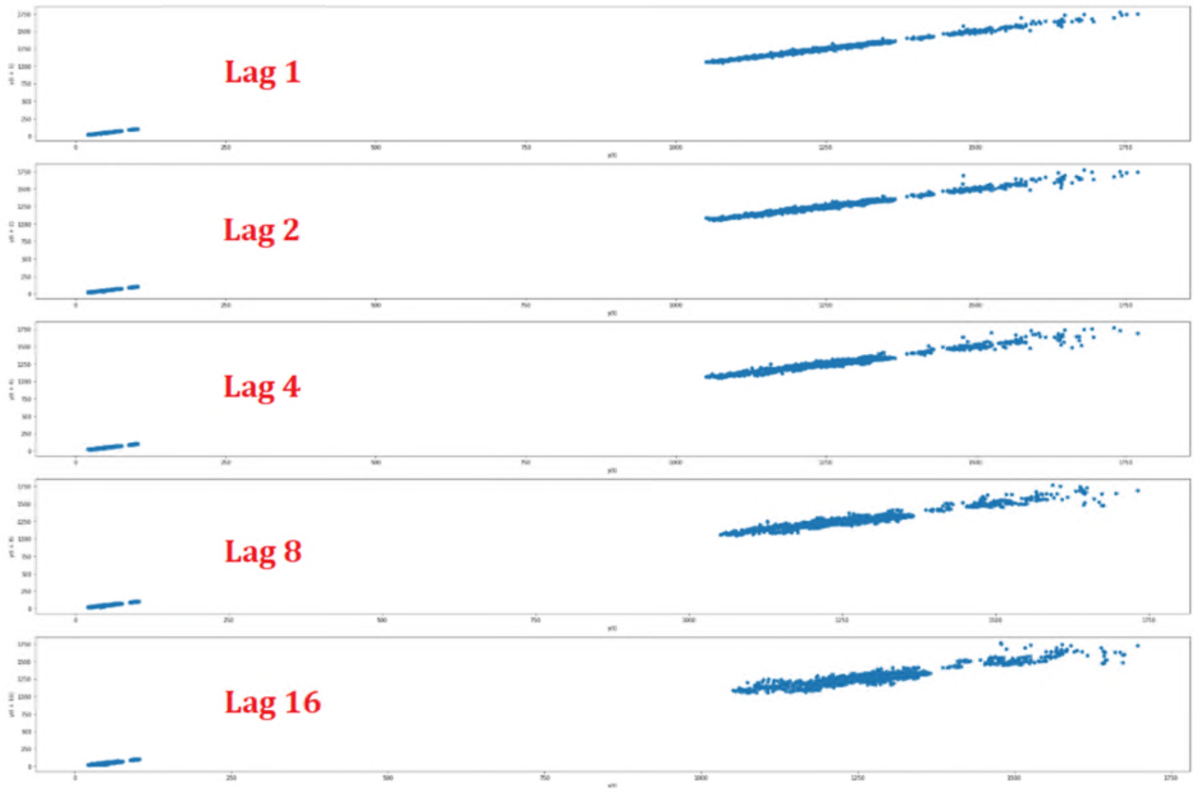
Fig 4: US Dollar Index Prices over Time



2.2. Is Data Random?

Lag plots are effective ways of checking for non-randomness of data, since it is highly unlikely random data would exhibit any structure in its lag plots. The lag plots would be performed for the dataset from **Lag 1, 2, 4, 8** and **16** in **Fig 5** below:

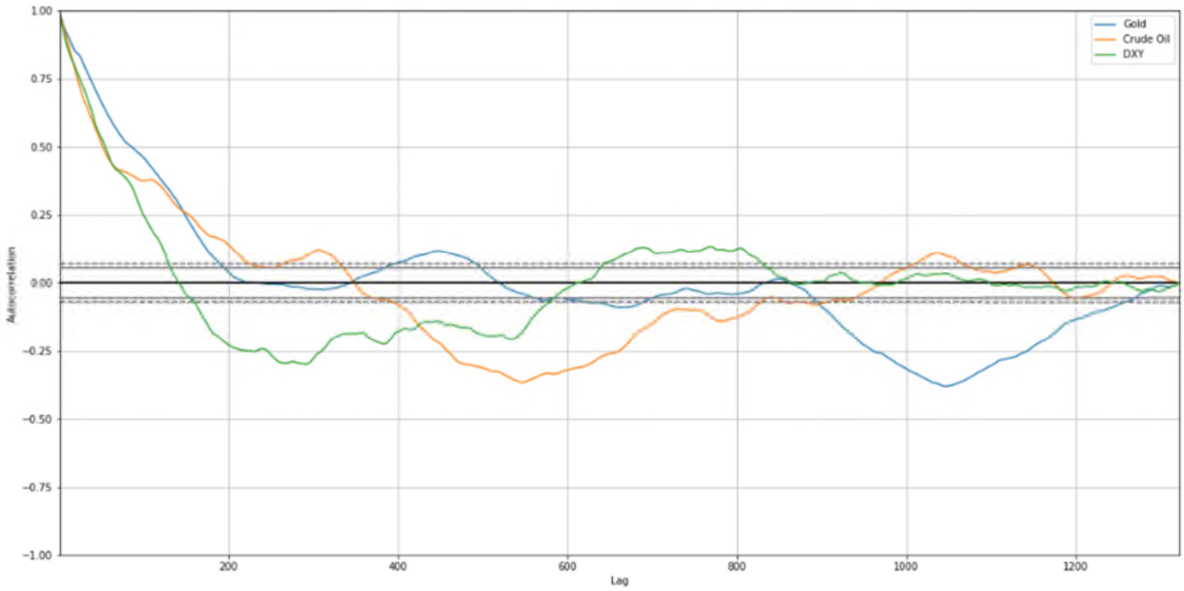
Fig 5: Dataset Lag-Plot Analysis



A linear structure can be seen in the lag-plots of the dataset, indicating non-randomness of data and possibly, presence of positive autocorrelations. It also suggests that an **Autoregressive (AR) model** would be well-suited for the modelling of the time series here. Besides, it can be observed that past **Lag 4,** the linearity of the lag-plot structure starts to "fade", which is expected in any AR model with older data typically providing little incremental information.

To confirm the existence of the positive autocorrelations, autocorrelation plots would be performed in **Fig 6** below:

Fig 6: Autocorrelation Plots



The above plots confirms the observations found in the lag plots, with the plots not falling within the **95 and 99% confidence intervals**, i.e. the horizontal lines above, and that autocorrelation exists.

2.3. Is Data Stationary?

Any time-series statistical model would require stationary checks, an ideal condition implied in most models, to ensure that the underlying statistical properties remains unchanged over time, i.e. the removal of dependencies between data in each time period (4).

To check for stationary processes, Augmented Dickey-Fuller (ADF) unit root test would be performed on each of the time series here, where the null hypothesis would be the presence of a unit root, and the alternative (hypothesis) the presence of a weakly stationary process. The results have been presented in **Fig 7** to **9** below:

Fig 7: ADF Test for Gold

```
Augmented Dickey-Fuller Test on Gold
-------------------------------------------------

Critical Values:         -3.4354 (1%), -2.8638 (5%), -2.5680 (10%)

Null Hypothesis:         The process contains a unit root.
Alternative Hypothesis:  The process is weakly stationary.


Test Statistic           = 0.3279
No of Lags Chosen        = 16

P-Value                  = 0.9786
Is Stationary?           Non-Stationary (1%), Non-Stationary (5%), Non-Stationary (10%)
```

Fig 8: ADF Test for Crude Oil

```
Augmented Dickey-Fuller Test on CrudeOil
-------------------------------------------------

Critical Values:        -3.4353 (1%), -2.8637 (5%), -2.5679 (10%)

Null Hypothesis:        The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.


Test Statistic          = -1.1644
No of Lags Chosen       = 7

P-Value                 = 0.6887
Is Stationary?          Non-Stationary (1%), Non-Stationary (5%), Non-Stationary (10%)
```

Fig 9: ADF Test for US Dollar Index

```
Augmented Dickey-Fuller Test on DXY
-------------------------------------------------

Critical Values:        -3.4354 (1%), -2.8638 (5%), -2.5679 (10%)

Null Hypothesis:        The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.


Test Statistic          = -2.3889
No of Lags Chosen       = 13

P-Value                 = 0.1449
Is Stationary?          Non-Stationary (1%), Non-Stationary (5%), Non-Stationary (10%)
```

From the figures above, it is evident the time series data for Gold, Crude Oil and US Dollar Index are **non-stationary**, whether at **1%, 5% or 10% significance levels**. The null hypothesis that the process contains a unit root **cannot be rejected** for all 3 time series data at all 3 significance levels.

2.4. Data Processing

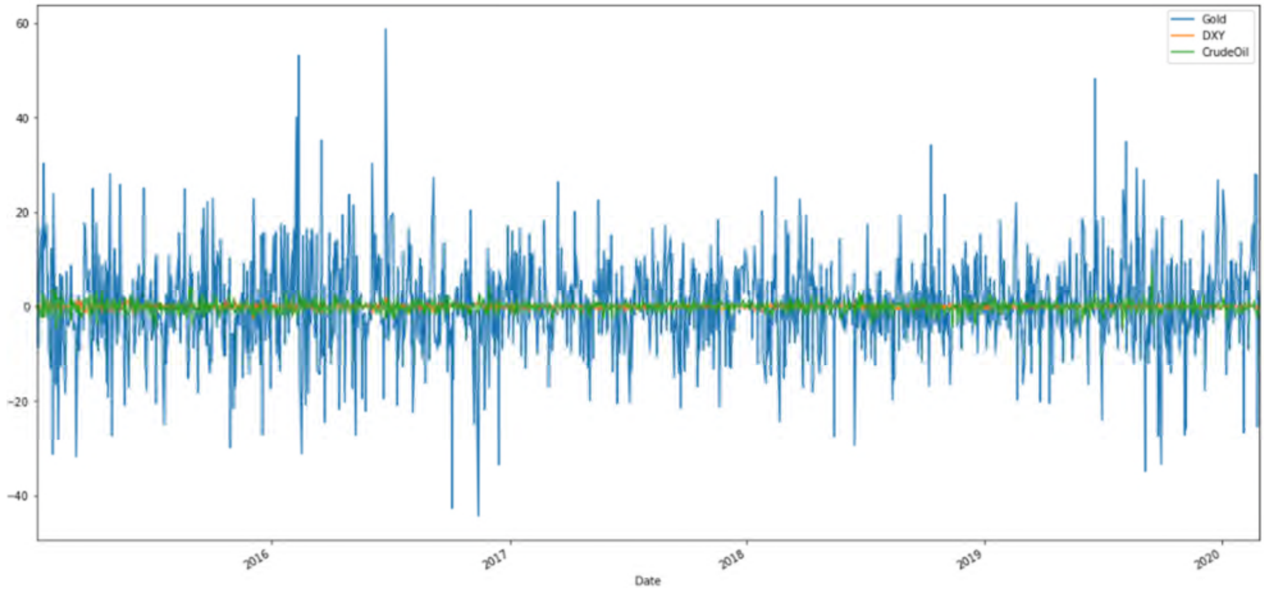
A common strategy to convert non-stationary processes into stationary ones involves first-differencing of the dataset. Prior to differencing of the dataset, the particularly volatile period from **1st March to 18th April 2020** would be removed from the dataset, and used only for forecasting evaluation. The summary statistics of the first differenced dataset is:

Fig 10: Summary Statistics of First Differenced Dataset

	DXY	Gold	CrudeOil
count	1286.000000	1286.000000	1286.000000
mean	0.005482	0.354977	-0.004355
std	0.412148	10.304586	1.160364
min	-2.369995	-44.300049	-4.630001
25%	-0.230003	-5.099976	-0.677501
50%	0.010002	0.150024	0.059998
75%	0.239998	5.800049	0.677501
max	1.919998	58.800049	8.050003

For visualisation purposes, the differenced dataset plots have been provided in **Fig 11** below:

Fig 11: Differenced Dataset Plots



To ensure that the first differenced datasets are now stationary, ADF tests would be performed on them again, with the corresponding results provided in **Fig 12** to **14** below:

Fig 12: ADF Test for Differenced Gold

```
Augmented Dickey-Fuller Test on Gold
--------------------------------------------------

Critical Values:         -3.4354 (1%), -2.8638 (5%), -2.5680 (10%)

Null Hypothesis:         The process contains a unit root.
Alternative Hypothesis:  The process is weakly stationary.


Test Statistic           = -36.4287
No of Lags Chosen        = 0

P-Value                  = 0.0000
Is Stationary?           Stationary (1%), Stationary (5%), Stationary (10%)
```

Fig 13: ADF Test for Differenced Crude Oil

```
Augmented Dickey-Fuller Test on CrudeOil
--------------------------------------------------

Critical Values:         -3.4354 (1%), -2.8638 (5%), -2.5680 (10%)

Null Hypothesis:         The process contains a unit root.
Alternative Hypothesis:  The process is weakly stationary.


Test Statistic           = -38.2505
No of Lags Chosen        = 0

P-Value                  = 0.0000
Is Stationary?           Stationary (1%), Stationary (5%), Stationary (10%)
```

Fig 14: ADF Test for Differenced US Dollar Index

```
Augmented Dickey-Fuller Test on DXY
--------------------------------------------------

Critical Values:        -3.4354 (1%), -2.8638 (5%), -2.5680 (10%)

Null Hypothesis:        The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.


Test Statistic          = -35.6592
No of Lags Chosen       = 0

P-Value                 = 0.0000
Is Stationary?          Stationary (1%), Stationary (5%), Stationary (10%)
```

From **Fig 12** to **14**, upon first differencing, the null hypothesis **can now be rejected** at all 3 significance levels of **1%, 5%** and **10%** for all 3 time series data, and the originally non-stationary time series data have been successfully converted to a stationary one. Further checks would also be made to ensure no serial autocorrelations exists within the differenced dataset, with each time series' ACF and PACF plotted in **Fig 15** to **17** below to confirm this:

Fig 15: Differenced Gold ACF and PACF:

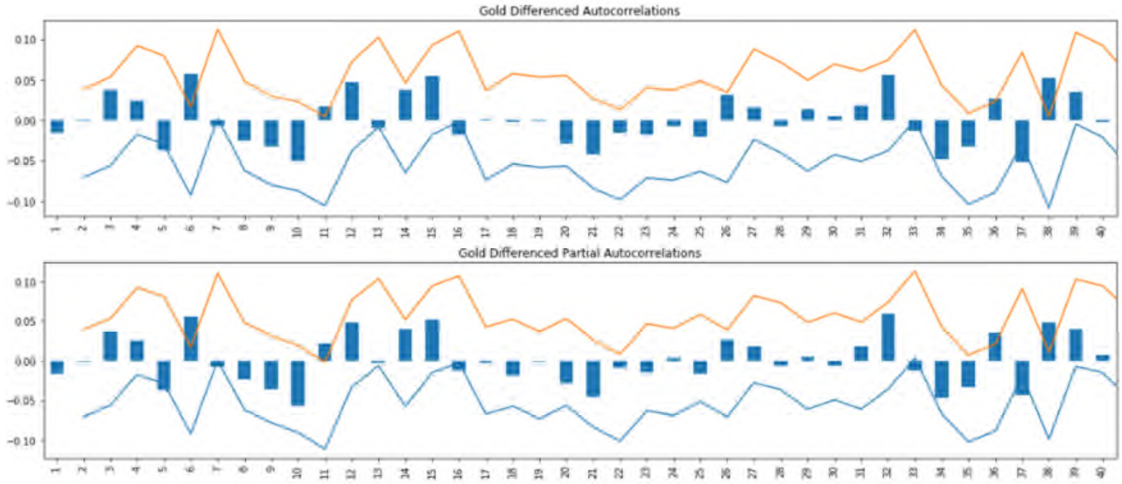


Fig 16: Differenced Crude Oil ACF and PACF:

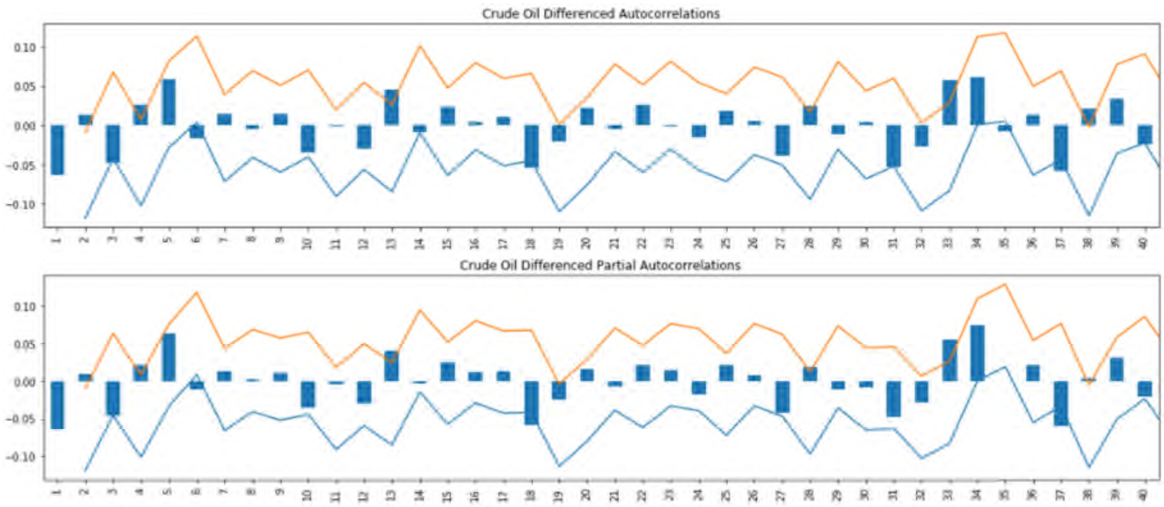


Fig 17: Differenced DXY ACF and PACF:

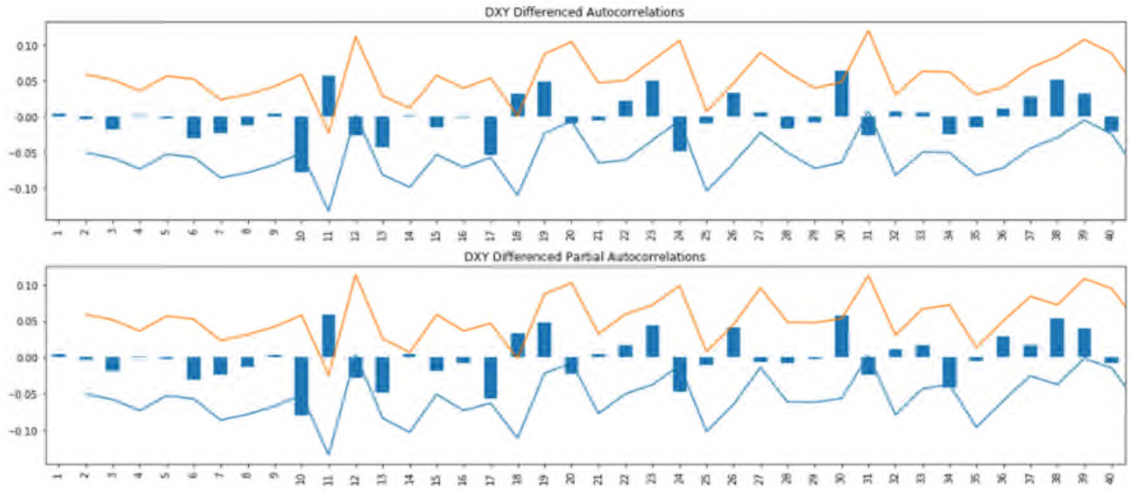


Fig 15 to **17** have shown that the PACF and ACF plots all falls within the 5% confidence intervals, indicating the absence of serial autocorrelations within all 3 time series data, confirming the success of first-differencing in removing serial autocorrelations between periods.

2.5. Granger Causality

Prior to implementing any regression models, it's critical to elicit some underlying relationships between the variables of interest, i.e. Crude Oil, Gold and US Dollar Index Prices. A simple way to achieve this would be running Granger Causality Tests up to lag 10 for the first-differenced dataset, with the results tabulated in **Table 1 to 3** below:

Table 1: Granger Causality for Gold

	Crude Oil		US Dollar Index	
Lag	**F-test**	**P-Value**	**F-test**	**P-Value**
1	0.5572	0.4555	6.01	0.0144
2	1.4649	0.2315	3.1976	0.0412
3	1.0712	0.3602	2.9125	0.0334
4	1.1031	0.3536	2.2643	0.0603
5	1.0273	0.4	2.0438	0.07
6	0.8275	0.5486	2.6337	0.0153
7	0.8986	0.5067	2.3803	0.0203
8	0.8394	0.5678	2.1315	0.0303
9	0.7024	0.7072	1.9291	0.0443
10	0.8612	0.5695	1.8592	0.0469

Table 2: Granger Causality for Crude Oil

	Gold		US Dollar Index	
Lag	**F-test**	**P-Value**	**F-test**	**P-Value**
1	0.0389	0.8437	0.0054	0.9412
2	0.3206	0.7258	0.038	0.9627
3	0.2773	0.8418	0.049	0.9856

4	0.5766	0.6797	0.3105	0.8711
5	0.9861	0.4249	0.3165	0.9033
6	0.7496	0.6098	0.3522	0.9088
7	0.6621	0.7044	0.3219	0.9443
8	0.6724	0.7162	0.4014	0.9202
9	0.8512	0.5688	0.4352	0.9165
10	0.8831	0.5485	0.4356	0.9295

Table 3: Granger Causality for US Dollar Index

	Crude Oil		**Gold**	
Lag	**F-test**	**P-Value**	**F-test**	**P-Value**
1	2.8069	0.0941	0.4919	0.4832
2	2.9439	0.053	1.3963	0.2479
3	2.7666	0.0406	1.132	0.3349
4	2.2203	0.0648	1.4971	0.2007
5	1.9868	0.0779	1.6676	0.1394
6	1.6603	0.1273	1.6294	0.1354
7	1.4858	0.1681	1.4621	0.1768
8	2.3353	0.0172	1.3165	0.2308
9	2.14	0.0238	2.2953	0.0148
10	1.8894	0.0427	1.9951	0.0306

From the above tables, at a **5% significance level**, the overall relationships between the first-differenced time series data can be broadly summarised (for most lags) as:

a) Crude Oil "Granger-cause" Gold prices and vice versa.
b) US Dollar Index does not "Granger-cause" prices of Gold, though Gold in turn "Granger-cause" US Dollar Index.
c) Crude Oil does not "Granger-cause" prices of US Dollar Index, though US Dollar Index "Granger-cause" Crude Oil prices.

There's a few interesting findings above:

a) Gold seems to be a "forward-looking" indicator when it comes to affecting prices of the other two time series data. Instead of Gold reacting to lags in DXY as one would expect, the reverse is true.
b) DXY lags affects Crude Oil prices as how one would expect in practice.

3. Model Selection

With the differenced dataset, the appropriate model would need to be selected to fit the dataset. Given linearity of lag-plots found in **Section 2.2**, in addition to the time series interchangeably causing each other found by Granger Causality studies made in **Section 2.5**, a **VAR model** is likely well-suited to model the co-interactions between **Gold, Crude Oil** and **US Dollar Index**.

Before proceeding with its implementation, in any reduced-form VAR model, the ordering of variables plays a crucial significance, since it constitutes an implicit identification of restrictions as variables are computed in recursive fashion within the model (5). Therefore, to potentially identify the correct ordering, findings from **Section 2.5** Granger-causality tests would be utilised here for guidance. Given Gold "Granger-causing" the remaining variables, and US Dollar Index "Granger-causing" Crude Oil, the ordering of the variables in the VAR Model would be as follows: Gold first, followed by US Dollar Index, then Crude Oil.

The reduced-form VAR will be estimated as per the following equation:

$$Y = BZ + U$$

The VAR model, selected to minimise the Akaike Information Criterion (AIC) up to max lags of 15, is estimated here. The resulting summary statistics of the estimated VAR model is described in **Fig 18** below:

Fig 18: Summary of VAR Model

```
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
Date:           Sun, 26, Apr, 2020
Time:                     23:12:44
--------------------------------------------------------------------
No. of Equations:         3.00000    BIC:                    3.04711
Nobs:                     1285.00    HQIC:                   3.01702
Log likelihood:          -7384.82    FPE:                    20.0642
AIC:                      2.99894    Det(Omega_mle):         19.8780
--------------------------------------------------------------------
Results for equation Gold
===============================================================================
                 coefficient       std. error           t-stat            prob
-------------------------------------------------------------------------------
const               0.368466         0.287153            1.283           0.199
L1.Gold            -0.045416         0.030379           -1.495           0.135
L1.DXY             -1.821671         0.764179           -2.384           0.017
L1.CrudeOil         0.120749         0.248843            0.485           0.628
===============================================================================

Results for equation DXY
===============================================================================
                 coefficient       std. error           t-stat            prob
-------------------------------------------------------------------------------
const               0.005542         0.011510            0.482           0.630
L1.Gold            -0.000874         0.001218           -0.718           0.473
L1.DXY             -0.009694         0.030631           -0.316           0.752
L1.CrudeOil        -0.016773         0.009975           -1.682           0.093
===============================================================================

Results for equation CrudeOil
===============================================================================
                 coefficient       std. error           t-stat            prob
-------------------------------------------------------------------------------
const              -0.002206         0.032323           -0.068           0.946
L1.Gold            -0.000844         0.003420           -0.247           0.805
L1.DXY             -0.014319         0.086018           -0.166           0.868
L1.CrudeOil        -0.063886         0.028010           -2.281           0.023
===============================================================================

Correlation matrix of residuals
                Gold       DXY  CrudeOil
Gold        1.000000 -0.403164  0.041197
DXY        -0.403164  1.000000 -0.116611
CrudeOil    0.041197 -0.116611  1.000000
```

This translates to a reduced-form VAR form of:

$$\textbf{Eqn (1)}:\ \Delta\mathrm{Gold}_q = 0.368466 - 0.045416\,\Delta\mathrm{Gold}_{q-1} - \mathbf{1.821671\,\Delta DXY_{q-1}} + 0.120749\,\Delta\mathrm{CrudeOil}_{q-1} + \varepsilon_t$$

$$\textbf{Eqn (2)}:\ \Delta\mathrm{DXY}_q = 0.005542 - 0.000874\,\Delta\mathrm{Gold}_{q-1} - 0.009694\,\Delta\mathrm{DXY}_{q-1} - 0.016773\,\Delta\mathrm{CrudeOil}_{q-1} + \varepsilon_t$$

$$\textbf{Eqn (3)}:\ \Delta\mathrm{CrudeOil}_q = -0.002206 - 0.000844\,\Delta\mathrm{Gold}_{q-1} - 0.014319\,\Delta\mathrm{DXY}_{q-1} + \mathbf{0.063886\,\Delta CrudeOil_{q-1}} + \varepsilon_t$$

**** Statistically insignificant at two-tailed 5% significance level**

From the equations found above, a few key relationships can be observed:

a) The fitted VAR model only persists up to 1 lag.
b) Gold and US Dollar Index prices tend to "mean-revert", i.e. prior increases of their respective lags tend to lead to smaller decreases in following periods and vice versa.
c) Crude Oil prices tend to persists, i.e. prior increases tend to lead to following period increases. This could potentially be attributed to the price inelasticity of supply for Crude Oil (6), which tends to exacerbate its own price movements for some time. However, do note that this coefficient is **statistically insignificant** at 5% confidence intervals.
d) Gold prices tend to increase with prior increases in Crude Oil prices, though the reverse is not true. It is also inversely related to prior US Dollar Index price movements, as what theory suggests, though it is **statistically insignificant** at 5% confidence intervals.
e) US Dollar Index is inversely related to prior Gold and Crude Oil price movements, in line with theory.
f) Crude Oil is inversely related to prior Gold and US Dollar Index price movements.

4. Analysis

To interpret VAR models better, impulse responses and variance decompositions would be performed to analyse the dynamics from any applied shocks on the equations derived above:

4.1. Impulse Response

The impulse response function has been performed and plotted in **Fig 19** at 95% confidence intervals, with **Fig 20** showing cumulative effects of these impulses. Such an analysis provides insights into the estimated responses on the independent variables from a unit impulse to each of the dependent variables in the VAR model. From **Fig 19 and 20**, it can be seen that the effects of these unit impulses rapidly dissipate, not lasting more than two periods. This can potentially give credence to the theory of efficient markets, and that price deviations are arbitraged off in little time. Besides, in line with the VAR model described in **Section 3**, the findings found are:

a) Gold and DXY impulses does not affect Crude Oil prices, however minimally.
b) Crude Oil price movements significantly affect Gold and DXY price movements.
c) DXY impulse price movements affect Gold prices significantly, though it must be noted that DXY affecting that of Gold via **Eqn (1)** could potentially be statistically insignificant.
d) Gold impulse price movements also significantly affect DXY price movements.

Fig 19: Impulse Response

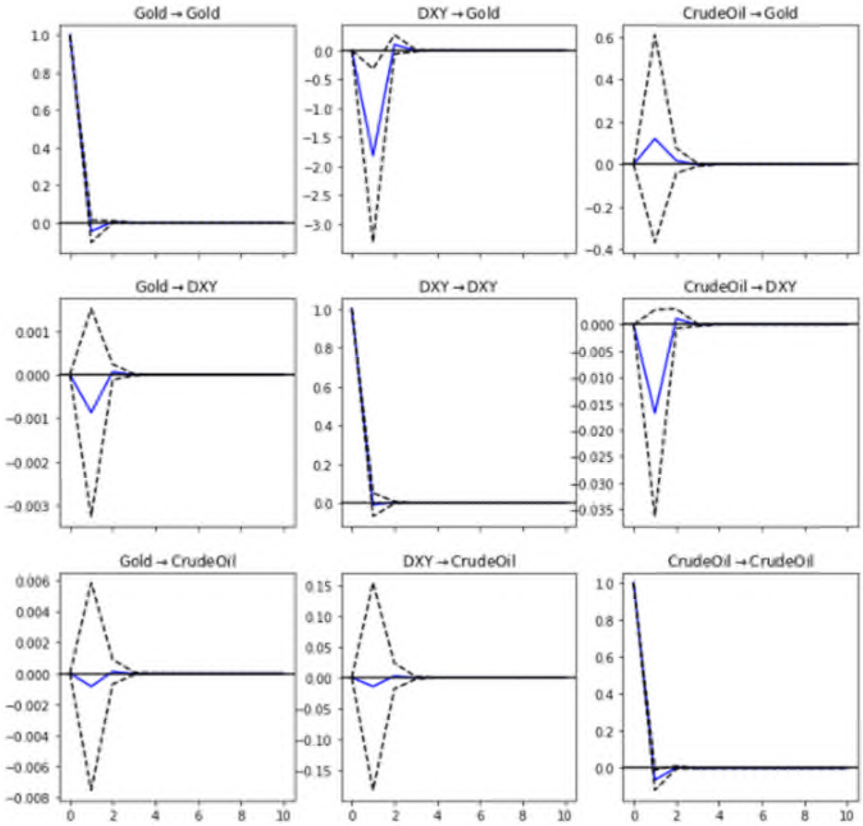
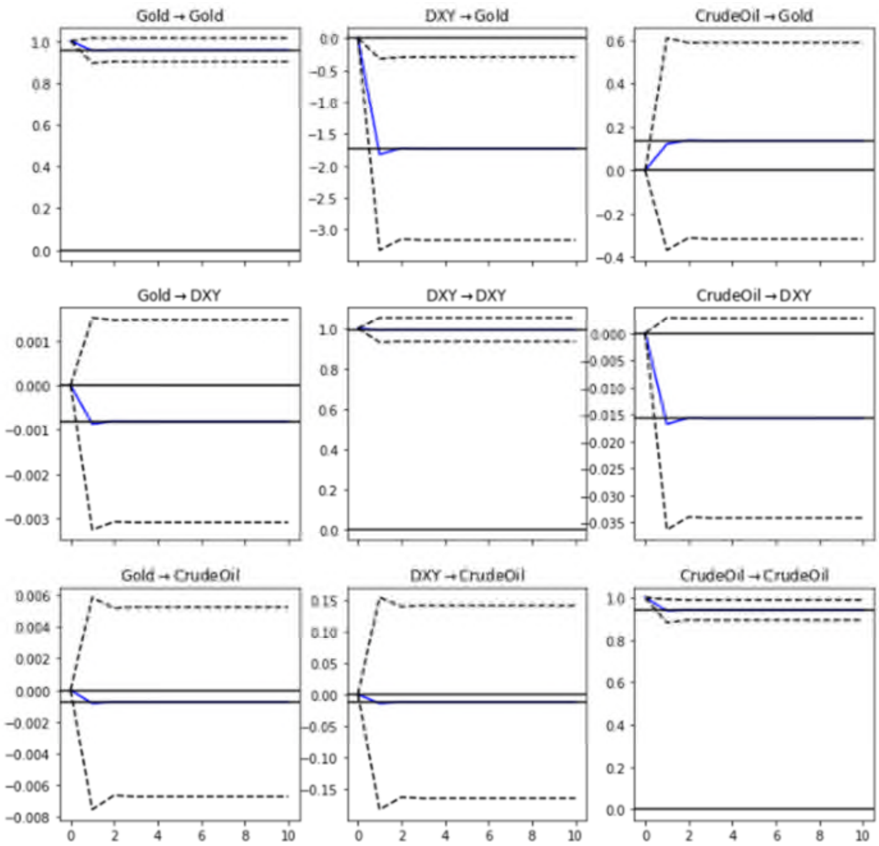


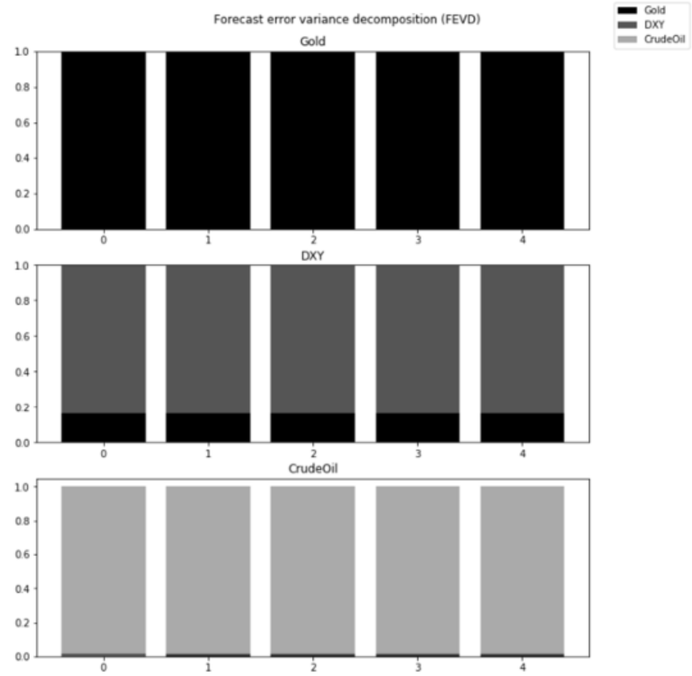
Fig 20: (Cumulative) Impulse Response



4.2. Forecast Error Variance Decomposition (FEVD)

An alternative to Impulse Response Analysis, FEVD can provide insights as to how much each variable's forecast error variance can be attributed to exogenous shocks to other variables, i.e. their relative shock importance (7). The figures have been provided in **Fig 21 and 22** for up to 5-steps ahead forecast, since forecast error remains almost constant post one lag given the estimated **VAR(1)** model:

Fig 21: FEVD Results



FEVD for Gold

	Gold	DXY	CrudeOil
0	1.000000	0.000000	0.000000
1	0.995176	0.004642	0.000181
2	0.995161	0.004654	0.000185
3	0.995161	0.004654	0.000185
4	0.995161	0.004654	0.000185

FEVD for DXY

	Gold	DXY	CrudeOil
0	0.162541	0.837459	0.000000
1	0.162513	0.835304	0.002183
2	0.162512	0.835296	0.002193
3	0.162512	0.835296	0.002193
4	0.162512	0.835296	0.002193

FEVD for CrudeOil

	Gold	DXY	CrudeOil
0	0.001697	0.011941	0.986362
1	0.001755	0.011898	0.986347
2	0.001756	0.011898	0.986346
3	0.001756	0.011898	0.986346
4	0.001756	0.011898	0.986346

From the results above, it can be observed that each of the three variables mostly explain their own variances, especially for Gold which explains its own variances almost completely. This Gold phenomenon observed is in line with the Granger-causality findings in **Section 2.5**, where it seems to be a "forward-looking" predictor. To some extent, DXY and Gold does explain Crude Oil variance to the tune of 1%, however minimal. Interestingly, Gold explains around 16% of DXY variance, like what a "forward-looking" predictor should.

5. Evaluation

The limitations with the above model fitted would be provided in this section and potential improvements suggested and detailed

5.1. Limitations & Improvements

Two model limitations would be identified below:

a) Omission of potentially correlated variables

There are potentially time series correlated with the above three variables (8; 9), for example major stock indices such as **S&P500** and **NASDAQ**, other commodities such as **Silver** and **Copper**, long-dated treasury bond indices such as **TLT**, and volatility indices such as **VIX** etc. Such omissions might lead to model biases and could result in unreliability of model generated. To illustrate this, suppose a VAR model of the form:

$$y_t = A_o + A_1 y_{t-1} + \varepsilon_t$$

The omission of a pertinent variable from **y** itself would mean $\boldsymbol{y_{t-1}}$would be correlated with $\boldsymbol{\varepsilon_t}$ and might cause the violation of implicit assumption of reduced-form VAR, which is no serial autocorrelations between error terms over time (4).

Improvements could include fitting the model with the inclusion of these potentially correlated time series and evaluating the appropriate improvements in the model using various criterions such as AIC and possibly Bayesian information criterion (BIC), to find the right combinations of variables.

b) Forecast efficacy on validation dataset

The implicit disadvantage of any **VAR(n)** model is that accuracy tends to declines as forecasting horizon increase, especially past the n lags it is generated for, where any forecast beyond n-steps tend to yield little incremental information and error variances approaching that of white-noises.

An example has been done for the generated VAR model on the first-differenced validation dataset during the COVID-19 period of **1st March to 18th April 2020** in **Table 4** below:

Table 4: Subsection of Forecasted Values

Date	Gold	DXY	CrudeOil
2020-03-02	0.995459	0.004208	-0.015788
2020-03-03	0.313684	0.004896	-0.002098
2020-03-04	0.345046	0.005256	-0.002407
2020-03-05	0.342930	0.005230	-0.002419
2020-03-06	0.343071	0.005233	-0.002416
2020-03-09	0.343061	0.005232	-0.002416
2020-03-10	0.343062	0.005232	-0.002416
2020-03-11	0.343062	0.005232	-0.002416
2020-03-12	0.343062	0.005232	-0.002416
2020-03-13	0.343062	0.005232	-0.002416
2020-03-16	0.343062	0.005232	-0.002416
2020-03-17	0.343062	0.005232	-0.002416
2020-03-18	0.343062	0.005232	-0.002416
2020-03-19	0.343062	0.005232	-0.002416
2020-03-20	0.343062	0.005232	-0.002416
2020-03-23	0.343062	0.005232	-0.002416
2020-03-24	0.343062	0.005232	-0.002416
2020-03-25	0.343062	0.005232	-0.002416

It can be observed the forecasted first-difference values quickly converge and do not change much beyond 5 steps. When one actually "un-difference" the values and line it up side-by-side with actual values that occurred (with forecast metrics computed), the predictive efficacy of the model is generally underwhelming as it attempts to extrapolate the values outside of regressed intervals, as shown in **Table 5**, **Fig 22 and 23** below:

Table 5: Forecast vs Actual Comparison Subsection

Date	Gold_Actual	DXY_Actual	CrudeOil_Actual	Gold_Pred	DXY_Pred	CrudeOil_Pred
2020-03-02	1566.699951	97.360001	44.759998	1567.695410	97.364209	44.744210
2020-03-03	1594.800049	97.150002	46.750000	1568.009094	97.369105	44.742112
2020-03-04	1644.400024	97.339996	47.180000	1568.354141	97.374361	44.739705
2020-03-05	1643.000000	96.820000	46.779999	1568.697070	97.379591	44.737286
2020-03-06	1668.000000	95.949997	45.900002	1569.040141	97.384824	44.734870
2020-03-09	1672.400024	94.900002	41.279999	1569.383202	97.390056	44.732454
2020-03-10	1675.699951	96.410004	31.129999	1569.726264	97.395289	44.730037
2020-03-11	1660.300049	96.510002	34.360001	1570.069325	97.400521	44.727621
2020-03-12	1642.300049	97.470001	32.980000	1570.412387	97.405754	44.725204
2020-03-13	1590.300049	98.750000	32.930000	1570.755449	97.410986	44.722788
2020-03-16	1516.699951	98.089996	31.730000	1571.098510	97.416218	44.720372
2020-03-17	1486.500000	99.580002	28.700001	1571.441572	97.421451	44.717955
2020-03-18	1525.800049	101.160004	27.330000	1571.784633	97.426683	44.715539
2020-03-19	1477.900024	102.760002	20.830000	1572.127695	97.431916	44.713122
2020-03-20	1479.300049	102.820000	25.910000	1572.470757	97.437148	44.710706
2020-03-23	1575.699951	102.489998	24.490000	1572.813818	97.442381	44.708290
2020-03-24	1696.099976	101.739998	24.490000	1573.156880	97.447613	44.705873
2020-03-25	1638.099976	101.050003	24.190001	1573.499941	97.452845	44.703457

Fig 22: Plots of Forecast vs Actual

Fig 23: Forecast Metrics Results over Period

```
Gold Bias: 64.2297
Gold Mean Absolute Error: 86.2131
Gold Mean Squared Error: 9803.5582
Gold Root Mean Squared Error: 99.0129

Crude Oil Bias: -15.9770
Crude Oil Mean Absolute Error: 16.4283
Crude Oil Mean Squared Error: 331.6473
Crude Oil Root Mean Squared Error: 18.2112

DXY Bias: 1.7459
DXY Mean Absolute Error: 2.1352
DXY Mean Squared Error: 6.6148
DXY Root Mean Squared Error: 2.5719
```

The model particularly fare badly for such volatile periods, where the price movements of the forecasted prices of the three variables were all minute compared to actual movements of the underlying variables.

Improvements could include using alternative forms of model such as VARMA, which have been found in a study to forecast macroeconomic variables better than equivalent VAR models (10). Equivalently, VARMA could potentially be used to forecast such time series variables that are also heavily influenced by macroeconomic variables as well.

6. Conclusion

In this research proposal, the co-interactions of **Gold, Crude Oil** and **US Dollar Index** were studied using a Vector Autoregression (VAR) model, where underlying relationships between the time series were elicited. Here, Gold was found to be generally a "forward-looking" predictor of **US Dollar Index** especially. Crude Oil was found to largely to move on its own, with little influence from the other two time series variables. US Dollar Index, however, was found to be significantly influenced by lagged price movement of Gold prices. Overall, the following key relationships were identified:

a) Crude Oil price movements are momentum-based, i.e. prior price movements tend to persist, and is mostly unaffected by prior US Dollar Index and or Gold price movements. Its price movements though, affects future US Dollar Index and Gold price movements in a negative and positive manner respectively.
b) Gold price movements tend to "mean-revert", and is generally affected by prior Crude Oil and US Dollar Index price movements in a positive and negative manner respectively. It is also found to significantly influence future US Dollar Index price movements negatively.
c) US Dollar Index price movements also "mean-revert" as well, and is negatively affected by prior Gold and Crude Oil price movements. Its own price movements tend to negatively affect future Gold price movements as well.

Using the Impulse Response Analysis, the price shocks are found to dissipate quickly, indicating little scope for continuous arbitrage opportunities over time from any applied shocks. It also found the generated VAR model generally provides ineffective forecast of future prices, especially for volatile periods, with its usefulness being limited to only the understanding of price transmission relationships discussed above. Improvements to the model could include the inclusion of other potentially correlated time series variables and the use of alternative models such as VARMA, which has shown to be a more effective predictor of macroeconomic variables.

7. Bibliography

1. *Dynamic transmission effects between the interest rate, the US dollar, and gold and crude oil prices.* **Wang, Yu Shan and Chueh, Yen Ling.** s.l. : ScienceDirect, January, 2013, Economic Modelling, Vol. 30, pp. 792-798.

2. *Oil, gold, US dollar and stock market interdependencies: a global analytical insight.* **Arfaoui, Mongi and Rejeb , Aymen Ben.** s.l. : Emerald Publishing Limited, October, 2017, European Journal of Management and Business Economics, Vol. 26, pp. 278-293.

3. *Interactive relationships between crude oil prices, gold prices, and the NT–US dollar exchange rate—A Taiwan study.* **Chang, Hsiao-Fen, Huang, Liang-Chou and Chin, Ming-Chin.** s.l. : ScienceDirect, December, 2013, Energy Policy, Vol. 63, pp. 441-448.

4. **Studenmund, A.H.** *Using Econometrics.* 6. s.l. : Pearson, 2011. p. 648.

5. *On the Identification of Structural Vector Autoregressions.* **Sarte, Pierre-Daniel G.** s.l. : Federal Reserve Bank of Richmond, 1997, FRB Richmond Economic Quarterly, Vol. 83, pp. 45-67.

6. *Determinants of the crude oil futures curve: Inventory, consumption and volatility.* **Nikitopoulos, Christina Sklibosios, et al., et al.** s.l. : ScienceDirect, November, 2017, Journal of Banking & Finance, Vol. 84, pp. 53-67.

7. *Wikipedia.* [Online] [Cited: 24 April, 2020.] https://en.wikipedia.org/wiki/Variance_decomposition_of_forecast_errors.

8. *The Interactions among Gold, Oil, and Stock Market: Evidence from S&P500.* **Gokmenoglua, Korhan K. and Fazlollahia, Negar.** s.l. : ScienceDirect, May, 2015, Procedia Economics and Finance, Vol. 25, pp. 478 – 488.

9. *Hedges and safe havens: An examination of stocks, bonds, gold, oil and exchange rates.* **Ciner, Cetin, Gurdgiev, Constantin and Lucey, Brian M.** s.l. : ScienceDirect, September, 2013, International Review of Financial Analysis, Vol. 29, pp. 202-211.

10. *VARMA versus VAR for Macroeconomic Forecasting.* **Athanasopoulos, George and Vahid, Farshid.** 2, s.l. : Taylor & Francis Online, 2008, Journal of Business & Economic Statistics, Vol. 26, pp. 237-252.