University of Warsaw

Data Science & Business Analytics

The market price analysis of gold and petrol: Evidences with cointegration and causality using Time Series data

Project report

in the field of Econometrics

Semester 2

Authors: Anirban Das (454449),

Antek Piotrowski (407067).

Warsaw, 14th June – 2023

Contents

Abstract

This study looks into the relationship between oil and gold prices trying to verify two hypotheses. (1) the existence of a stable equilibrium relationship between the prices over the long run, with the two variables moving and fluctuating together and (2) the possibility of defining this relationship explicitly using a linear model. The first hypothesis is supported by cointegration analysis, which shows that there is an equilibrium link between oil and gold prices over the long run. The second hypothesis is also supported by the error correction term (ECT) analysis, which shows that the linear model successfully captures short-term dynamics and returns to the long-term equilibrium. The Ramsey Reset Test further establishes that the employed regression model is free of specification flaws. However, the Breusch-Pagan and Breusch-Godfrey tests' findings of heteroscedasticity and autocorrelation point to possible problems with the traditional linear regression assumptions.

1. Introduction

The price of crude oil and gold has been one of the two most important representatives in the large commodity market, which are not completely effected or driven by the supply and demand market trends. According to (W.-C. Lee & Lin, 2012), both of the prices should move and get affected together as the macroeconomic factors like interest rates, and exchange rates. It is also frequently established that the gold is used as a reserve asset which proves to be a "safe haven" in compared to other financial assets such as stocks. Specific evidences from the 2008 global economic crisis showed that the crude oil price nose-dived from $147 to $30 per barrel, simultaneously the gold declined steady from $1000 per ounce to $700 per ounce. It also showed changes and affects over one another over large period of time as a part of practicality in response to shocks and fluctuations. While gold being the most dominantly traded precious metal, the crude oil takes the stage for most raw material, both playing highly important role in shaping the global economy.

Historical evidences proved that the during a time of conflicts, chaos in the world economies and geographical areas, the oil prices drastically changed simultaneously the price of gold. As per (Gil-Alana et al., 2017), "During the 70 s and 80 s, the average real gold prices increased by 2.04 and 1.76 times compared to the previous period. There was a temporary spike in the gold/oil ratio from 12.5 in autumn 1979 to 21 in winter 1980. This was due to a $400 per troy ounce price jump in gold from September 1979 to January 1980 resulting from the Soviet Union's invasion of Afghanistan. In real terms, gold hit an all-time high of $1537.94 per troy ounce in the year 1980". Followed by Iraq's invasion in Kuwait, Persian Gulf wars, Crisis in East Africa, then unrests in Venezuela in early 2000s, lead to huge changes in the global crude oil pricing which is also highly affected by the global inflation. Simultaneously the prices of the gold changed and the gold-crude oil dynamics kept proving it's worth to study due to ever changing dynamics of the relationship and gather theoretical with experimental relationship between the two time series. Thus it is very important to analyse the oil-gold price relationship over decades of changes and their fluctuations.

This paper revolves around the idea of the interactive mechanism between the two largest commodities in the world, leading to the primary hypotheses that there exists a long term relationship or co-integration between the oil and gold price. It suggests that there is a stable equilibrium relationship between the prices in the long run, where the two variables moves and fluctuate together. The data will be further tested to establish that there is a causal relationship between both the variables.

Simultaneously, this paper also tries to establish that the oil and gold price relationship can be defined explicitly using a linear model. Thus the secondary hypotheses states that the relationship between the variables can be explained in a linear functional form.

Thus understanding and finding out the relationship between the oil and gold prices is highly important as of their economic significance, impacting the financial markets. The oil is key for industrial, transportation and energy sector, thus fluctuations in it's prices can affect the backbone of a country's economy. On the other hand, the gold is a reserve/hedge or a safe haven asset against the ever changing currency and inflation rates. The study on the long term relationship and the key findings can be a base of other works related to the field of finance, investment management, insights of risk management techniques, thus the functioning of the global economy.

2. Literature Review

The literature on this work is on a varied scale. Using ordinary regression models with correlation analysis on bivariate time series data by early works of (Pindyck & Rotemberg, 1990) established the evidences of the co-movements in the prices time series of gold and oil in international market. (Samanta & Zadeh, 2012) used the time series data for the gold, crude oil prices, and the US stock exchange rates, established significant correlation. Then one of the works of (Basit, 2013), pointed out no co-movement between the Gold and Oil prices in Pakistan.

Then using the multivariate linear time series and co-integration analysis, the co-movement is further worked on by (Zhang & Wei, 2010), (Simakova & Šimáková, 2016), (Y.-H. Lee et al., 2012). Where, (Zhang & Wei, 2010) proved that there exists a positive correlation with coefficient 0.9295 between the prices of gold and crude oil. The author also established that the price of the crude oil Granger caused on the volatility of the prices of the gold. In order to examine asymmetric cointegration and causal linkages between West Texas Intermediate Crude Oil and gold prices in the futures market, Lee et al. (2012) used the momentum threshold error-correction model with generalized autoregressive conditional heteroscedasticity. The study looked at information between May 1, 1994, and November 20, 2008. Their empirical findings demonstrated that there is an asymmetric long-run adjustment between oil and gold. The causation relationship also demonstrated that West Texas Intermediate Crude Oil plays a significant role. (Simakova & Šimáková, 2016) concentrated on the correlation between the levels of the oil and gold prices. The work used synthesis and analysis techniques to combine theoretical knowledge from books, journals, and other sources. The study demonstrated that the investigated variables had a lasting association. Based on these works, this research paper is further developed and experimented.

3. Data definition

The data is obtained from Federal Reserve Economic Data (FRED) webpage. FRED is an online database maintained by Federal Reserve Bank of St. Louis, MO, USA.

The gold price dataset:
Link: https://fred.stlouisfed.org/series/PCU2122212122210
Max time series range: 1985-06-01 - 2017-12-01.
Frequency: Monthly.

The oil dataset:
Link: https://fred.stlouisfed.org/series/DCOILBRENTEU
Max time series range: 1987-05-20 - 2023-06-05.
Frequency: Daily.

The price of oil is calculated in USD per Barrel, meanwhile price of gold is in the Producer Price Index (PPI).
The data transformations are done in a very simplistic approach to keep data as close to the original as possible. The very first step included in changing the column names to 'DATE', 'OIL' and 'GOLD' to keep the same notation and for the code to be easier to read and specifying the date column with *"%Y-%m-%d"* format using the *as.Date()* function. In order to create .xts object, the 'DATA' columns are transformed in both data sets from character to date type. This allowed to merge both sets by date. Because both prices were collected with different frequency (and also because oil prices were sometimes missing) merged data set had a lot of missing values. To omit them, we saved both columns as numeric (this created *NA*s where prices were missing) and used *na.omit* function. Now an easy to read .xts dataset was created, with all the necessary values data set ready to be analysed with approximate span of 20 years containing 2 variables with 234 monthly observations.

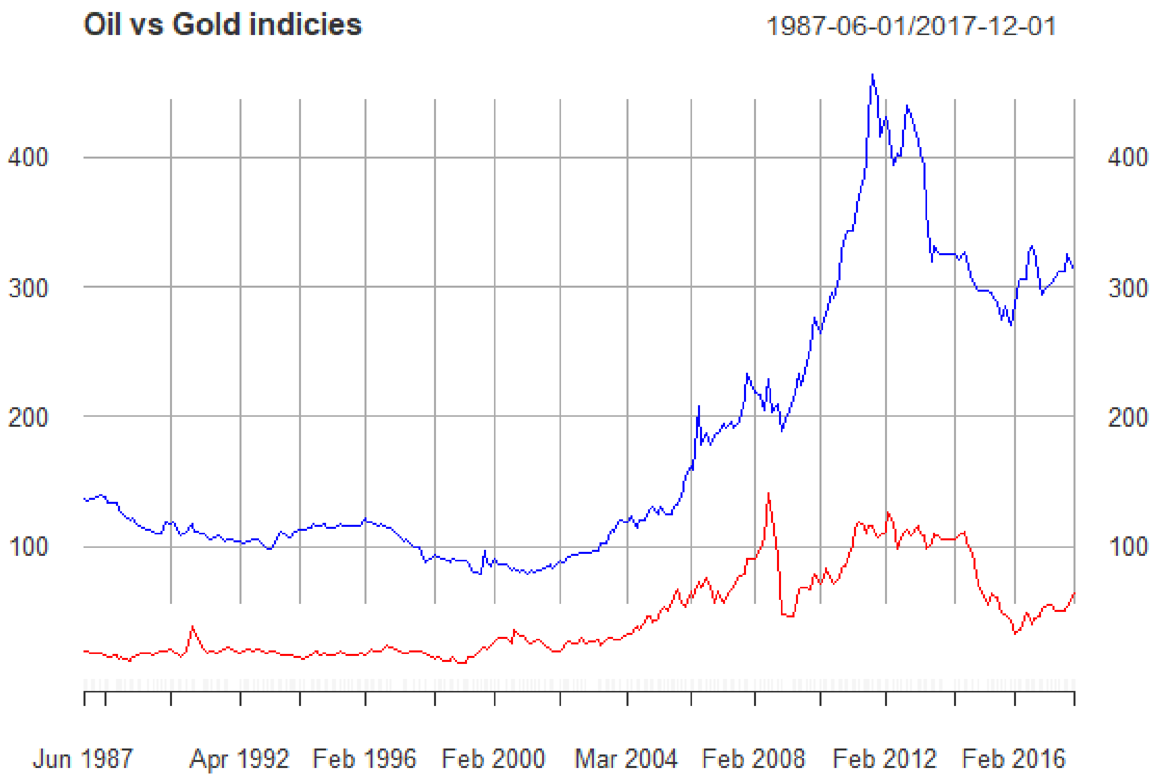

Fig1. Oil and Gold Prices: Monthly frequency

Red lines visualizes the change of the oil price and the blue is the change of gold prices. At first glance data looks very promising - it is not exactly in the same shape, but it is somewhat similar - especially spikes are in similar periods. That gives us the reason to dive deeper into the relationship between those variables. Evidently the data is not stationary as seen on the graph. To proceed with co-integration, confirming stationarity conditions, further discussions will be carried out in the next section.

4. Methodology

The analysis begins with simple data preparation, and after necessary changes done in Data definition section, the conitegration analysis is ready to be performed. The general to specific (GETS) model in this dataset containing only two variable is not directly applicable as, GETS is a iterative process of adding and removing variables to refine the model criteria.

As per the definition of Co-integration, it happens if there exists two or more time series which are non-stationary and at the same time exhibit stationarity at same integrating order (first differences), have a long term equilibrium, moves in a way that the stationary time series is obtained by their linear combination.

$$\beta Y_t = \beta_1 y_{1t} + \beta_2 y_{2t} + \beta_3 y_{3t} \sim I(0)$$

The above equation implies that y_1, y_2, y_3 are nonstationary when they are independent but their linear combination leads to stationarity at integrating order 0. The β implies the cointegrating vector dictating the cointegrating series combination.

The selected dataset is checked and confirmed that they are non-stationary using the *testdf2* function described at the Appendix. The *p_adf* values are greater than 5% significant level, pointing out that the time series are non-stationary condition and failing to reject the null hypotheses.

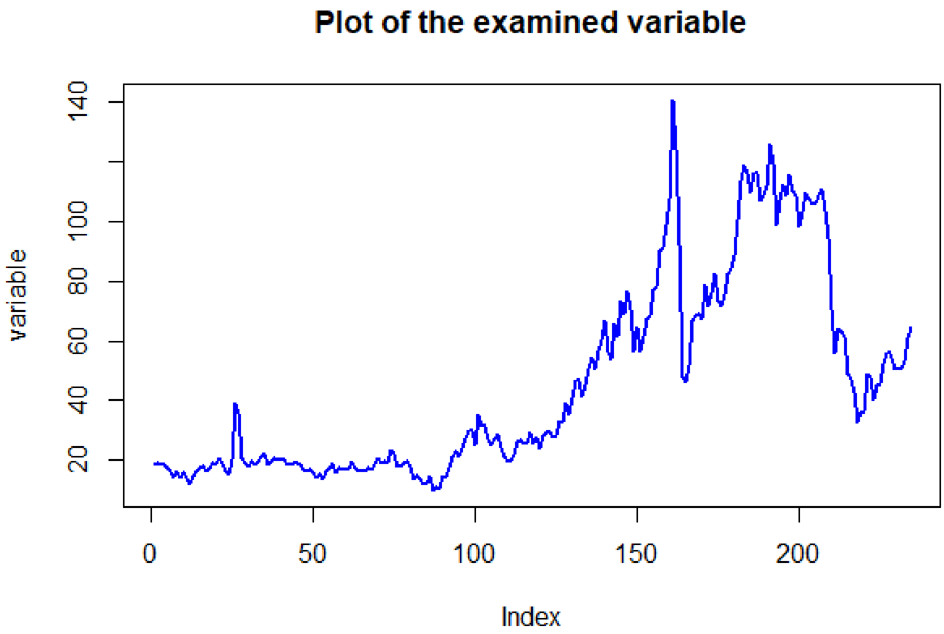

Fig 2. Non-Stationarity in oil data

Thus the first difference is performed using *diff.xts()* creating the 'dOIL' column in 'merged_ts' dataset and the same ADF test is performed. This time, the stationarity condition is achieved at second augmentation, integrating the data from Oil time series at Order one.

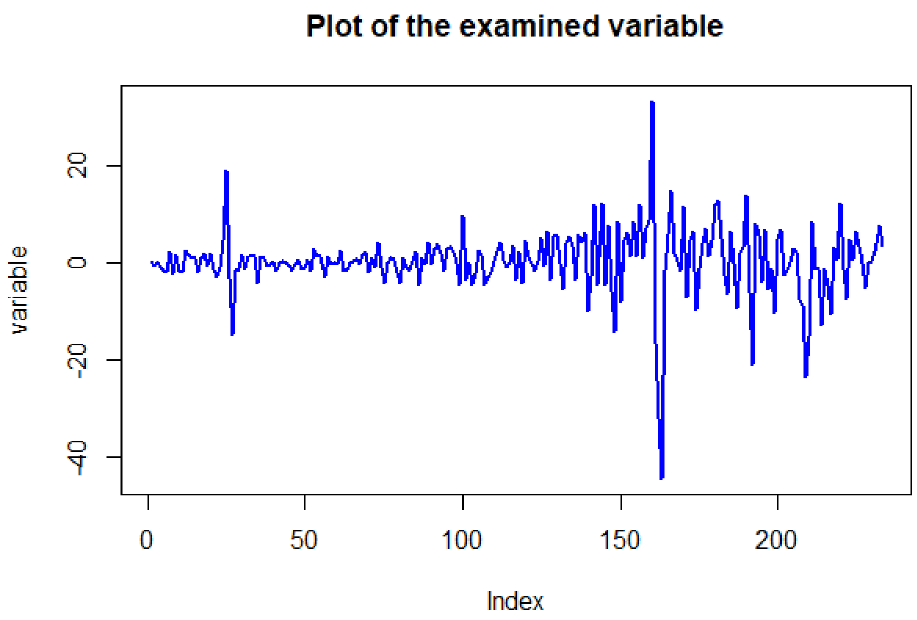

Fig 3. Stationarity in oil data of first difference

The same steps are followed for the 'GOLD' column in 'merged_ts', where initially the *p_adf* values are greater than 5% significance level, failing to reject the null hypotheses of non-stationarity. Simultaneously, the ADF test on the first difference is performed. The stationarity condition is achieved at eighth augmentation, integrating the data from Gold time series at Order one.

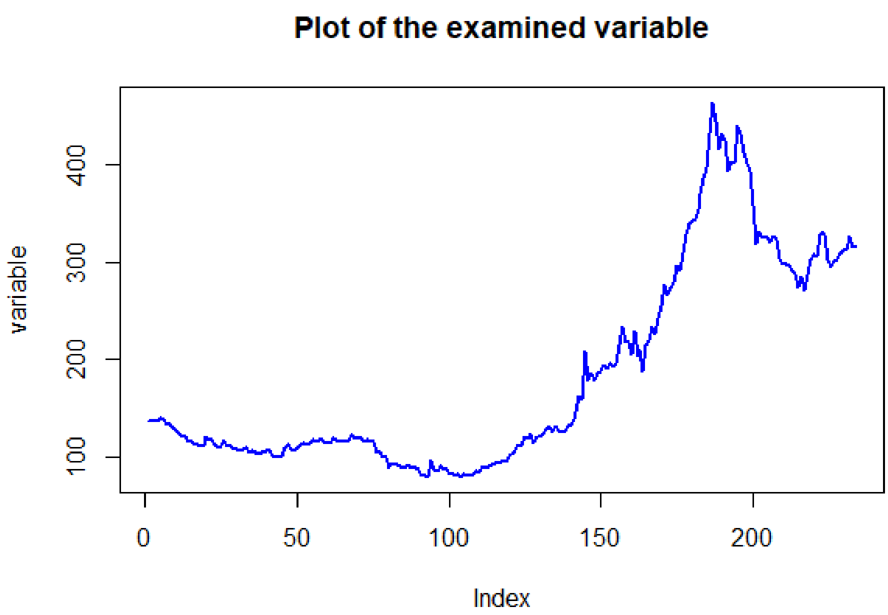

Fig 4. Non-Stationarity in gold data

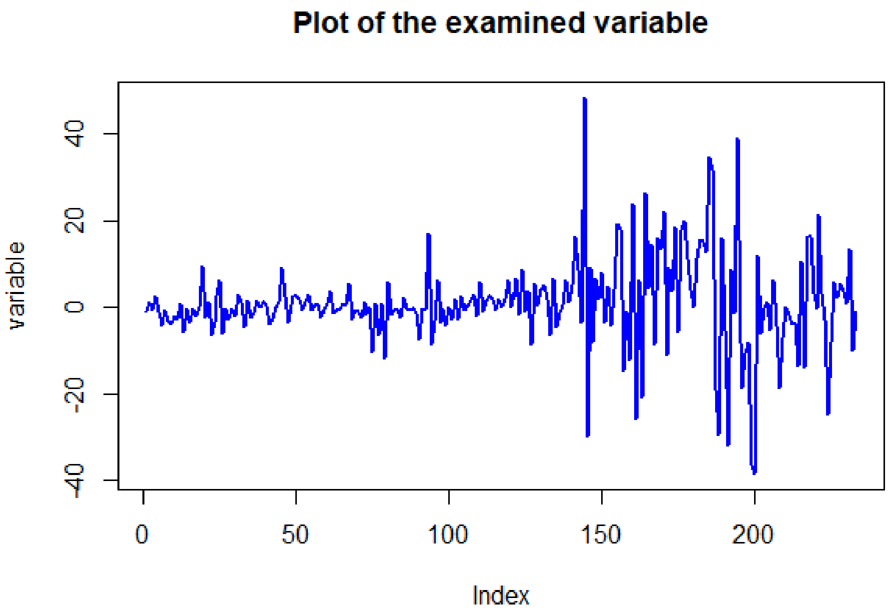

Fig 5. Stationarity in gold data of first difference

1. Cointegration Test

Confirming the stationarity at same integrating order for both variables, the next step is to check if their linear combination is stationary. For this process, the Engle-Granger Cointegration Test is performed. At first the linear regression model is prepared.

```
model.coint <- lm(OIL~GOLD, data = merged_ts)
```

Then the residuals from this model is tested for stationarity.

```
testdf2(variable = residuals(model.coint), test.type="nc",
        max.augmentations = 9, max.order = 9)
```

From the test it was found out the residuals are stationary at the test statistics -3.727834 at first augmentation. Checking the Critical Value Table provided during the lab,

Number of Variables $N+1$	Sample Size	Critical Values 10%	5%	1%
2	50	3.28	3.67	4.32
	100	3.03	3.37	4.07
	200	3.02	3.37	4.00

As the number of observation is 234 and number of variables is two, at 5% significance level the critical value –(3.37) is considered. As the test statistics -3.727834 falls in the rejection zone, the null hypotheses is rejected in favour of the time series being stationary. Thus it can be said the linear combination of two non-stationary time series is stationary, confirming co-integration.

Checking the residual summary, the cointegrating vector can be interpreted as:

$$[1, 0, -0.27005(Gold)]$$

This vector defines the linear combination of the Gold and Petrol prices after stationarity condition is proven. It can be interpreted as follows:

In the long term, if the price of gold increases by a unit, then the price of petrol also increases by 0.27005 unit. Thus defining the long term relationship between the variables.

2. Error Correction Model

The error correction mechanism is performed and tested to check the short term dynamics of the model. According to Error Correction Mechanism model:

$$\Delta Oil = \beta_1 \Delta Gold + \beta_2 \, (lagged\ residuals) + \epsilon$$

After running the model, the obtained equation is:

$$\Delta Oil = 0.1913 \Delta Gold - 0.07114(lagged\ residuals) + \epsilon$$

Here, the value 0.1913 is known as the short term dynamics, which can be interpreted as follows: if $\Delta Gold$ is increased by one unit, the ΔOil will be increased by 0.1913 unit.

$-0.07114(lagged\ residuals)$ describes that if unitary shock is introduced for the steady state relationship between oil and gold, the system goes out of the steady state. After one period, the system will correct itself by7.011% of the shock. This is the error correction mechanism where the system teaches itself on previous error and corrects itself on the next one. Thus for this case the system can take approximately 14 periods to return to the steady state. As the coefficient of lagged residual is the interval of the range [-2,0], it means that the shock will be corrected and the error correction mechanism works. Thus the error correction vector can be represented by:

$$[1, -0.19313, 0.07114]$$

In terms of short term relationship, the unit change in ‘dGold’ will lead to 0.19313unit positive change in oil. This is backed up by the a very small p-value confirming a strong relationship between the variables.

3. Granger Causality Test

The Granger causality test is used to determine the causal relationship between variables. It is used to predict the possible outcomes whether one variable can be assigned to predict the other, which indicates the potential causal relationship.

For this study, the H0 null hypotheses was that the lags of oil are insignificant, i.e. does not have the power to prove causal relation with gold. At lag3, the model result failed to reject the null hypotheses. But at lag 4, the H0 hypotheses is rejected and it is evident that the Oil Granger causes the Gold, which means that the past values of Oil can be used as a source of information to predict future Gold values.

```
grangertest(x=merged_ts$dOIL, y=merged_ts$dGOLD, order = 4)
```

Similarly testing the other way, it was found that the Gold Granger can cause the Oil, i.e. past values of Gold can be a source of useful information to predict future Oil values.

```
grangertest(x=merged_ts$dGOLD, y=merged_ts$dOIL, order = 4)
```

4. Ramsey Reset Test

The Ramsey Reset test is used to check the form of function of a model i.e. it asses if including additional powers to independent variables get the model improved fit.

```
resettest(model.coint, power = 2, type="fitted")
```

The null hypotheses of the test is that any tested model is correctly specified. In this case, the p value is 0.1613 which is greater than the 5% significance level, thus failing to reject the null hypotheses and concluding that there is least to none specification errors in the regression model ‘model.coint’.

5. Breusch-Pagan's test for homoscedasticity

Continuing the Breusch-Pagan's test, the model is checked for heteroscedasticity. It is used to examine if the residuals of a given model is exhibiting the unequal variances for different levels of independent variables. So the H0 null hypotheses for this test is homoscedasticity which means that the residuals has constant variance. For this case:

```
bptest(model.coint, studentize=TRUE)
```

the p value is 0.001499, rejecting the null hypothesis in favour of the variance being heteroscedastic.

6. Breusch-Godfrey test

The Breusch Godfrey test is used to check for autocorrelation in the residuals for a regression model. The serial correlation violates the independence of the residuals, suggesting strong correlation. For this case:

```
bgtest(model.coint, order = 1)
```

The p-value is extremely small, indicating strong correlation up to order 1 for the 'model.coint', rejecting the null hypothesis.

5. Results and interpretation

Based on the tests, model tuning and performance, it is evident that there is a stable long term relationship between the prices of Gold and Oil. The cointegration is established and tested. Thus verifying that there is a stable equilibrium relationship between the commodity prices as it moves and fluctuates together, supporting the primary hypotheses of the stable relationship. Moreover, the effects of shock being mitigated at 14 periods, effectively capturing the short term dynamics of the model was proved by the Error Correction Mechanism. The ECM also adds up with the results from Ramsey Reset Test where at 2nd or 3^{rd} power, the model concluded no specification errors, i.e. the functional form is correctly specified for the regression model, which supports the secondary hypotheses of the relationship can be explicitly defined using linear model. Moreover, the causality tests also provided strong evidence that both of the variables cause each other at 4^{th} order. This means that past information of one variable can be used to predict the future of other.

However, the Breusch-Pagan and Breusch-Godfrey tests were failed, which indicated presence of unequal variances and serial correlation in residuals of the model. This is contradictory to the secondary hypotheses to define the relationship simply using the linear regression model. Thus there might be question on reliability of the estimated coefficients, which might further lead to biased parameter estimates.

	H0	**Test statistics**	**aug**	**p_adf**
Oil	Fail to reject	-1.97	1	0.32
dOil	Rejected	-10.93	2	0.01
Gold	Fail to reject	-0.28	1	0.918
dGold	Rejected	-3.08	8	0.01
model.coint	Rejected	-3.73	1	0.01

Table 1. Cointegration analysis table (Unit Root Test)

	p-value	**H0**
Ramsey Reset Test	0.1613	Fail to reject
Breusch-Pagan's Test	0.001499	Rejected
Breusch-Godfrey Test	0.2e-10	Rejected

Table 2. Test results

6. Findings

With the establishment of the cointegration analysis, the primary hypotheses of long term stable relation between the oil and gold prices are validated. The secondary hypotheses is kind off having contradictory conclusions from the results. The ECM and Ramsey Reset test though confirms the oil and gold price relationship being defined by a linear model, but the failure of Breusch-Pagan and Breusch-Godfrey tests indicated the presence of unequal variances and serial correlation in residuals of the model which can question the reliability of the estimated coefficients, which might further lead to biased parameter estimates.

Thus to mitigate this concerns, alternative regression techniques needs to be later employed that can account for the autocorrelation and heteroscedasticity. The ARCH/GARCH models can be applied for heteroscedasticity and autoregressive models for autocorrelation. Thus by addressing these points and re-estimating the model, further robust validation can be achieved for the hypotheses verification.

Bibliography

Basit, A. (2013). Impact of KSE-100 index on oil prices and gold prices in Pakistan. *IOSR J Bus Manag (IOSR-JBM)*, *9*(5), 66–69.

Gil-Alana, L. A., Yaya, O. S., & Awe, O. O. (2017). Time series analysis of co-movements in the prices of gold and oil: Fractional cointegration approach. *Resources Policy*, *53*, 117–124. https://doi.org/https://doi.org/10.1016/j.resourpol.2017.06.006

Lee, W.-C., & Lin, H.-N. (2012). Threshold effects in the relationships between USD and gold futures by panel smooth transition approach. *Applied Economics Letters*, *19*(11), 1065–1070.

Lee, Y.-H., Huang, Y.-L., & Yang, H.-J. (2012). The asymmetric long-run relationship between crude oil and gold futures. *Global Journal of Business Research*, *6*(1), 9–15.

Pindyck, R. S., & Rotemberg, J. J. (1990). The excess co-movement of commodity prices. *The Economic Journal*, *100*(403), 1173–1189.

Samanta, S. K., & Zadeh, A. H. M. (2012). *Co-movements of oil, gold, the US dollar, and stocks*.

Simakova, J., & Šimáková, J. (2016). *Analysis of the Relationship between Oil and Gold Prices*. https://www.researchgate.net/publication/266005958

Zhang, Y.-J., & Wei, Y.-M. (2010). The crude oil market and the gold market: Evidence for cointegration, causality and price discovery. *Resources Policy*, *35*(3), 168–177. https://doi.org/https://doi.org/10.1016/j.resourpol.2010.05.003

Appendix

1. The Main Code

```
setwd("C:\\Users\\mraer\\Desktop\\UW\\Semester-2\\Advanced
Econometrics\\New_Eco")
Sys.setenv(LANG = "en")


library(xts)
library(lmtest)
library(fBasics)
library(urca)
library("lmtest")
library(ggplot2)
source("function_testdf2.R") # <- Find it on section 2



dataset1 <- read.csv('Aurum.csv')
dataset2 <- read.csv('petrol.csv')



#dataset1 <- dataset1[, c(1, 5)]

colnames(dataset1) <- c('DATE', 'GOLD')
colnames(dataset2) <- c('DATE', 'OIL')

dataset1$DATE <- as.Date(dataset1$DATE, format = "%Y-%m-%d")
dataset2$DATE <- as.Date(dataset2$DATE,  format = "%Y-%m-%d")

merged_ts <- merge(dataset1, dataset2, by = "DATE", all = TRUE)


merged_ts$OIL <- as.numeric(merged_ts$OIL)
merged_ts$GOLD <- as.numeric(merged_ts$GOLD)

merged_ts <- na.omit(merged_ts)

#merged_ts <- merged_ts[seq(1, nrow(merged_ts), 2), ]

#merged_ts <- merged_ts[1:500,]

merged_ts <- xts(merged_ts[, -1], merged_ts$DATE)


merged_ts$dOIL <- diff.xts(merged_ts$OIL)
merged_ts$dGOLD <- diff.xts(merged_ts$GOLD)

plot(merged_ts$OIL,
     main = 'Oil vs gold indicies', lwd = 0.7, col = "red")

lines(merged_ts$GOLD, col = "blue", lwd = 0.5)
```

```
# legend("topleft", c("Mortgage", "Inflation"),
#        col = c("black", "blue"), lty = 1)

testdf2(variable = merged_ts$OIL, test.type = "c",
        max.augmentations = 12, max.order = 12)

testdf2(variable = merged_ts$dOIL, test.type = "nc",
        max.augmentations = 12, max.order = 12)

# 3rd Row = Aug 2 Order 1


testdf2(variable = merged_ts$GOLD, test.type = "c",
        max.augmentations = 12, max.order = 12)

testdf2(variable = merged_ts$dGOLD, test.type = "nc",
        max.augmentations = 12, max.order = 12)

# 9th Row = Aug 8 Order 1


model.coint <- lm(OIL~GOLD, data = merged_ts)
summary(model.coint)

testdf2(variable = residuals(model.coint), test.type="nc",
        max.augmentations = 9, max.order = 9)

# Row 2 thus 1 augmentation (m=1) test statistics = -3.727834

length(residuals(model.coint)) #=234

#Checking the chart provided by Prof. Rafal, at 5% (N=2) the critical
value is -3.37. Our test statistic = -3.727834 is in the rejection
zone/ intervl
#Thus we reject the NUll Hypothesis that the TS is non-stationary and
conclude Stationary TS with
#Co-integration. So price of Gold and Petrol are cointegrated

### Cointegratiing Vector: We check residual summary
#[1, 3.70444 (intercept), -0.27005(Gold)] <- Remember opposite sign of
there
#petrol = -3.7044+0.27005*Gold + Epsilon

#This vector defines the linear combination of the Gold and Petrol
price that is stationary.
#This also defines the long term relationship between the petrol and
gold price
#In the long term, if Gold increases by 1 unit then Petrol increases by
0.27005 unit


####################################Cointegration Model: ERROR
Correction

#1. Creating Lagged Residuals

merged_ts$lresid <- lag.xts(residuals(model.coint))
```

```
#2. Estimating ECM

model.ecm <- lm(dOIL~dGOLD + lresid -1, data = merged_ts)
summary(model.ecm)

# Error-Correction Mechanism model
# delta(PETROL)=beta_1*delta(GOLD)+beta_2*(lagged residuals) + epsilon
# delta(PETROL) = 0.19313*delta(GOLD) -0.07114*(lagged residuals) +
epsilon

#0.19313 is known as the short term dynamics. If delta(GOLD) is
increased by one unit, delta(PETROL) will be increased 0.19313 unit

# -0.07114*(lagged residuals) descrbes if there is a unitary shock to
the relation between Petrol and Gold in steady state solution, the
system goes out of the steady
#state solution.After one period, the system will correct itself by
7.011% of the shock
# THIS IS THE ERROR CORRECTION MECHANISM as the system Teaches on the
previous errors and corrects on the next period.So the system needs
approximately 14 periods to
# return to the steady state solution

#The Lagged Residual Parameter (coefficient alpha_2) should be within
the interval [-2,0]. It means that the shock is corrected and error
correction mechanism works

# lresid = petrol +3.70444 -0.27005*Gold <- From Cointegrating Vector
Model


## LONG TERM RELATION: described here by Cointegrating vector
## Short TERM RELATION: described here by ECM Vector


########################### General to Specific (GETS Approach)

# Granger Causalty is used to determine the Causal relationship between
the variables based on their predictive power.
# There there are only two variables, GETS using backward or forward
elimination is not possible.
# Rather, Granger Causalty can determine the possible outcome where one
variable can be used to predict another variable, indicating a
potential causal relationship.
# It helps in the process of model selection by identifying the lagged
variables that have predictive power for the dependent variable. By
examining the significance
# of the lagged variables in the model, you can determine if they
contribute to the explanatory power of the model and provide evidence
of causal relationships.

names(merged_ts)

grangertest(x=merged_ts$dOIL, y=merged_ts$dGOLD, order = 3)
# H0: the first, the second, and the third lags of OIL are
insignificant
# OIL lags are inignificant
```

```
# OIL does not cause GOLD:  where OIL variable cannot be used to
predict another variable GOLD

grangertest(x=merged_ts$dOIL, y=merged_ts$dGOLD, order = 4)
# Now at Lag3, Price of Oil is helpful to explain the model Price of
Gold. So Price of Oil is a cause to Price of Gold
#H0 is rejected > OIL lags at order 3 are significant

# Checking the other way:
grangertest(x=merged_ts$dGOLD, y=merged_ts$dOIL, order = 3)
# H0: the first, the second, and the third lags of GOLD are
insignificant
# GOLD lags are inignificant
# GOLD does not cause OIL:  where GOLD variable cannot be used to
predict another variable OIL

grangertest(x=merged_ts$dGOLD, y=merged_ts$dOIL, order = 4)
# Now at Lag3, Price of GOLD  is helpful to explain the model Price of
OIL. So Price of GOLD is a cause to Price of OIL
#H0 is rejected > GOLD lags at order 3 are significant


##^^^^^^^^^^^^^^^^^^^^^  CAUSAL RELATIONSHIP BETWEEN BOTH VARIABLE AT
ORDER 3 IS ESTABLISHED


################### Ramsey Rest TEST: CHECKING Linearity of the Model

resettest(model.coint, power = 2, type="fitted")

## p-value = 0.1613/0.9 if power is 3 > above 5% significance level >
Failing to Reject the Null Hypothesis and establishing that the model
is okay
## ALSO The Ramsey Reset test is typically used to detect
nonlinearities in the relationship between the dependent variable and
the independent variables in a linear regression framework.
## SINCE Cointegration is already established, it is not actually
fruitful. As long term and short term relationship was established,
this test can be omitted.


###################### Breusch-Pagan’s and White’s tests-
homoscedasticity

bptest(model.coint, studentize=TRUE)

# p-value = 0.001499 > null hypothesis of homoscedasticity is rejected
in favour of heteroscedasticy
# it suggests that the assumption of constant variance is violated.
This can affect the reliability of the estimated cointegration
relationship.


#################### Breusch-Godfrey test-no autocorrelation

bgtest(model.coint, order = 1)
```

```
## p-value < 0.00000000000000022 > Null hypothesis is rejected: There
is strong autocorrelation
```

---

2. function_testdf2.R

```
require(fUnitRoots)
require(lmtest)


# order<-1
# max.order<-5
# variable<-SP500$SP500
# max.augmentations<-1
# augmentations<-0
testdf2 <- function(variable, test.type, max.augmentations, max.order)
      {
      results_adf <- NULL
      variable <- coredata(variable[!is.na(variable)])

      for(augmentations in 0:max.augmentations)
            {
            df.test_ <- adfTest(variable, lags = augmentations, type =
test.type)
            df_ <- as.numeric(df.test_@test$statistic)
            p_adf <- as.numeric(df.test_@test$p.value)
            resids_ <- df.test_@test$lm$residuals


            bgtest_<-list()
            bgodfrey<-list()
            p_bg<-list()


            for (order in 1:(max.order+1))
            {
            bgtest_[[order]] <- bgtest(resids_~1, order = order-1)
            bgodfrey[[order]]<- bgtest_[[order]]["statistic"]
            # names(bgodfrey[[order]]) <- NULL
            p_bg[[order]]<- bgtest_[[order]]["p.value"]
            }
            results_adf <- rbind(results_adf, data.frame(augmentations =
augmentations, adf = df_, p_adf = p_adf,
                                                  bgodfrey = bgodfrey,
p_bg = p_bg))
            rm(df.test_, df_, resids_, bgtest_, bgodfrey, p_bg)
            }

      results_adf <- results_adf[results_adf$augmentations >= 0,]

      row.names(results_adf) <- NULL

      plot(variable, type = "l", col = "blue", lwd = 2, main = "Plot of
the examined variable")

      results_adf$p_bg.p.value<-NULL
      return(results_adf)}
```