Empirical Analysis Report

Gwendoline Hays-Valentin & Hugo Michel

Professor: Catherine Bruneau

Financial Econometrics

-

M2 Finance Technology Data - Sorbonne School Of Economics

Université Paris 1 Panthéon-Sorbonne

Paris, France, December 2023

Contents

List of Tables

1 Empirical Analysis 1 : Analysis of series

1.1 Data

The objective of our study is to explore the complex relationships between these selected stock indices and the macroeconomic series. We aim to identify and analyze patterns, correlations, and potential cointegration relationships and causal effects among these time-series data.

To do so, we use the FRED API to download to download the inflation rate (10-Year Breakeven Inflation Rate) financial time series (T10YIE).

Then we chose to use the Yahoo Finance API to collect stocks data from the FedEx company and oil market prices. FedEx Corporation is an American company and airline specializing in international freight transport. The ticker of the stocks market FedEx company is ‘FDX‘ and the ticker of the oil stocks market is (‘CL=F‘).

Our objective is to analyze the impact of oil prices and inflation on the market price of an international logistics/freight transport company whose main raw material is oil.

Regarding the period we chose the following period: from 2012-12-31 to 2022-12-31. The period we chose allow us to analyze the evolution of oil stock price when extreme event (covid and war in ukraine) occurs. Such a time frame allows for a comprehensive study of the indices and economic indicators under different economic conditions: pre-crisis stability, the crisis period, and the post-crisis recovery phase. Analyzing data from this period provides critical insights into how stock indices and economic indicators behave under stress and can help in identifying early warning signs of financial instability.

The frequency of our time series is daily frequency for the two stocks price series and monthly frequency for some macro-economic time series (i.e inflation rate). These frequencies are sufficiently granular to observe the medium to long-term trends and reactions of the markets to macroeconomic changes, while also being detailed enough to pinpoint significant events and shifts in investor sentiment. To ensure that our time series are all based on the same frequency (i.e. a daily frequency), we resample the macroeconomic series to daily frequency by propagating the month’s value to all days of the month.

To sum up we have 2 daily stocks prices time series (Oil and FedEx company) and one macroeconomic time series which is the inflation rate at a daily frequency. We would like to show what are the interactions, en effects among these three time series.

Figure 1.1: Data

	CL=F	FDX	inflation_rate
2012-12-31	91.820000	81.141144	2.45
2013-01-02	93.120003	83.379295	2.48
2013-01-03	92.919998	83.688965	2.46
2013-01-04	93.089996	83.963188	2.48
2013-01-07	93.190002	83.795113	2.52
...	...	...	...
2022-12-23	79.559998	173.121033	2.20
2022-12-27	79.529999	174.223160	2.26
2022-12-28	78.959999	170.287018	2.28
2022-12-29	78.400002	172.747116	2.27
2022-12-30	80.260002	170.434616	2.30

2500 rows × 3 columns

IMPORTANT NOTE:

It's important to note that we will use these 3 time series during the Empirical Application 1, Empirical Application 2 and Empirical Application 3. Therefore, in this fist empirical application we fully analyze these time series. To do so, we proceed as follow:

- Analyze trend and seasonality of the time series
- Plot the series
- Analyze the stationarity: Check for unit root by running Augmanted Dickey Fuller Test
- Decompose the series and analyze each component of the series
- Analyze the ACF and PACF plots of each time series
- Check for Unit Root after taking first difference

Please note that analyses will not be apply a second time in the next Empirical Application. Indeed during the Empirical Application 2 and 3 we will directly dive in main tasks of this 2 and 3 Emiricial Application.

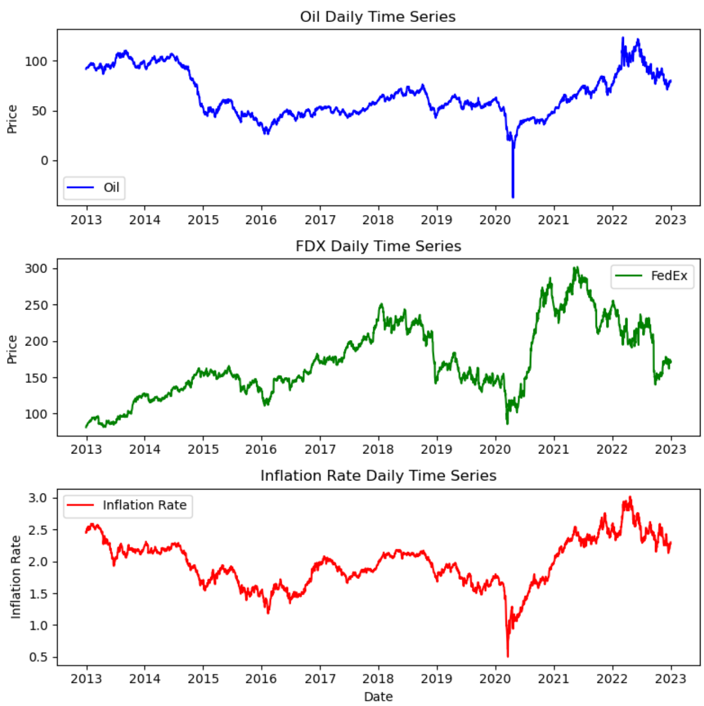
1.2 Time Series Analysis

To analysis our 3 time series we implement the following methodology:

- Plot the time series to have a preview of the pattern (i.e trend, seasonality,...) of the series
- Preprocessing the time series by applying the log transformation
- Decompose the time series in order to analyze each component of our 3 time series
- Analyze Seasonality thanks to ACF plot
- Check for the presence of unit root and deterministic trend by appyling Augmanted Dickey Fuller Test (ADF)
- Take the first difference
- Check for the presence of unit to assess whether after taking the first difference our time series are stationary
- Anlayze ACF and PACF plots after taking the first difference

P̂lotting the time series with daily frequency

Figure 1.2: Daily Time series



Visually, we can see that our series do not show any particularly strong positive or negative trends. In addition it's seems to be no periodic pattern.

Hence, these time series seems to follow a stochastic trend such as a random walks without constant and trend.

1.2.1 Transform the time series into *log*

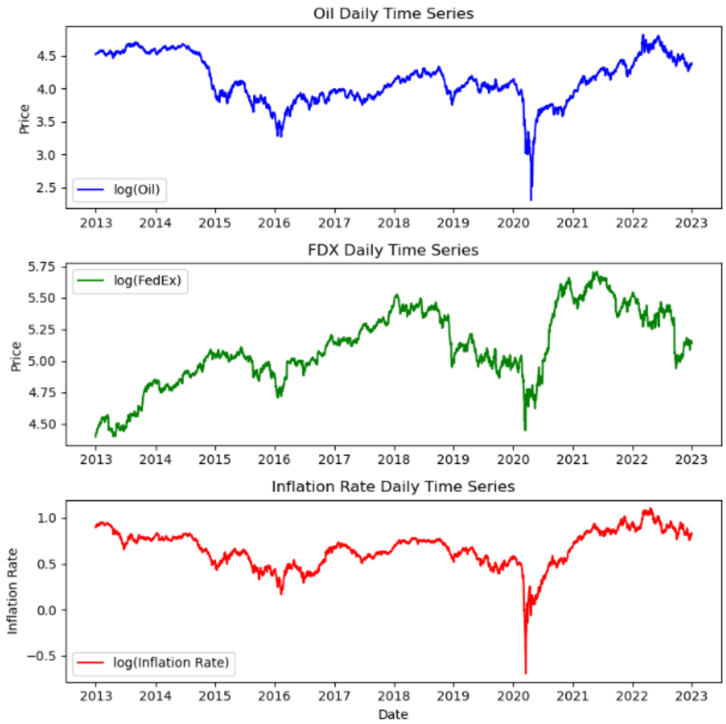
We apply the *log* operator to each time series in order to smoothing out the noise and extreme values of our time series. In fact by taking the log, we can stabilize the variance of our series, making the data more amenable to statistical analysis.

Figure 1.3: Daily Time series after log transformation

	CL=F	FDX	inflation_rate
2012-12-31	4.519830	4.396190	0.896088
2013-01-02	4.533889	4.423400	0.908259
2013-01-03	4.531739	4.427107	0.900161
2013-01-04	4.533567	4.430378	0.908259
2013-01-07	4.534640	4.428375	0.924259
...	...	...	...
2022-12-23	4.376511	5.153991	0.788457
2022-12-27	4.376134	5.160337	0.815365
2022-12-28	4.368941	5.137485	0.824175
2022-12-29	4.361824	5.151829	0.819780
2022-12-30	4.385271	5.138352	0.832909

2500 rows × 3 columns

Figure 1.4: Daily Time series after log transformation

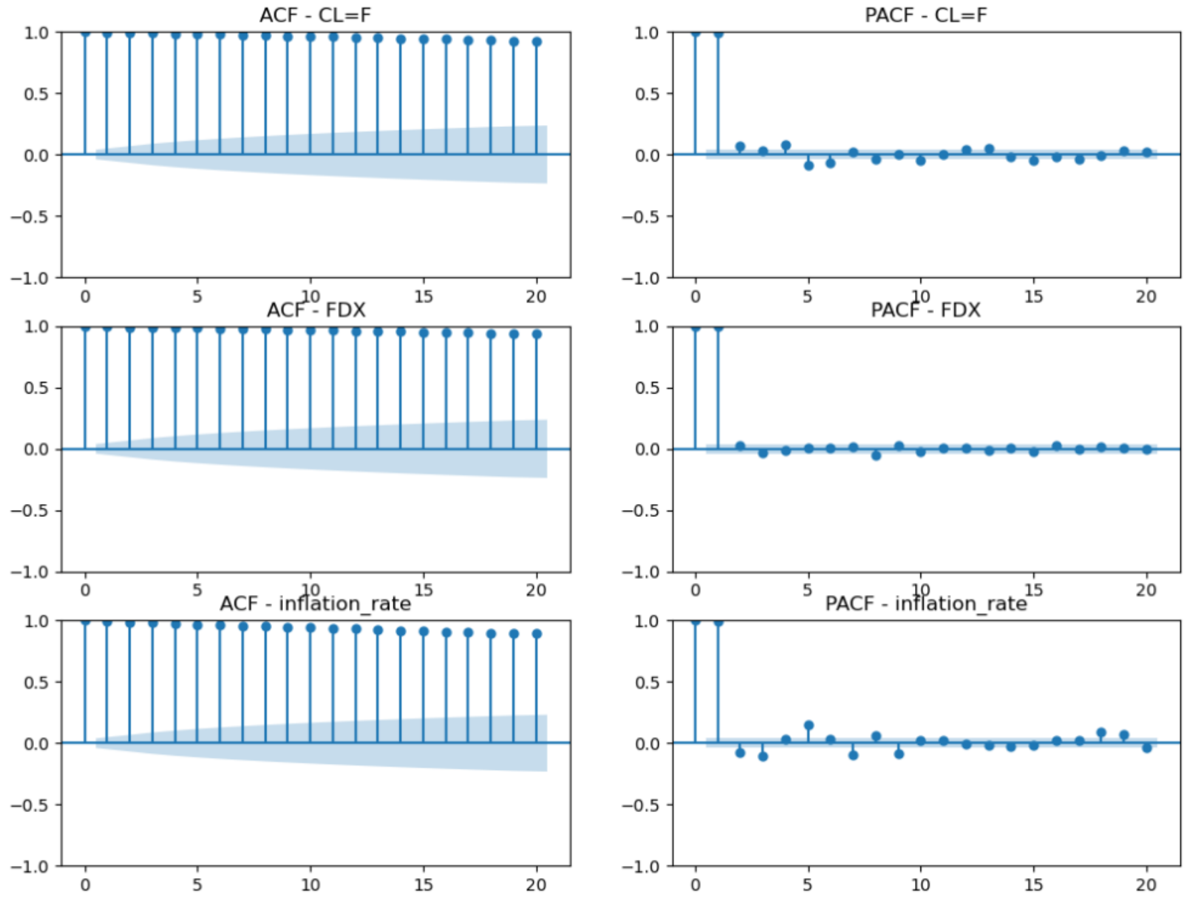


After *log* transformation, we can visually see that the time series appear to follow a stochastic process. They seem to follow random fluctuations, irregular patterns and no clear trend or seasonality. Overall, the data therefore appear erratic and may suggest a stochastic process.

We will confirm our assumptions by applying the relevant tests hypothesis and then decompose the series.

1.2.2 Plot ACF and PACF of log time series

Figure 1.5: ACF and PACF after log transformation

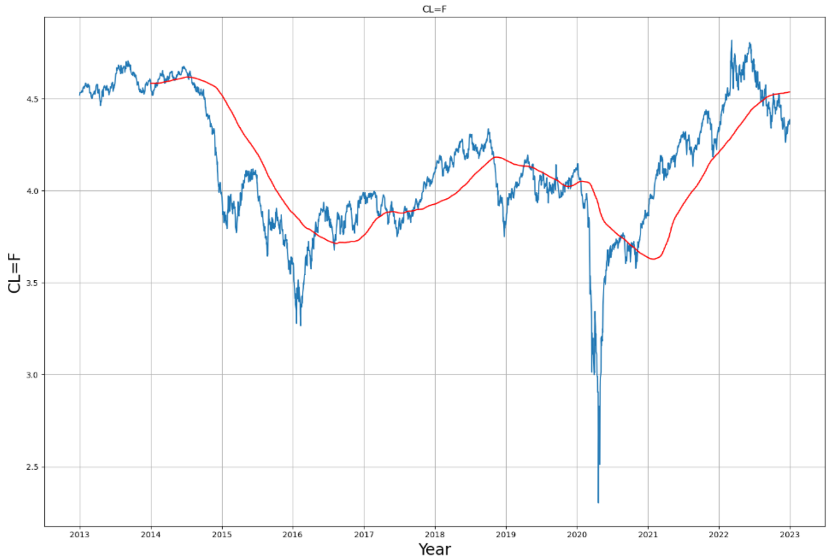


We can see that the ACF values are close to 1 from lags 1 to 20. This indicate a very strong positive correlation between the current value and the previous 20 values. In other words, the time series is highly correlated with its recent past. This suggests a strong memory or persistence in the data. In fact, the ACF values are high and decrease very, very slowly, this is a sign that the data are not stationary, so they need to be differentiated. But before to take the first difference we will perform a ADF test to test the presence of an unit root for each time series and confirm that our time teries are non-stationary and then follow a stochastic trend.

1.2.3 Plot trend of series with moving average

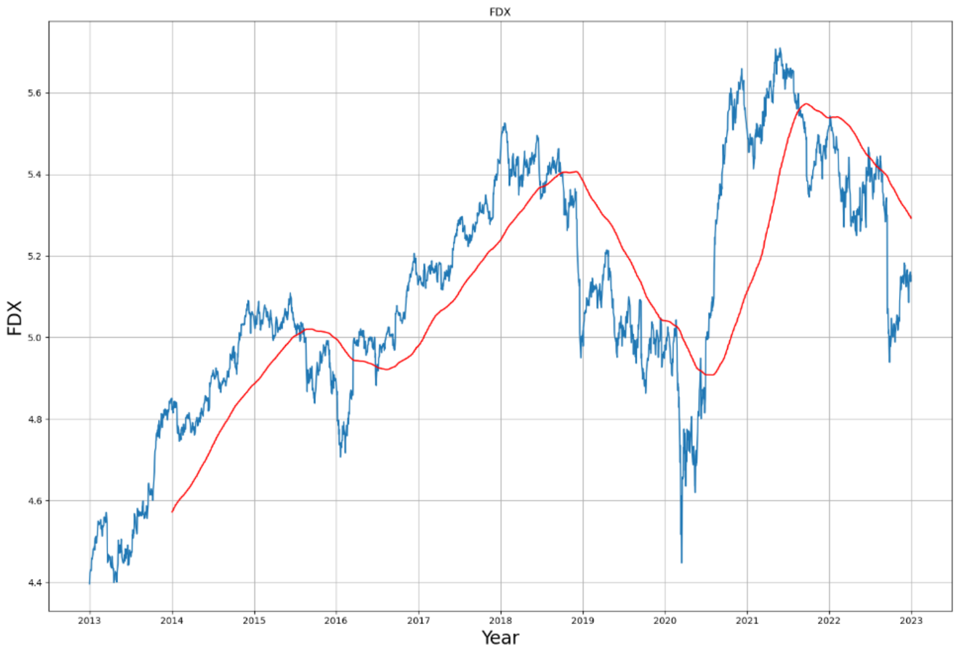
Oil stock price series

Figure 1.6: Oil stock price series with moving average



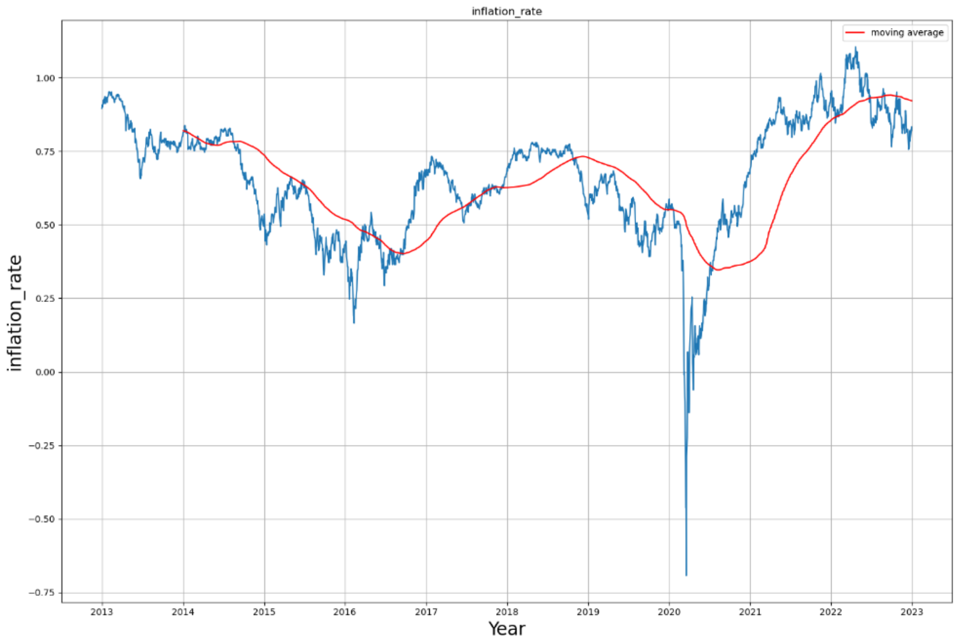
FedEx stock price series

Figure 1.7: FedEx stock price series with moving average



Inflation rate series

Figure 1.8: Inflation rate series with moving average

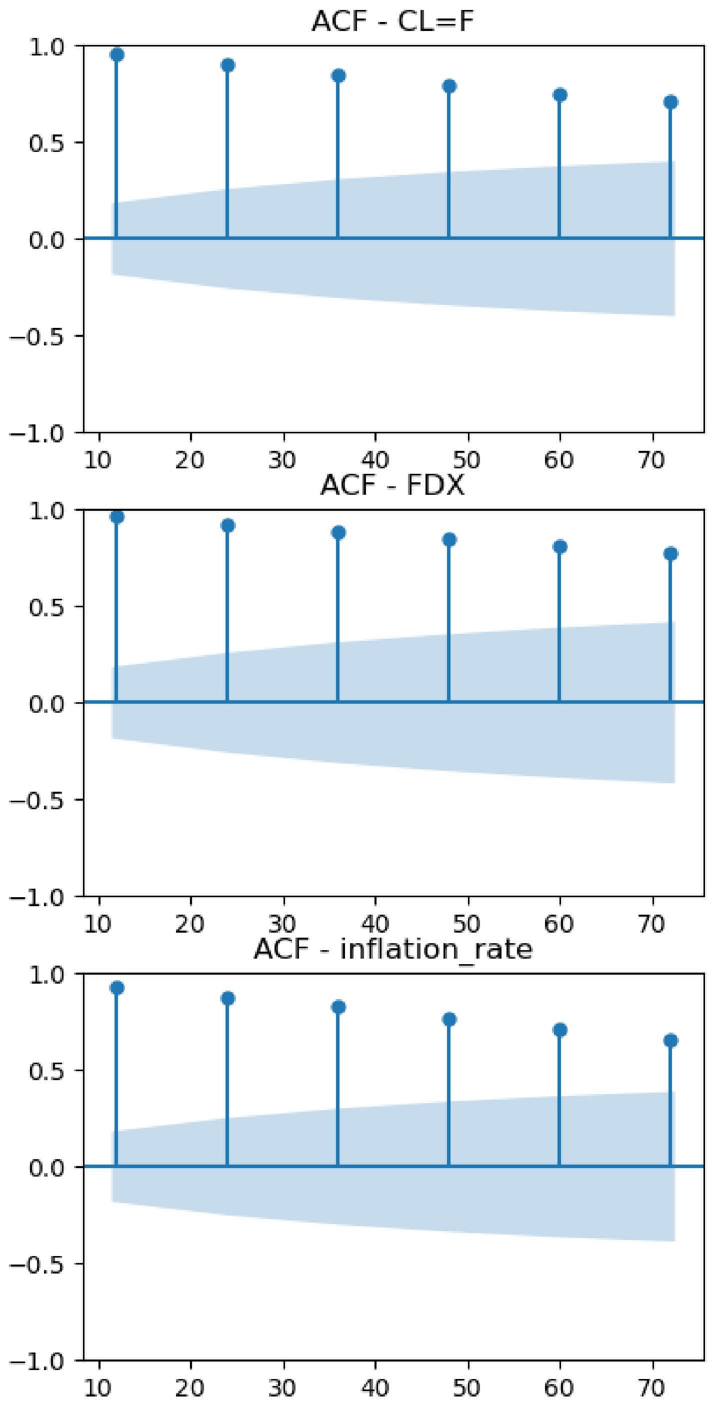


1.2.4 Identifying seasonal data using ACF

In order to break down the data, we need to know how often the cycles are repeated. Often wa can guess, but we can also use the ACF to identify the period.

To do so we set the following list of lags : $[12, 24, 36, 48, 60, 72]$. This list of lags is used as input of the ACF plot python functions. We chose the number of lags because we would like to see if there is seasonal pattern which are repeated each month.

Figure 1.9: Seasonal ACF plots



The ACF plots does not shows a periodic correlation pattern. Commonly, to find the period, we look for a lag greater than one, which is a peak in the ACF graph. Here, we can't see any a peak at a specific lags, which means that our time three time series doesn't have any seasonal pattern.

1.2.5 Perform ADF test without taking first difference

We perform an ADF test in order to test the presence of an unit root for each raw series.

To do so we consider the following hypothesis in the following order:

$\Delta X_t = b_0 + b_1 t + \rho X_{t-1} + \sum_{j=1}^{p-1} \phi_j \delta X_{t-j} + \epsilon_t$

where:

- b_0 : is the constant (i.e the drift) of the time series
- b_1 : is the trend of the time series
- ρ : is coefficient to test integrated order of the time series for example (i.e. $I(0)$)

REGRESSION: CONSTANT AND TREND

For this test we will implement first the most general regression:

$\Delta X_t = b_0 + b_1 t + \rho X_{t-1} + \sum_{j=1}^{p-1} \phi_j \delta X_{t-j} + \epsilon_t$

By performing this test, we assumes that the time series has a deterministic constant offset and a linear trend.

Test for deterministic trend:

- H_0 : The trend coefficient is not significant ($b_1 = 0$)
- H_1 : The trend coefficient is significant ($b_1 \neq 0$)

Otherwise if H_0 is rejected, then we accept H_1 and check the presence of unit root

Test for unit root with trend:

- H_0: There is a unit root (i.e. $\rho = 0$ non stationarity) with a significant trend. The time series is non-stationary with a deterministic trend.
- H_1: There is no unit root (i.e. $\rho \neq 0$ stationarity) no stochastic trend but a deterministic trend.

Otherwise if H_0 is not rejected (b_1 is not significant) we have to turn the regression with constant (without trend)

REGRESSION: CONSTANT

$\Delta X_t = b_0 + \rho X_{t-1} + \sum_{j=1}^{p-1} \phi_j \delta X_{t-j} + \epsilon_t$

By performing this test, we assumes that the time series has a deterministic constant offset but does not consider any trend.

Test for constant:

- H_0 : The constant coefficient is not significant ($b_0 = 0$)
- H_1 : The constant coefficient is significant ($b_0 \neq 0$)

If H_0 is rejected, then we accept H_1 (i.e constant coefficient is significant $b_0 \neq 0$) and test the presence of unit root with constant:

- H_0: There is a unit root (i.e. $\rho = 0$ non stationary) with a significant constant. The series is a random walk with a drift
- H_1: There is no unit root (i.e. $\rho \neq 0$ stationary) and no significant constant

Otherwise, if H_0 is not rejected (i.e constant coefficient is not significant $b_0 = 0$) we have to turn the third regression with no constant and no trend and test directly the presence of a unit root.

REGRESSION: NO CONSTANT AND NO TREND

$\Delta X_t = \rho X_{t-1} + \sum_{j=1}^{p-1} \phi_j \delta X_{t-j} + \epsilon_t$

By performing this test, we assumes that the time series does not have a deterministic constant offset or trend components.

- H_0: There is a unit root (i.e. non stationary $\rho = 0$). The series is a random walk
- H_1: There is no unit root (i.e. stationary $\rho \neq 0$). Tje series is stationary

The p-value is the probability to wrongly reject H_0

If the p-value are lower than the threshold 5%, hence we can reject the null hypothesis H_0 and accept H_1 and affirm that the time series is likely stationary

Otherwise we fail to reject the null hypothesis H_0 and don't accept H_1 and affirm that the time series may be non-stationary

ADF results

Figure 1.10: ADF test on Oil stock prices (CL=F)

```
############################################
ADF TEST: OIL SROCK PRICES  (log(CL=F))
############################################

______________________________
REGRESSION : CONSTANT AND TREND
______________________________
ADF Statistic: -2.4569387154315274
p-value for the trend: 0.3498148230685654

Trend coefficient is not sginificatif 0.3498148230685654 > 0.05
We fail to reject H0 (The trend coefficient is not signiciant (b1=0)
We turn to the second REGRESSION WITH CONSTANT

______________________________
REGRESSION : CONSTANT
______________________________
ADF Statistic: -2.5342535929265395
p-value for the constant: 0.10736090751335386
The constant coefficient is not significant (b0=0)
ADF test with Constant: p-value=0.10736090751335386 > 0.05.
We fail to reject the null hypothesis H0 (b0=0)
We turn to the third REGRESSION WITH NO CONSTANT AND NO TREND and test the presence of UR

______________________________
REGRESSION : NO CONSTANT AND NO TREND
______________________________
ADF test with No Constant and No Trend: p-value=0.5706582761076945 > 0.05. Fail to reject the null hypothesis H0, we can't acce
pt H1 at 95% confidence level.
The time series may be non-stationary. There is a presence of a unit root and stochastic trend.
The time series is a random walk without constant and deterministic trend
```

Figure 1.11: ADF test on FedEx stock prices (FDX)

```
############################################
ADF TEST: FedEx STOCK PRICES (log(FXD))
############################################

______________________________
REGRESSION : CONSTANT AND TREND
______________________________
ADF Statistic: -2.384692068826278
p-value for the trend: 0.3877881431337533

Trend coefficient is not sginificatif 0.3877881431337533 > 0.05
We fail to reject H0 (The trend coefficient is not signiciant (b1=0)
We turn to the second REGRESSION WITH CONSTANT

______________________________
REGRESSION : CONSTANT
______________________________
ADF Statistic: -2.3998545291046365
p-value for the constant: 0.14177747098194665
The constant coefficient is not significant (b0=0)
ADF test with Constant: p-value=0.14177747098194665 > 0.05.
We fail to reject the null hypothesis H0 (b0=0)
We turn to the third REGRESSION WITH NO CONSTANT AND NO TREND and test the presence of UR

______________________________
REGRESSION : NO CONSTANT AND NO TREND
______________________________
ADF test with No Constant and No Trend: p-value=0.8270176181906397 > 0.05. Fail to reject the null hypothesis H0, we can't acce
pt H1 at 95% confidence level.
The time series may be non-stationary. There is a presence of a unit root and stochastic trend.
The time series is a random walk without constant and deterministic trend
```

Figure 1.12: ADF test on the Inflation Rates

```
###########################################
ADF TEST: INFLATION RATE (log(inflation_rate))
###########################################

______________________________
REGRESSION : CONSTANT AND TREND
______________________________
ADF Statistic: -2.4275273589316124
p-value for the trend: 0.365089751853672

Trend coefficient is not sginificatif 0.365089751853672 > 0.05
We fail to reject H0 (The trend coefficient is not signiciant (b1=0)
We turn to the second REGRESSION WITH CONSTANT

______________________________
REGRESSION : CONSTANT
______________________________
ADF Statistic: -2.2904798809760902
p-value for the constant: 0.1750567854964667
The constant coefficient is not significant (b0=0)
ADF test with Constant: p-value=0.1750567854964667 > 0.05.
We fail to reject the null hypothesis H0 (b0=0)
We turn to the third REGRESSION WITH NO CONSTANT AND NO TREND and test the presence of UR

______________________________
REGRESSION : NO CONSTANT AND NO TREND
______________________________
ADF test with No Constant and No Trend: p-value=0.37886600904080786 > 0.05. Fail to reject the null hypothesis H0, we can't acc
ept H1 at 95% confidence level.
The time series may be non-stationary. There is a presence of a unit root and stochastic trend.
The time series is a random walk without constant and deterministic trend
```

Regarding the three time series, by analyzing the 2 first regression tests results CONSTANT AND TREND, CONSTANT we can see that for the 3 time series we fail to reject the null hypothesis H_0 and then we can't accept H_1 at 95% confidence level. Therefore, we apply the third regression test NO CONSTANT AND NO TREND and we can affirm that the 3 series may be non-stationary without constant and deterministic trend. There is a presence of a unit root and stochastic trend. Overall, our 3 series are a random walk without constant and deterministic trend. It is consistent with what we said previously (i.e the series follow a stochastic trend) by analysis visually the 3 time series.

We can deduce that we have to take the first difference for each time series and check the presence of a unit root (i.e. stochastic trend) again thanks to ADF test.

1.2.6 Decompose the series

We will decompose each series separately and analyze the different components.

The time series can be decompose with the following 3 components:

$x(t) = Tend + Seasonality + Residual$

Trend: The trend component represents the long-term, systematic, and often nonlinear movement in the data over time. It captures the underlying direction in the time series, whether it's increasing, decreasing, or remaining relatively constant. Trends can be caused by various factors, such as economic growth, population changes, or technological advancements. A time series with a clear trend component may exhibit a consistent upward or downward movement, which is not related to short-term fluctuations or seasonality.

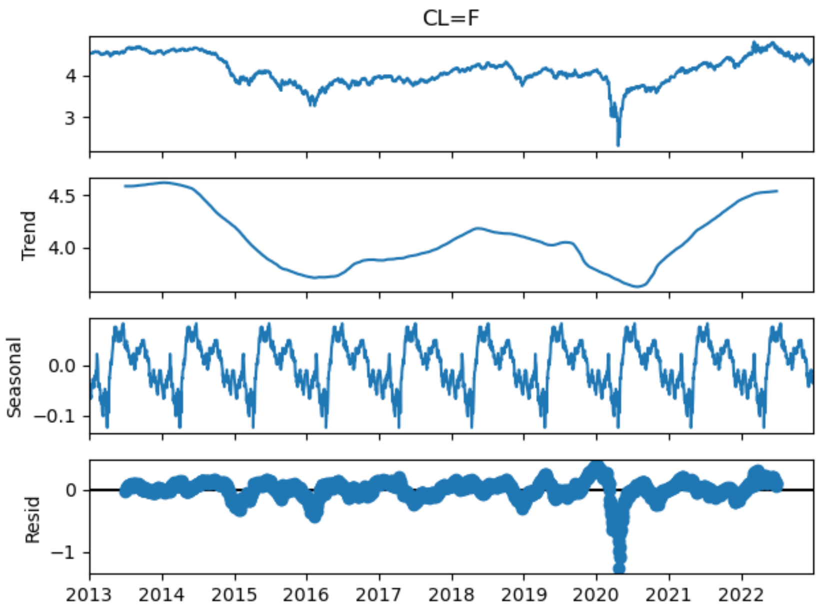
Seasonality: Seasonality refers to the regular, repeating patterns in the data that occur at fixed intervals. These intervals can be daily, weekly, monthly, quarterly, or any other specific time frame. Seasonality is often associated with external factors, such as the calendar (e.g., holidays), weather, or cultural events. Time series with seasonality will show periodic patterns that repeat within a particular time frame.

Residuals: The residual component, also known as the irregular component, represents the unexplained or random variation in the time series that cannot be attributed to the trend or seasonality. It includes noise, unexpected events, and other random factors. Residuals are essentially what remains after removing the trend and seasonality components. Analyzing the residuals is important because they contain valuable information about the inherent uncertainty and unpredictability in the data.

To decompose the series we choose the following period 252 for daily jobs.

Oil stock prices series decomposition

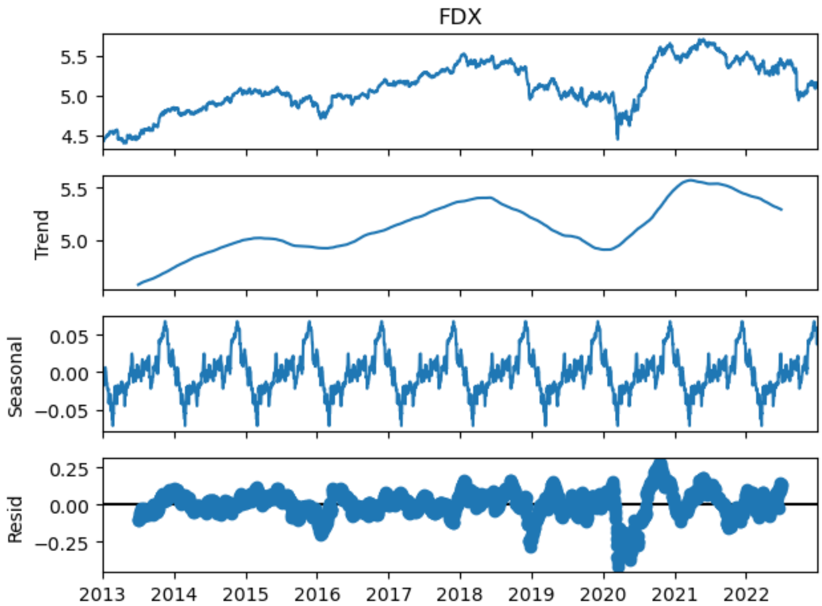
Figure 1.13: Oil stock prices series decomposition




```
------------------------
Are the Residuals a WN ?
------------------------
Residuals is NOT a WN
```

FedEx stock prices series decomposition

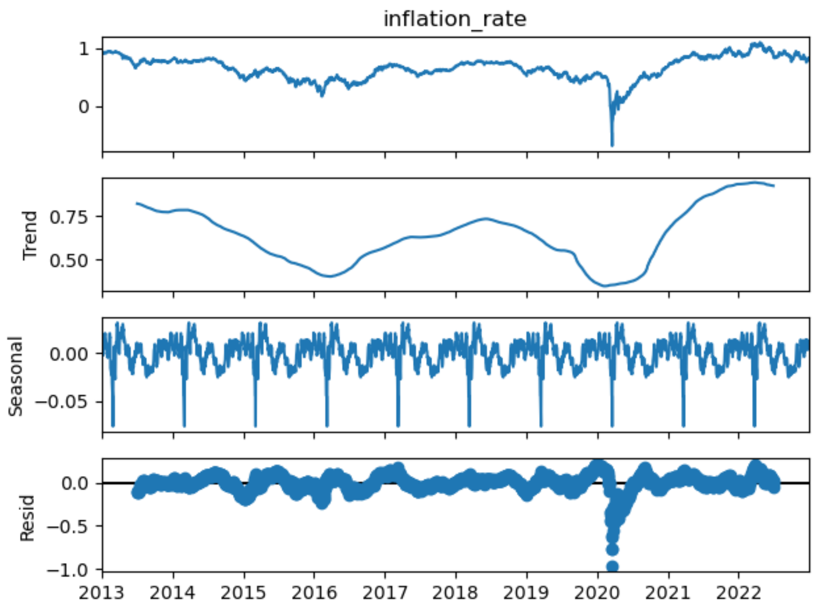
Figure 1.14: FedEx stock prices series decomposition




```
------------------------
Are the Residuals a WN ?
------------------------
Residuals is NOT a WN
```

Inflation rates stock prices series decomposition

Figure 1.15: Inflation rates stock prices series decomposition




```
------------------------
Are the Residuals a WN ?
------------------------
Residuals is NOT a WN
```

1.2.7 Taking the first difference

We take the first difference for each time series and check that there is no more stochastic trend.

In order to take a first difference we apply the following formula:

$\text{diff}[t] = x[t] - x[t-1]$

where:

- diff[t] is the difference at time t
- $x[t]$ is the value of the time series at time t
- $x[t-1]$ is the value of the time series at time $t-1$

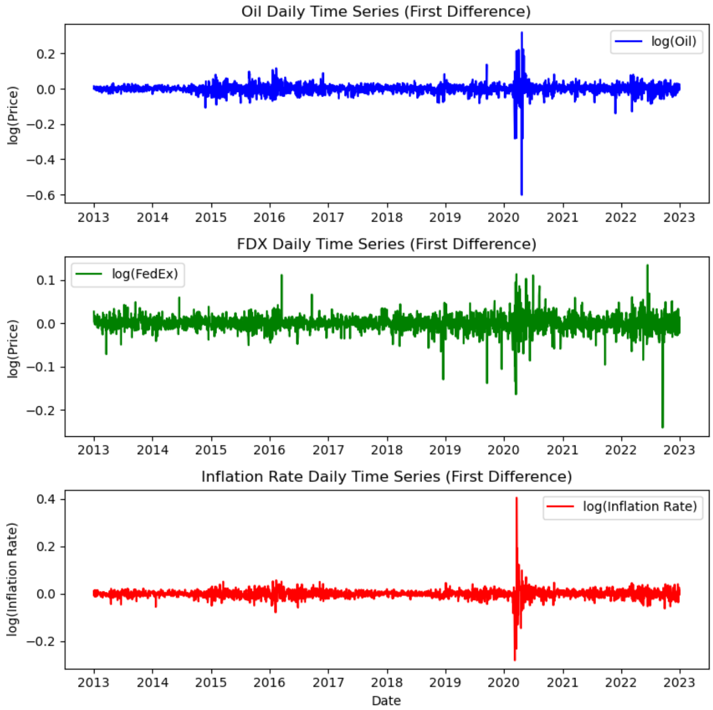
Figure 1.16: First difference for each time series

	CL=F	FDX	inflation_rate
2013-01-02	0.014059	0.027210	0.012171
2013-01-03	-0.002150	0.003707	-0.008097
2013-01-04	0.001828	0.003271	0.008097
2013-01-07	0.001074	-0.002004	0.016000
2013-01-08	-0.000429	0.004109	-0.011976
...	...	...	...
2022-12-23	0.026363	0.001365	0.004556
2022-12-27	-0.000377	0.006346	0.026907
2022-12-28	-0.007193	-0.022852	0.008811
2022-12-29	-0.007117	0.014343	-0.004396
2022-12-30	0.023447	-0.013477	0.013129

2499 rows × 3 columns

Plot the first difference of each series

Figure 1.17: Plotting after first difference for each time series



We can see visualy that the series seems to be now stationary after taking the first difference. We will perform a ADF test to check if there are no more stochastic trend.

Perform ADF test on each series after taking first difference

Figure 1.18: ADF test on Oil stock prices (CL=F) plot (After first difference)

```
############################################
ADF TEST (First difference): OIL SROCK PRICES  (log(CL=F))
############################################

______________________________
REGRESSION : CONSTANT AND TREND
______________________________
ADF Statistic: -9.104015740103508
p-value for the trend: 2.1147474425019912e-13

Trend coefficient is sginificatif 2.1147474425019912e-13 < 0.05
We reject H0 (The trend coefficient is not signiciant (b1=0) and accept H1 (b0≠0)
ADF test with Constant and Trend: adf_statistic=-9.104015740103508 <-3.4122646594651993 Reject the null hypothesis H0, we accep
t H1 at 95% confidence level.
The time series is likely stationary but a determistic trend
```

Figure 1.19: Oil stock prices series (CL=F) plot (After first difference)

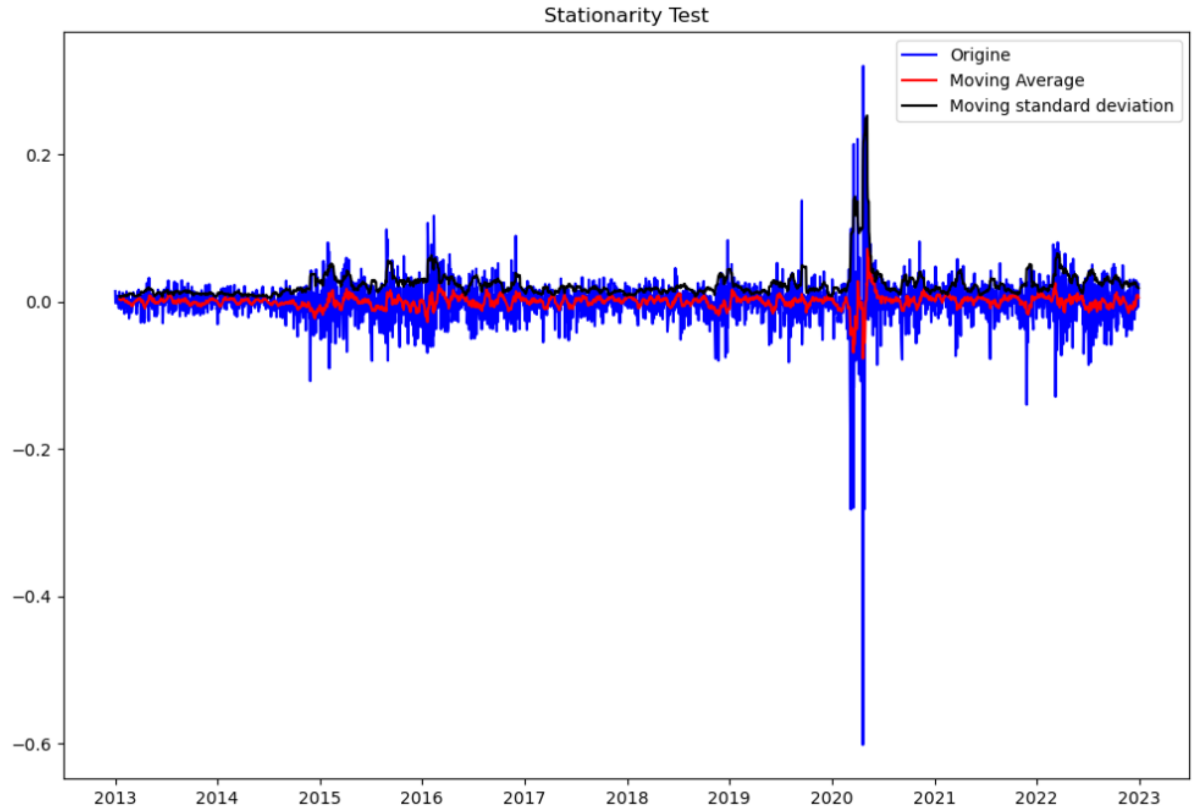


Figure 1.20: ADF test on FedEx stock prices (FDX) (After first difference)

```
###########################################
ADF TEST (First difference): FedEx STOCK PRICES (log(FXD))
###########################################

______________________________
REGRESSION : CONSTANT AND TREND
______________________________
ADF Statistic: -15.871765952802278
p-value for the trend: 1.1638316694938227e-22

Trend coefficient is sginificatif 1.1638316694938227e-22 < 0.05
We reject H0 (The trend coefficient is not signiciant (b1=0) and accept H1 (b0≠0)
ADF test with Constant and Trend: adf_statistic=-15.871765952802278 <-3.4122546731863177 Reject the null hypothesis H0, we acce
pt H1 at 95% confidence level.
The time series is likely stationary but a determistic trend
```

Figure 1.21: FedEx stock prices (FDX) plot (After first difference)

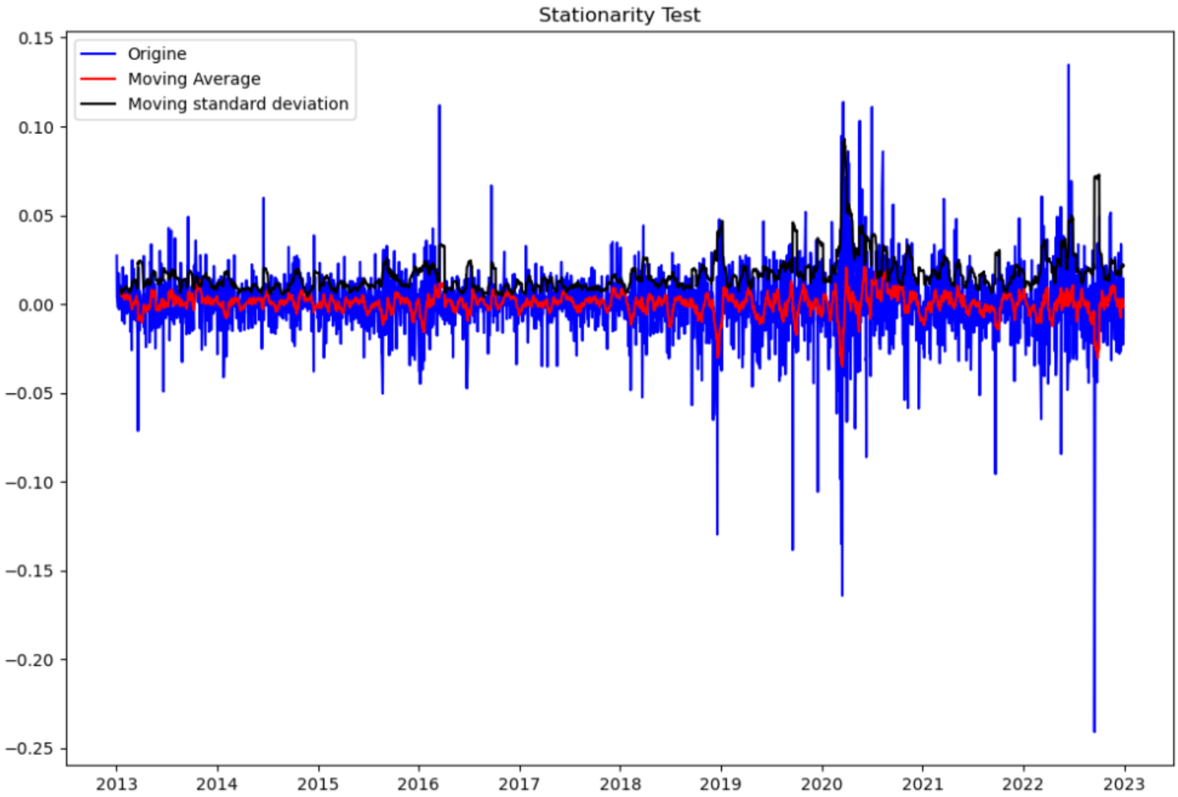


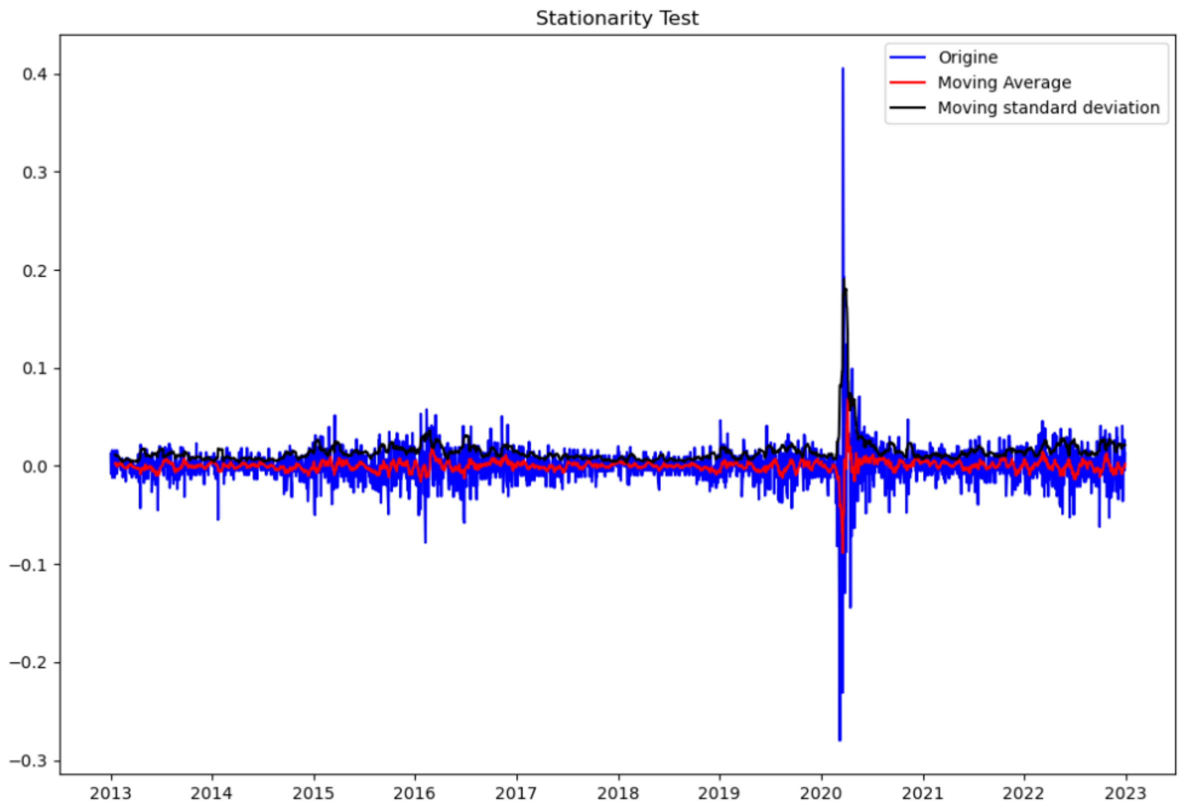
Figure 1.22: ADF test on the Inflation rates plot (After first difference)

```
###########################################
ADF TEST (First difference): INFLATION RATE (log(inflation_rate))
###########################################

______________________________
REGRESSION : CONSTANT AND TREND
______________________________
ADF Statistic: -13.432391872954224
p-value for the trend: 4.1156929556948056e-21

Trend coefficient is sginificatif 4.1156929556948056e-21 < 0.05
We reject H0 (The trend coefficient is not signiciant (b1=0) and accept H1 (b0≠0)
ADF test with Constant and Trend: adf_statistic=-13.432391872954224 <-3.4122617947281424 Reject the null hypothesis H0, we acce
pt H1 at 95% confidence level.
The time series is likely stationary but a determistic trend
```

Figure 1.23: Inflation rates series plot (After first difference)



According the ADF test, we can see that after applying first difference, the 3 time series are stationary but a determistic trend. Therefore, we affirm that for the 3 series are likely stationary. There is no presence of a unit root and stochastic trend.

In addition, the visual plots of the series after taking the first difference, we can see that the plot of the mean and variance for each series, seem to confirm that the series are stationary, since the variance seems almost constant over time and the mean is close to zero over time for each of the 3 time series.

1.2.8 Using ACF and PACF to choose model order

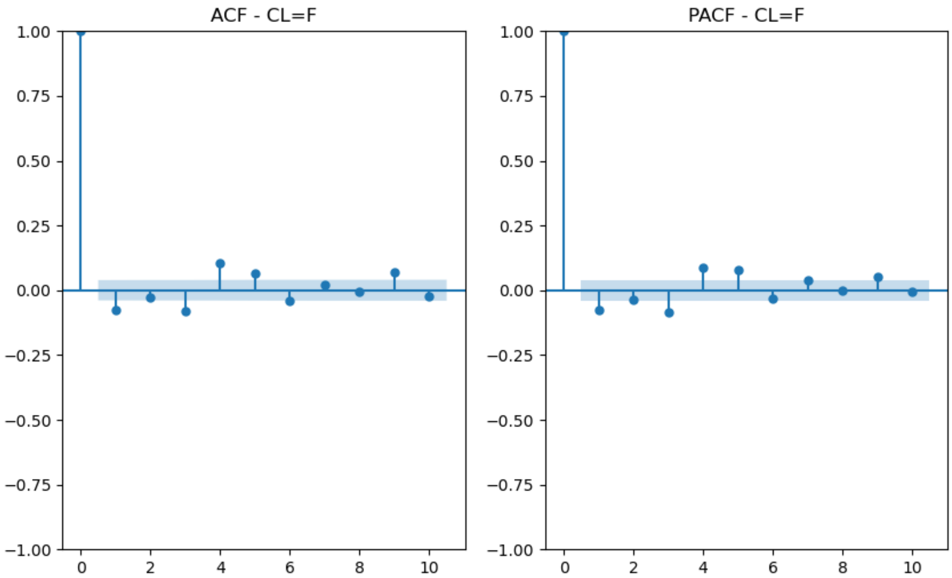
By comparing the ACF and PACF for a time series, we can deduce the order of the ARIMA model. If the amplitude of the ACF decreases with increasing lag and the PACF cuts off after a certain lag p, then we have an $AR(p)$ model.

If the ACF amplitude cuts off after a certain lag q and the PACF amplitude decreases, then we have an $MA(q)$ model.

In summary:

AP(p)	MA(q)	ARMA(p,q)
Tails off	Cuts off after lag q	Tails off
Cuts off after lag p	Tails off	Tails off

Figure 1.24: ACF and PACF for Oil stock prices (After first difference)

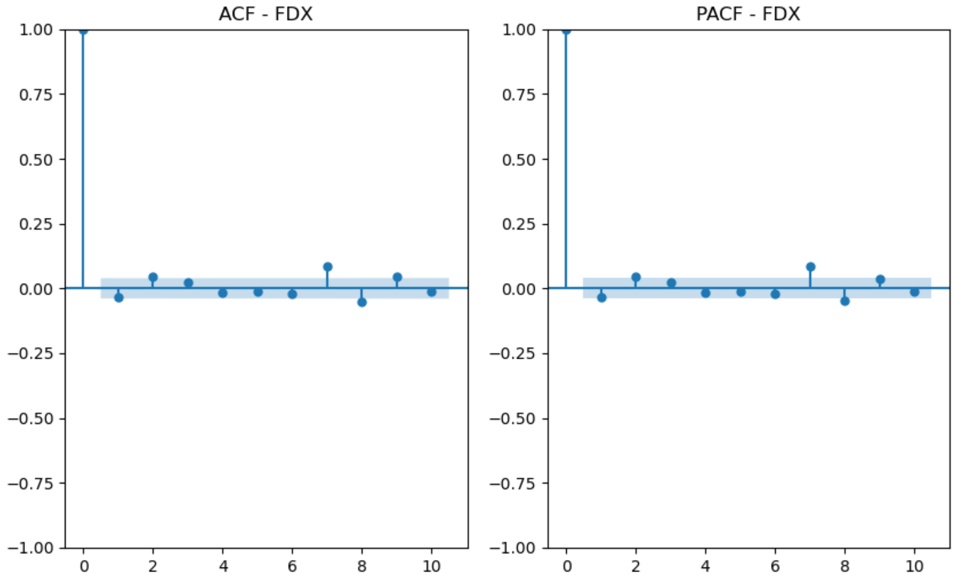


According to the PACF and ACF plots, we can see that for Oil stock prices (CL=F):

- ACF cuts off after lag 1
- PACF cuts off after lag 1

Hence, the best order seems to be $P = 1$ and $Q = 1$

Figure 1.25: ACF and PACF for FedEx stock prices (After first difference)

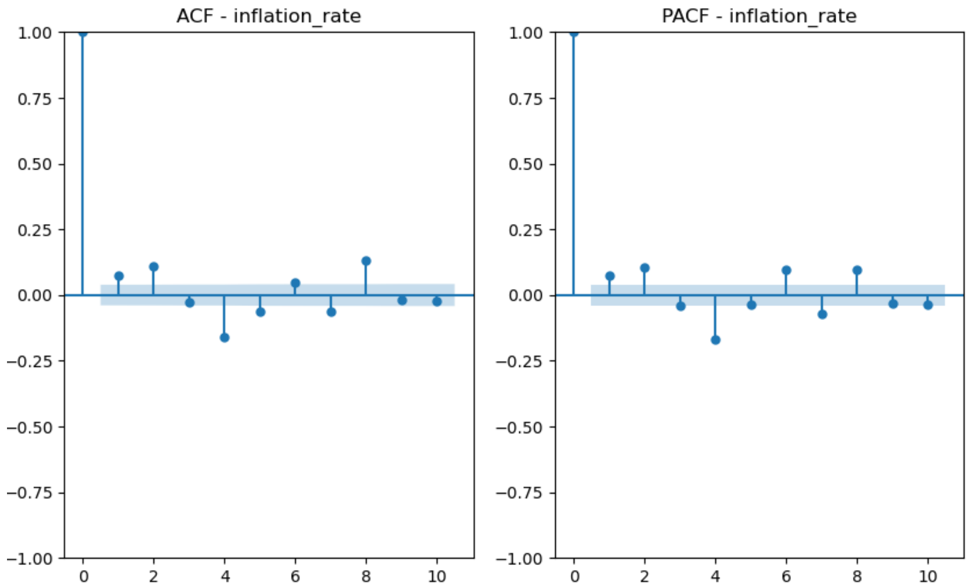


According to the PACF and ACF plots, we can see that for FedEx stock prices (FDX):

- ACF cuts off after lag 2
- PACF cuts off after lag 2

Hence, the best order seems to be $P = 2$ and $Q = 2$

Figure 1.26: ACF and PACF for Inflation rates (After first difference)



According to the PACF and ACF plots, we can see that for the Inflation rates series:

- ACF cuts off after lag 2
- PACF cuts off after lag 2

Hence, the best order seems to be $P = 2$ and $Q = 2$

We will verify the results we obtained from ACF and PACF plot by searching the best orders for ARIMA model over AIC and BIC criterions

Searching best parameters for ARIMA model over AIC and BIC criterions

To search the best order we iterate over the p and q parameter and fit the model for each time series. Then, we evaluate the performance of each ARIMA with several information criterions such as BIC, AIC, MSE and RMSE. Lower the information criterion are, better the performance of ARIMA model are.

- p**:** number of auto-regressive terms (AR order)
- d**:** number of non-seasonal differences (differentiation order)
- q**:** number of moving average terms (MA order)

Given that all our series are differentiated two times, we set the order d at 2 because we took the second difference of our time series

ARIMA for Oil stock prices

Figure 1.27: ARMA for Oil stock prices (After first difference)

	p	q	AIC	BIC	MSE	RMSE
8	2	2	-10268.548119	-10239.431891	0.000954	0.030882
2	0	2	-10267.506151	-10250.036414	0.000956	0.030914
7	2	1	-10267.263245	-10243.970263	0.000955	0.030903
4	1	1	-10266.253639	-10248.783902	0.000956	0.030922
5	1	2	-10255.177816	-10231.884834	0.000959	0.030975
1	0	1	-10253.086260	-10241.439768	0.000962	0.031017
6	2	0	-9391.364660	-9373.894923	0.001360	0.036877
3	1	0	-9138.933885	-9127.287393	0.001506	0.038805
0	0	0	-8343.796716	-8337.973470	0.002072	0.045519

According to AIC, BIC, MSE and MSE criterions, The best order for a ARMA model are $P = 2$ and $Q = 2$

ARMA for FedEx stock prices

Figure 1.28: ARMA for FedEx stock prices (After first difference)

	p	q	AIC	BIC	MSE	RMSE
7	2	1	-12583.977992	-12560.685009	0.000378	0.019452
4	1	1	-12580.654751	-12563.185014	0.000379	0.019473
2	0	2	-12580.437334	-12562.967597	0.000379	0.019474
1	0	1	-12580.108127	-12568.461635	0.000380	0.019483
8	2	2	-12578.686701	-12549.570473	0.000379	0.019464
5	1	2	-12576.973223	-12553.680240	0.000379	0.019479
6	2	0	-11930.968227	-11913.498490	0.000492	0.022189
3	1	0	-11632.845572	-11621.199081	0.000555	0.023562
0	0	0	-10781.086915	-10775.263669	0.000781	0.027951

According to AIC and MSE criterions, the best order for a ARMA model are $P = 2$ and $Q = 1$.

Let's fit a ARMA(2,1,1) for the FedEx stock prices serie

ARMA for Inflation rates stock prices

Figure 1.29: ARMA for Inflation rates stock prices (After first difference)

	p	q	AIC	BIC	MSE	RMSE
7	2	1	-12394.890124	-12371.597142	0.000408	0.020194
8	2	2	-12381.557887	-12352.441658	0.000410	0.020240
4	1	1	-12368.862072	-12351.392335	0.000412	0.020306
2	0	2	-12366.430772	-12348.961035	0.000413	0.020316
5	1	2	-12358.037302	-12334.744319	0.000414	0.020342
1	0	1	-12357.341482	-12345.694990	0.000415	0.020361
6	2	0	-11759.106769	-11741.637032	0.000527	0.022959
3	1	0	-11610.914360	-11599.267869	0.000560	0.023660
0	0	0	-10826.035439	-10820.212194	0.000767	0.027697

According to BIC and AIC criterions, The best order for a ARMA model are $P = 2$ and $Q = 1$.

Let's fit a ARMA(2,1,1) for the Inflation Rate serie

1.2.9 Check the residuals of the ARMA

Finally we check that the residuals of the ARMA you retain are not serially correlated. To do so we will perform Liung-Box-test and plot differents graphics such as ACF plot, Normal Q-Q plot, Histogram + estimated density and correlogram.

Residuals plot

This plot shows the standardized residuals. If our model works correctly, there should be no obvious structure in the residuals.

Histogram + estimated density

The histogram shows the distribution of residuals by examine if the residual follow a normal distribution or not.

Normal Q-Q

The normal Q-Q diagram is another way of showing how the distribution of model residuals compares to a normal distribution. If our residuals are normally distributed, all points should lie along the line.

Correlogram

The correlogram is simply an ACF graph of the residuals rather than the data. 95% of correlations for a shift greater than zero should not be significant. If there is a significant correlation in the residuals, it means that there is information in the data that our model has not captured.

In summary:

- Standardized residual: There are no obvious pattern in the residuals
- Histogram + KDE : The KDE curve should be very similar to the normal distribution
- Normal Q-Q: Most of the data points should be lie on the straight line
- Correlogram: 95% of correlations for lag grater than zero should be significant

Ljung-Box test

The Ljung-Box test is checking whether the autocorrelations in the time series are jointly significantly different from zero up to the specified lag.

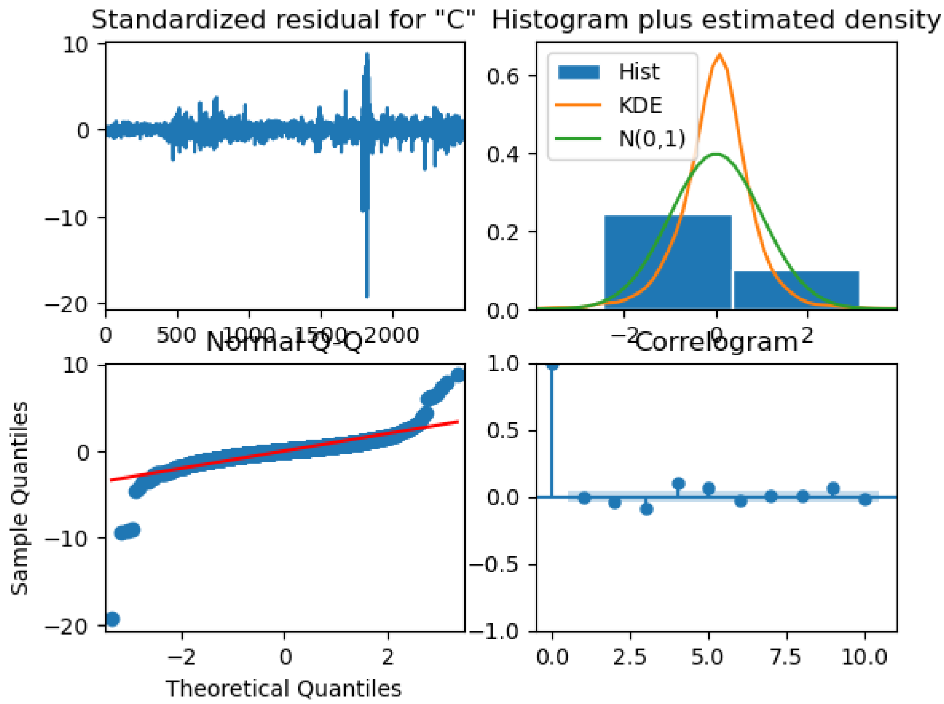
Ljung-Box test is used to assess the adequacy of a time series model, especially in the context of ARIMA modeling, where it is important to have residuals with no significant autocorrelation. If the test yields a low p-value, it suggests evidence against the null hypothesis of no serial correlation, indicating that the model's residuals may exhibit some significant autocorrelation.

- H_0: The residuals are independently distributed. There is no serial correlation in the time series up to a certain lag
- H_1: The residuals are not independently distributed; they exhibit serial correlation.

Check residuals of the $ARMA(2,1,2)$ for Oil stock prices

Figure 1.30: Residuals of the $ARMA(2,1,2)$ for Oil stock prices (After first difference)

```
Mean of Residuals: 0.00024320361812085295
We reject H0 and accept H1:The residuals are not independently distributed
Serial correlation detected in residuals.
```

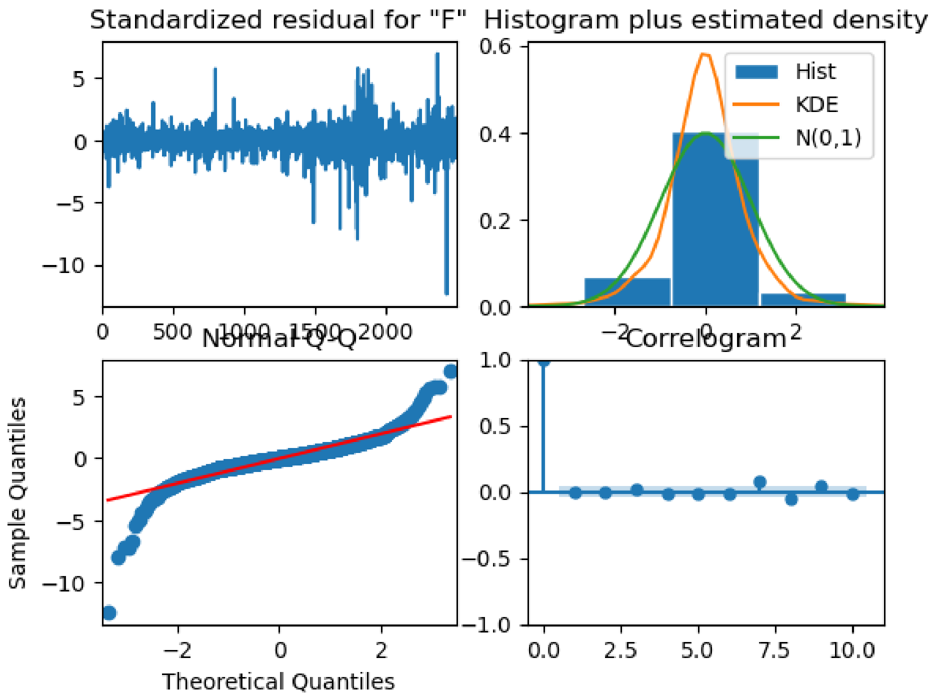


	lb_stat	lb_pvalue
1	0.021629	0.883078
2	3.502415	0.173564
3	19.882686	0.000180

Check residuals of the $ARMA(2, 1, 1)$ for FedEx stock prices

Figure 1.31: Residuals of the $ARMA(2, 1, 1)$ for FedEx stock prices (After first difference)

```
Mean of Residuals: -0.00048308230303752814
We fail to reject H0: The residuals are independently distributed.
No significant serial correlation in residuals.
```

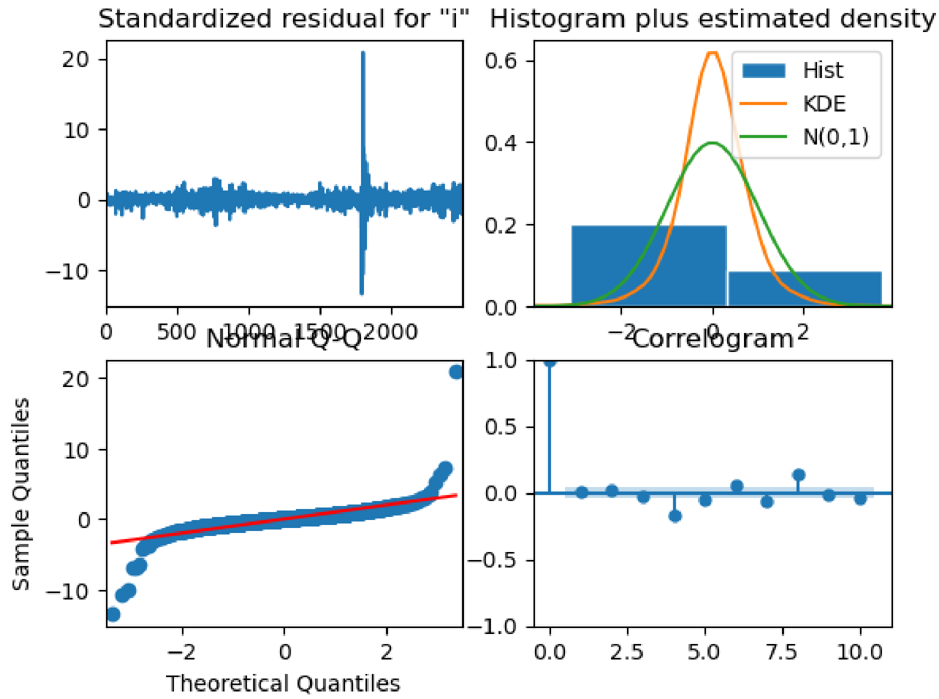


	lb_stat	lb_pvalue
1	0.010331	0.919042
2	0.011835	0.994100
3	1.406640	0.703979

Check residuals of the $ARMA(2,1,1)$ for Inflation rates series

Figure 1.32: Residuals of the $ARMA(2,1,1)$ for Inflation rates series (After first difference)

```
Mean of Residuals: 0.00016226042700602015
We fail to reject H0: The residuals are independently distributed.
No significant serial correlation in residuals.
```



	lb_stat	lb_pvalue
1	0.033711	0.854323
2	1.102634	0.576190
3	2.802910	0.423021

We can draw the same conclusion for our 3 series.

The Q-Q plot of each series looks like a heavy tailed q-q plot. This means that, compared with the normal distribution, there is much more data in the extremities than in the center of the distribution.

Furthermore, according to the density histogram, we can see that the green line showing a normal distribution is far from the KDE distribution (i.e. orange line).

This suggests that our ARIMA model can be improved.

Otherwise, according to the correlogram, there is no significant correlation in the residuals; our model seems to capture well the information in the data.

According the Ljung-Box test, for FedEx and Inflation Rate time series the residuals our ARIMA model there are no significant serial correlation in residuals.

Conversely, for the Oil stock price time series, there are serial correlation detected in residuals.

1.3 Conclusion

We analysed extensively 3 series. This work will serve as a basis for implementing more advanced models such as VAR, VECM, STR for taking into account of non linearity and Markov Switching model.

Please refer to the following link to see the full code of this assignment: Empirical Analysis 1

2 Empirical Analysis 2 & 3 : VAR in "level" and VECM

2.1 DATA

For this Empirical Application 2 and Empirical Application 3 we focus on the same time series than we use for the first Empirical Application.

For reminder, the three series we use are:

- 10-Year Breakeven Inflation Rate (T10YIE)
- Crude Oil Dec 23 (CL=F) stock index
- FedEx Corporation (FDX) stock index

Regarding the period we chose the following period: from 2012-12-31 to 2022-12-31.

We use FRED API to collect the inflation rate time series and Yahoo Finance API to collect the oil stock index and FedEx stock index.

Please refer to the first Empirical Analysis to see fully time series analysis (Time series decomposition, unit root analysis, ACF and PACF plots, trend analysis,...)

IMPORTANT NOTE:

Given that Empirical Application 2 and Empirical Application 3 are closely linked, we will here jointly work on the Empirical Application 2 and Empirical Application 3 at the same time.

2.2 Methodology

The methodology we implement is the following:

1- Cointegration Johansen rank test

The first step is to apply a cointegration test and examine whether the series are cointegrated. To do so, we apply a Cointegration Johansen rank test.

If there is no cointegration, we turn to a VAR model with first differences of the I(1) components:

- Order of the VAR

- Estimation
- Causality tests

If we validate cointegration, we will consider the VAR without differencing the I(1) components and apply a VAR in level and VECM model:

- **2 - For VAR in level**
  - Estimation
  - Causality tests
- **3 - VECM**
  - Apply a VECM model
  - Apply Causality test
  - Plot impulse response function

2.3 Cointegration Test

Cointegrated time series refer to a set of multiple time series variables that exhibit a long-term or equilibrium relationship with each other, even though individually, they may be non-stationary. In other words, cointegration is a statistical property that allows multiple non-stationary time series to move together in the long run, despite potentially having short-term fluctuations and trends. From a more mathematical point of view, cointegration implies that a linear combination of non-stationary time series variables results in a stationary time series. While individual variables may not be stationary, there exists a combination of them that is stationary.

Cointegration implies that there is a stable, long-term relationship among the variables. This means that any short-term deviations from this relationship will be corrected over time, and the variables will return to their equilibrium (i.e the mean).

In fact, cointegration is commonly used in economics and finance. For example, in finance, it's applied to pairs trading, where two or more assets that are cointegrated are traded against each other. In economics, cointegrated variables might represent different economic indicators like inflation and unemployment, which should theoretically move together in the long run.

To test if our 3 series are cointegrated we will apply the Johansen cointegration test. The Johansen cointegration test is a statistical method used to determine the presence and number of cointegrating relationships among a set of time series variables. Cointegration is a crucial concept in time series analysis, particularly in the context of multivariate data, where variables may exhibit long-term relationships even if they are not individually stationary.

The hypothesis of the Johansen cointegration test are :

- H_0**:** there are at most r cointegrating relationships in the system, where r is less than or equal to the number of variables K.
- H_1**:** there are more than r cointegrating relationships in the system, suggesting a higher degree of long-term dependencies among the variables.

The rank r is a measure of how many linearly independent cointegrating relationships exist among the variables. If the computed test statistics exceed the critical values, the null hypothesis H_0 is rejected in favor of the alternative hypothesis, indicating the presence of cointegration and providing insights into the number of cointegrating relationships in the system. In fact, rejecting the null hypothesis suggests that there is evidence of cointegration, implying that there are linear combinations of the variables that result in stationary relationships.

Johansen Cointegration Test results

Figure 2.1: Johansen Cointegration Test results

```
==== Results Johansen cointegration test ====
Trace Statistic at [90%, 95%, 99%] Confidence Level: [59.63450555 14.23244478  5.48212509]
Max Eigenvalue Statistic at [90%, 95%, 99%] Confidence Level: [45.40206077  8.75031969  5.48212509]
Critical Values for Trace Statistic: [18.8928 12.2971  2.7055]
Critical Values for Max Eigenvalue Statistic: [21.1314 14.2639  3.8415]
Rank (Number of Cointegrating Relationships): 3


==== Hypoyhesis Testing ====
For significance level: 0.05
Reject the null hypothesis: There is evidence of cointegration.
```

Figure 2.2: Johansen Cointegration Test results

```
#### COINTEGRATION TEST SUMMARY ####
Johansen cointegration test using trace test statistic with 95% significance level
```

Johansen cointegration test using trace
test statistic with 5% significance level

r_0	r_1	test statistic	critical value
0	3	59.63	29.80
1	3	14.23	15.49

Trace Statistic

The trace statistic is a test statistic associated with the null hypothesis that there are at most r cointegrating relationships in the system. The larger the trace statistic, the more evidence there is against the null hypothesis. This statistic is compared to critical values to make decisions regarding the presence of cointegration.

Max Eigenvalue Statistic

The maximum eigenvalue statistic is another test statistic associated with the null hypothesis. It specifically tests whether there are exactly r cointegrating relationships. Like the trace statistic, a larger value indicates stronger evidence against the null hypothesis.

Critical Values for Trace Statistic

Critical values for Trace Statistic are thresholds beyond which we would reject the null hypothesis. The trace statistic is compared to a set of critical values to determine whether to reject the null hypothesis of at most r cointegrating relationships.

Critical Values for Max Eigenvalue Statistic

Similar to the trace statistic, Critical Values for Max Eigenvalue Statistic are the critical values associated with the max eigenvalue statistic. They serve as thresholds for rejecting the null hypothesis of exactly r cointegrating relationships.

Rank (Number of Cointegrating Relationships)

The rank is the number of cointegrating relationships suggested by the test. It is determined based on the trace statistic and critical values. The rank gives insights into the long-term relationships among the variables.

Comment about the Cointegration Test results

In summary, we can see, that trace statistics and max eigen value statistic are higher compared to their respective critical values at 90%, 95% and 99% confidence level.

For example, in our case, both the trace statistic (59.6345) and the max eigenvalue statistic (45.4020) exceed the critical values at the 95% confidence level (18.8928 for trace and 21.1314 for max eigenvalue).

This results suggests evidence against the null hypothesis H_0, indicating the presence of cointegration. Hence, we reject the null hypothesis H_0 and accept H_1 hypothesis.

We conclude that, we have evidence to support the presence of cointegration among the variables. This implies that there are linear combinations of these variables that result in stationary relationships, providing a long-term relation among them. This information is valuable for understanding the underlying structure and dynamics of our multivariate time series data.

Interpretation of Eigenvalue Trace Test

The test compares the test statistic (59.63) to the critical value (29.80). In this case, the test statistic (59.63) is greater than the critical value (29.80), suggesting that we reject the null hypothesis at the 5% significance level. This implies that there are at least r_0 cointegrating relationships among the variables.

Interpretation of Maximum Eigenvalue Test

The test compares the test statistic (14.23) to the critical value (15.49). In this case, the test statistic (14.23) is less than the critical value (15.49), suggesting that we do not reject the null hypothesis at the 5% significance level. This implies that there are at least r_1 cointegrating relationships among the variables. Otherwise as the test statistic (14.23) is close to the critical value (15.49) we can affirm that the series are marginally cointegrated.

Overall, the results suggest that there are cointegrating relationships among the variables, but the exact number is not clear. It's common to consider the lower rank r_1 as the more conservative estimate, so we might interpret this as suggesting there are at most r_1 cointegrating relationships.

Comment of the sign of cointegration coefficient

Figure 2.3: Johansen Cointegration coefficient

```
==== Coefficient Cointegration Relationships ====
Cointegrating Vectors:
[[  6.38836142   0.47222007   2.55381448]
 [  2.94745808  -3.27971911   0.86930487]
 [-12.51877619   0.26478404   0.4060579 ]]

Cointegrating Relationship 1:
Coefficient for CL=F: 6.3884
Coefficient for FDX: 2.9475
Coefficient for inflation_rate: -12.5188

Cointegrating Relationship 2:
Coefficient for CL=F: 0.4722
Coefficient for FDX: -3.2797
Coefficient for inflation_rate: 0.2648

Cointegrating Relationship 3:
Coefficient for CL=F: 2.5538
Coefficient for FDX: 0.8693
Coefficient for inflation_rate: 0.4061
```

The signs of the coefficients indicate the direction and strength of the relationships in each cointegrating combination. Interpretation may depend on the context of the study case and the specific economic or financial factors represented by the variables. Overall, positive coefficients imply positive relationships, while negative coefficients imply negative relationships. More precisely, the coefficients in the cointegrating vectors represent the weights of each variable in these relationships.

To comment the sign of the coefficients we will discuss and go through the 3 cointegrated relationship.

Cointegrating Relationship 1

- **CL=F (Crude Oil Futures):** The positive coefficient (6.3884) suggests a positive relationship with the cointegrated combination. An increase in CL=F is associated with an increase in this cointegrating relationship.
- **FDX (FedEx Stock):** The positive coefficient (2.9475) indicates a positive relationship. An increase in FDX is associated with an increase in this cointegrating relationship.
- **Inflation Rate** The negative coefficient (-12.5188) suggests a negative relationship. An increase in the inflation rate is associated with a decrease in this cointegrating relationship.

In other words, as Crude Oil and FedEx increase, the Inflation Rate tends to decrease in the long term, and vice versa. This cointegrating relationship seems to capture a combination of economic factors where the price of crude oil futures and FedEx stock move in the same direction, while the inflation rate moves in the opposite direction. This could be explained by the fact that higher oil prices might lead to increased transportation costs (affecting FedEx companies) and could also have an impact on inflation. The negative coefficient for Inflation Rate suggests an inverse relationship, indicating that as inflation increases, Crude Oil and FedEx are expected to decrease in the long term.

Cointegrating Relationship 2

- **CL=F (Crude Oil Futures):** The positive coefficient (0.4722) suggests a positive relationship with the second cointegrating combination.
- **FDX (FedEx Stock):** The negative coefficient (-3.2797) indicates a negative relationship.
- **Inflation Rate:** The positive coefficient (0.2648) suggests a positive relationship.

This second relationship implies a combination where the price of crude oil futures and the inflation rate move in the same direction, while the movement of FedEx stock is in the opposite direction. This relationship implies that as Crude Oil and Inflation increase, FedEx tends to decrease in the long term, and vice versa. In this relationship, the strong negative coefficient for FedEx suggests that FedEx stock is inversely related to the cointegrating factor. This could be interpreted as FedEx being sensitive to economic conditions or market sentiment. The positive coefficient for Inflation Rate indicates that as inflation increases, the cointegrating factor tends to increase, and the effect on Crude Oil is relatively weaker.

Cointegrating Relationship 3

- **CL=F (Crude Oil Futures):** The positive coefficient (2.5538) suggests a positive relationship with the third cointegrating combination.
- **FDX (FedEx Stock):** The positive coefficient (0.8693) indicates a positive relationship.
- **Inflation Rate:** The positive coefficient (0.4061) suggests a positive relationship.

In this relationship, all three variables—crude oil futures, FedEx stock, and the inflation rate—move in the same direction. This third relationship suggests that as Crude Oil, FedEx, and Inflation increase, they tend to move together in the long term. In this relationship, all coefficients are positive. Crude Oil and FedEx, in particular, show a positive connection with the cointegrating factor. This suggests that as Crude Oil and FedEx increase, the cointegrating factor (potentially representing broader economic or market trends) also tends to increase. The positive coefficient for Inflation Rate further supports the idea that these variables move together in the long term.

Overall Implications

- These cointegrating relationships indicate that these three variables are linked in the long term, and their movements are not independent.
- The relationships highlight potential economic dynamics, such as the impact of oil prices on transportation costs (affecting FedEx) and the interplay with inflation.
- Investors and policymakers may find these relationships useful for understanding how changes in Crude Oil, FedEx stock, and Inflation Rate may influence each other over time.

It's important to note that cointegration analysis provides information about long-term relationships but does not imply causation.

Overall conclusion, our series a cointegrated. Therefore we can turn to VAR in level and then VECM model.

2.4 Applying VAR model in level

The Vector Autoregressive (VAR) model is a statistical time series model used in econometrics to analyze the dynamic relationships among multiple variables over time. It is an extension of the univariate autoregressive (AR) model to multivariate time series data. VAR models are particularly useful when the variables under consideration are interrelated and influence each other.

Vector Notation

y_t denotes a vector on K variables at time t. The VAR model deals with p-th order autoregressive processes, denotes as $VAR(p)$ where p is the number of lags.

Formulation:

The general formula of a $VAR(p)$ model is given by:

$\mathbf{y}_t = \mathbf{c} + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{y}_{t-i} + \varepsilon_t$

where:

- y_t is the vector of time series variables at time t
- c is a vector of constants
- A_i represents the coefficient matrix for each lag i from 1 to p
- ϵ_t is a vector of error terms assumed to be white noise

2.4.1 Estimate the order of the VAR model without differencing the series

Determining the appropriate order p of the VAR model is crucial. Various criteria, such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), are commonly used to select the optimal lag order.

In our case to estimate the best order we will loop over p orders and choose the best order according the AIC and BIC criterions.

Figure 2.4: TOP 5 VAR in level best orders

TOP 5: Best orders

	p	AIC	BIC
5	5	-22.869593	-22.757585
2	2	-22.801647	-22.752693
6	6	-22.884176	-22.751124
8	8	-22.925335	-22.750149
3	3	-22.817049	-22.747091

According to the AIC and BIC information criterion, we can see that the best order for the VAR model is lag $p = 5$.

2.4.2 Estimate VAR model in Level

Results Summary

Figure 2.5: VAR model summary

```
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
Date:           Sat, 09, Dec, 2023
Time:                     13:54:31
--------------------------------------------------------------------
No. of Equations:         3.00000    BIC:                  -22.7576
Nobs:                     2495.00    HQIC:                 -22.8289
Log likelihood:           17957.1    FPE:               1.16913e-10
AIC:                     -22.8696    Det(Omega_mle):    1.14692e-10
```

Results for equation CL=F

Figure 2.6: Results for equation CL=F

Results for equation CL=F

	coefficient	std. error	t-stat	prob
const	0.093446	0.024243	3.855	0.000
L1.CL=F	0.885327	0.020752	42.662	0.000
L1.FDX	-0.069098	0.032198	-2.146	0.032
L1.inflation_rate	0.226277	0.032311	7.003	0.000
L2.CL=F	0.022627	0.027190	0.832	0.405
L2.FDX	0.151501	0.044573	3.399	0.001
L2.inflation_rate	-0.170936	0.045155	-3.785	0.000
L3.CL=F	-0.000488	0.027305	-0.018	0.986
L3.FDX	-0.111656	0.044670	-2.500	0.012
L3.inflation_rate	-0.030164	0.045178	-0.668	0.504
L4.CL=F	0.168442	0.027165	6.201	0.000
L4.FDX	0.083866	0.044684	1.877	0.061
L4.inflation_rate	-0.098346	0.045121	-2.180	0.029
L5.CL=F	-0.095888	0.020686	-4.635	0.000
L5.FDX	-0.061114	0.032470	-1.882	0.060
L5.inflation_rate	0.106701	0.031931	3.342	0.001

The constant term (0.0934) represents the estimated value of Crude Oil (CL=F) when all lagged values are zero.

The coefficient for the lagged value of CL=F at time $t-1$ (L1.CL=F: 0.8853) is positive and highly significant. This suggests a strong positive relationship between the current value of CL=F and its past value. This is consistent with the idea that Crude Oil prices are persistent and tend to follow trends.

The coefficient for the lagged value of FDX at time $t-1$ (L1.FDX: -0.0691) is negative and statistically significant. This implies a negative relationship between Crude Oil and FedEx in the short term. This could be due to the fact that higher oil prices might increase transportation costs for companies like FedEx.

The coefficient for the lagged value of inflation rate at time t-1 (L1.inflation rate: 0.2263) is positive and statistically significant. This suggests a positive relationship between Crude Oil and Inflation Rate in the short term. Higher oil prices may contribute to higher inflation.

Results for equation FDX

Figure 2.7: Results for equation FDX

Results for equation FDX

	coefficient	std. error	t-stat	prob
const	0.007727	0.015567	0.496	0.620
L1.CL=F	-0.008211	0.013325	-0.616	0.538
L1.FDX	0.964236	0.020675	46.638	0.000
L1.inflation_rate	0.012799	0.020748	0.617	0.537
L2.CL=F	-0.001156	0.017459	-0.066	0.947
L2.FDX	0.066417	0.028621	2.321	0.020
L2.inflation_rate	0.051203	0.028995	1.766	0.077
L3.CL=F	0.030541	0.017533	1.742	0.082
L3.FDX	-0.020646	0.028683	-0.720	0.472
L3.inflation_rate	-0.066447	0.029010	-2.291	0.022
L4.CL=F	-0.031427	0.017443	-1.802	0.072
L4.FDX	-0.028641	0.028692	-0.998	0.318
L4.inflation_rate	-0.016225	0.028973	-0.560	0.575
L5.CL=F	0.011888	0.013283	0.895	0.371
L5.FDX	0.016573	0.020850	0.795	0.427
L5.inflation_rate	0.013092	0.020503	0.639	0.523

The constant term (0.0077) represents the estimated value of FedEx (FDX) when all lagged values are zero.

The coefficient for the lagged value of CL=F at time t-1 (L1.CL=F: -0.0082) is negative but not statistically significant. There is weak evidence of a short-term relationship between the previous value of Crude Oil and current FedEx. The non-significant coefficient for the lagged value of CL=F implies that the previous value of Crude Oil does not have a statistically significant impact on the current value of FedEx in the short term.

The coefficient for the lagged value of FDX at time t-1 (L1.FDX: 0.9642) is positive and highly significant. This indicates a strong positive relationship between the current value of FedEx and its past value. This suggests that FedEx stock prices are persistent and tend to follow trends.

The coefficient for the lagged value of inflation rate at time t-1 (L1.inflation rate: 0.0128) is positive but not statistically significant. There is weak evidence of a short-term relationship between FedEx and Inflation Rate. The non-significant coefficient for the lagged value of inflation rate indicates that the previous value of Inflation Rate does not have a statistically significant impact on the current value of FedEx in the short term.

Results for equation Inflation Rate

Figure 2.8: Results for equation Inflation Rate

Results for equation inflation_rate

	coefficient	std. error	t-stat	prob
const	-0.043429	0.015654	-2.774	0.006
L1.CL=F	0.085092	0.013400	6.350	0.000
L1.FDX	0.070452	0.020791	3.389	0.001
L1.inflation_rate	1.007692	0.020864	48.298	0.000
L2.CL=F	-0.088568	0.017557	-5.045	0.000
L2.FDX	0.002322	0.028781	0.081	0.936
L2.inflation_rate	0.077943	0.029158	2.673	0.008
L3.CL=F	0.021742	0.017631	1.233	0.218
L3.FDX	0.036323	0.028844	1.259	0.208
L3.inflation_rate	-0.147964	0.029173	-5.072	0.000
L4.CL=F	-0.010440	0.017541	-0.595	0.552
L4.FDX	-0.134493	0.028853	-4.661	0.000
L4.inflation_rate	-0.116205	0.029135	-3.988	0.000
L5.CL=F	-0.001182	0.013357	-0.088	0.929
L5.FDX	0.030688	0.020967	1.464	0.143
L5.inflation_rate	0.161712	0.020619	7.843	0.000

The constant term (-0.0434) represents the estimated value of Inflation Rate when all lagged values are zero.

The coefficient for the lagged value of CL=F at time $t-1$ (L1.CL=F: 0.0851) is positive and highly significant. This suggests a positive relationship between the previous value of Crude Oil and the current Inflation Rate. Indeed, higher Crude Oil prices may contribute to higher inflation in the short term.

The coefficient for the lagged value of FDX at time $t-1$ (L1.FDX: 0.0705) is positive and statistically significant. This indicates a positive relationship between the previous value of FedEx and the current Inflation Rate. The positive and statistically significant coefficient for the lagged value of FDX implies a positive relationship between the past and current values of Inflation Rate could be due to the general economic activity reflected in the FedEx stock prices.

The coefficient for the lagged value of inflation rate at time $t-1$ (L1.inflation rate: 1.0077) is positive and highly significant. This suggests a strong positive relationship between the current Inflation Rate and its past value. This is consistent with the idea of inflation being persistent over time.

Overall interpretation

- Crude Oil prices play a significant role in short-term dynamics, impacting both FedEx and the Inflation Rate.
- FedEx stock prices are influenced by their own past values, but the impact of Crude Oil and Inflation Rate on FedEx is not statistically significant in the short term.
- Inflation Rate is influenced by its own past values, as well as the past values of Crude Oil and FedEx. This suggests a dynamic relationship where economic activity and transportation costs may contribute to changes in inflation.

Correlation matrix of residuals

The correlation matrix of residuals provides insights into the relationships between the residuals of different variables in your Vector Autoregression (VAR) model in level.

Figure 2.9: Correlation matrix of residuals

Correlation matrix of residuals

	CL=F	FDX	inflation_rate
CL=F	1.000000	0.146789	0.257619
FDX	0.146789	1.000000	0.220381
inflation_rate	0.257619	0.220381	1.000000

- **CL=F and FDX:** The positive correlation coefficient (0.146789) suggests a weak positive linear association between the residuals of Crude Oil (CL=F) and FedEx (FDX) stock indices. In econometric terms, this implies that when there is a positive deviation from the expected value in the CL=F series, there is a tendency for a positive deviation in the FDX series as well, and vice versa. However, the correlation is relatively low, indicating that the two variables' residuals are not strongly correlated. These findings suggests that a positive correlation between the residuals of Crude Oil and FedEx stock indices could be related to broader economic factors. For instance, both Crude Oil prices and FedEx stock prices might be influenced by economic growth or changes in global demand.

- **CL=F and inflation rate:** The positive correlation coefficient (0.257619) indicates a weak to moderate positive linear association between the residuals of Crude Oil (CL=F) and the Inflation Rate. This suggests that when there is a positive deviation from the expected value in the CL=F series, there is a tendency for a positive deviation in the inflation rate series, and vice versa. Indeed, the positive correlation between Crude Oil and Inflation Rate residuals is in line with economic theory. Changes in oil prices often impact overall price levels, contributing to inflation. This correlation might reflect the sensitivity of inflation to energy prices.
- **FDX and inflation rate:** The positive correlation coefficient (0.220381) implies a weak to moderate positive linear association between the residuals of FedEx (FDX) stock index and the Inflation Rate. This suggests that when there is a positive deviation from the expected value in the FDX series, there is a tendency for a positive deviation in the inflation rate series, and vice versa. In fact, the positive correlation between FedEx stock index and Inflation Rate residuals could be associated with economic conditions where increased economic activity, as reflected in higher stock prices, is accompanied by higher inflation.

2.5 Applying VECM model

Based on the previous Johansen cointegration test we know that our 3 series are cointegrated. Therefore, we can apply a VECM model.

A VECM (Vector Error Correction Model) is a specialized form of a Vector Autoregressive (VAR) model used for analyzing and modeling the behavior of multiple, cointegrated time series variables. VECM models are particularly useful when dealing with non-stationary time series data and are commonly used in economics and finance.

VECM models are designed to capture both short-term dynamics and the long-term equilibrium relationships among the variables in the system. In other words, a VECM combines short-term deviations from equilibrium (measured by error terms) with the long-term equilibrium relationship.

The "Error Correction Term" is a crucial component of a VECM model. It represents the mechanism by which the variables return to their long-term equilibrium after experiencing short-term shocks or deviations. This term accounts for the cointegration among the variables and ensures that the model reflects the concept of mean reversion, where variables revert to their long-term relationship after temporary divergences.

VECM Equation

$\Delta \mathbf{Y}_t = \mathbf{\Pi} \mathbf{Y}_{t-1} + \sum_{i=1}^{p-1} \mathbf{\Gamma}_i \Delta \mathbf{Y}_{t-i} + \mathbf{A} \mathbf{D}_t + \mathbf{B} \mathbf{X}_{t-1} + \mathbf{u}_t$

where:

- $\Delta \mathbf{Y}_t$ is the matrix of first differences of the variables at time t
- $\mathbf{Y}_{t-1}$ is the lagged levels of the variables at time
- $\mathbf{\Pi}$ is the matrix of cointegrating vectors.
- $\mathbf{\Gamma}_i$ are the matrices of adjustment coefficients for the short-run dynamics.
- A is a matrix that captures the impact of exogenous variables D_t, on the cointegrating relationship.
- B captures the impact of lagged values of exogenous variables X_{t-1} on the cointegrating relationship.
- u_t is the error term

The cointegrating vectors Π represent the long-term equilibrium relationships among the variables. These vectors are obtained from the Johansen cointegration test and indicate how the variables adjust to deviations from their long-term relationships.

The adjustment coefficients $\mathbf{\Gamma}_i$ represent the short-term dynamics that correct deviations from the long-term equilibrium. The matrices A and B capture the impact of exogenous variables on the cointegrating relationship.

In summary, the VECM is a powerful model for capturing both short-term and long-term relationships among cointegrated variables. It is widely used in econometrics and finance for modeling and forecasting.

2.5.1 VECM coefficients

Figure 2.10: Results for equation Inflation Rate

Det. terms outside the coint. relation & lagged endog. parameters for equation CL=F

	coef	std err	z	P>|z|	[0.025	0.975]
L1.CL=F	-0.1035	0.020	-5.077	0.000	-0.143	-0.064
L1.FDX	-0.0659	0.032	-2.032	0.042	-0.129	-0.002
L1.inflation_rate	0.2080	0.032	6.573	0.000	0.146	0.270

Det. terms outside the coint. relation & lagged endog. parameters for equation FDX

	coef	std err	z	P>|z|	[0.025	0.975]
L1.CL=F	-0.0042	0.013	-0.323	0.746	-0.030	0.021
L1.FDX	-0.0368	0.021	-1.787	0.074	-0.077	0.004
L1.inflation_rate	0.0221	0.020	1.100	0.271	-0.017	0.061

Det. terms outside the coint. relation & lagged endog. parameters for equation inflation_rate

	coef	std err	z	P>|z|	[0.025	0.975]
L1.CL=F	0.0906	0.013	6.791	0.000	0.064	0.117
L1.FDX	0.0614	0.021	2.893	0.004	0.020	0.103
L1.inflation_rate	0.0315	0.021	1.520	0.129	-0.009	0.072

For Equation CL=F:

- **Lagged Endogenous Variables:**
  - $L1.CL = F$: The coefficient is -0.1035. This suggests that there is a negative relationship between the lagged value of Crude Oil (CL=F) and the current value of Crude Oil. In other words, if Crude Oil was higher in the previous period, it tends to decrease in the current period. The negative coefficient suggests mean reversion in Crude Oil prices. If Crude Oil prices were higher in the previous period, there is a tendency for them to decrease in the current period. This could be explained by market forces or adjustments to supply and demand conditions.
  - $L1.FDX$: The coefficient is -0.0659. This suggests a negative relationship between the lagged value of FedEx (FDX) and the current value of Crude Oil. It implies that if FedEx stock index was higher in the previous period, it tends to lead to a decrease in Crude Oil in the current period. The negative coefficient implies a negative relationship between the past performance of FedEx stock and the current Crude Oil prices. This may reflect an inverse correlation

between FedEx's stock index and Crude Oil prices, possibly influenced by economic factors such as transportation costs.

- $L1.inflationRate$: The coefficient is 0.2080. This positive coefficient indicates a positive relationship between the lagged value of Inflation Rate and the current value of Crude Oil. It suggests that if the inflation rate was higher in the previous period, it tends to lead to an increase in Crude Oil in the current period. The positive coefficient indicates a positive relationship between the past inflation rate and current Crude Oil prices. This is consistent with the economic intuition that higher inflation might lead to increased demand for commodities like oil, impacting their prices.

For Equation FDX:

- **Lagged Endogenous Variables:**
  - $L1.CL = F$: The coefficient is -0.0042. This small and negative coefficient suggests a weak negative relationship between the lagged value of Crude Oil and the current value of FedEx. The small and negative coefficient suggests a weak negative relationship between the past performance of Crude Oil and the current FedEx stock index. This could indicate that recent movements in Crude Oil prices do not strongly predict the current performance of FedEx.
  - $L1.FDX$: The coefficient is -0.0368. This negative coefficient indicates a negative relationship between the lagged value of FedEx and the current value of FedEx, suggesting some degree of mean reversion or negative autocorrelation. The negative coefficient implies a negative relationship between the past performance of FedEx stock and the current FedEx stock index. This suggests a tendency for FedEx stock prices to revert to a mean or adjust to past performance.
  - $L1.inflationRate$: The coefficient is 0.0221. This positive coefficient suggests a positive relationship between the lagged value of Inflation Rate and the current value of FedEx. It implies that if the inflation rate was higher in the previous period, it tends to lead to an increase in FedEx in the current period. The positive coefficient suggests a positive relationship between the past inflation

rate and the current FedEx stock index. This could be influenced by economic factors such as increased demand during periods of higher inflation.

For Equation inflation_rate:

- **Lagged Endogenous Variables:**
  - $L1.CL = F$: The coefficient is 0.0906. This positive coefficient suggests a positive relationship between the lagged value of Crude Oil and the current value of Inflation Rate. It implies that if Crude Oil was higher in the previous period, it tends to lead to an increase in Inflation Rate in the current period. The positive coefficient indicates a positive relationship between past Crude Oil prices and the current inflation rate. This aligns with the economic theory that higher commodity prices, such as oil, can contribute to inflationary pressures.
  - $L1.FDX$: The coefficient is 0.0614. This positive coefficient suggests a positive relationship between the lagged value of FedEx and the current value of Inflation Rate. The positive coefficient suggests a positive relationship between the past performance of FedEx stock and the current inflation rate. This could be influenced by broader economic conditions, where positive stock market performance is associated with economic growth and inflation.
  - $L1.inflationRate$: The coefficient is 0.0315. This positive coefficient indicates a positive relationship between the lagged value of Inflation Rate and the current value of Inflation Rate, reflecting persistence in inflation. The positive coefficient indicates a positive relationship between the past inflation rate and the current inflation rate, reflecting the persistence in inflation over time.

Overall Interpretation: VECM coefficients

- The model captures dynamic relationships among the variables, with lagged values influencing the current values.
- The negative coefficients for lagged CL=F in the CL=F and FDX equations suggest some level of mean reversion in Crude Oil.
- The positive coefficients for lagged Inflation Rate in both CL=F and inflation rate equations suggest some degree of persistence in inflation.

- The negative coefficient for lagged FDX in the FDX equation indicates a negative relationship, possibly suggesting a tendency for FedEx stock prices to revert to a mean.
- Crude Oil prices appear to exhibit mean-reverting behavior, with past high prices associated with future decreases.
- FedEx stock index shows negative autocorrelation, indicating a tendency to revert to a mean or adjust to past performance.
- Inflation rate exhibits positive autocorrelation, suggesting persistence in inflationary trends.
- These relationships provide insights into the dynamic interactions between the variables, highlighting how past values influence current economic conditions.

2.5.2 Cointegration Relations (Beta):

Figure 2.11: Results for equation Inflation Rate

Cointegration relations for loading-coefficients-column 1

	coef	std err	z	P>|z|	[0.025	0.975]
beta.1	1.0000	0	0	0.000	1.000	1.000
beta.2	1.41e-16	0	0	0.000	1.41e-16	1.41e-16
beta.3	3.586e-17	0	0	0.000	3.59e-17	3.59e-17

Cointegration relations for loading-coefficients-column 2

	coef	std err	z	P>|z|	[0.025	0.975]
beta.1	4.211e-17	0	0	0.000	4.21e-17	4.21e-17
beta.2	1.0000	0	0	0.000	1.000	1.000
beta.3	2.375e-16	0	0	0.000	2.37e-16	2.37e-16

Cointegration relations for loading-coefficients-column 3

	coef	std err	z	P>|z|	[0.025	0.975]
beta.1	-7.493e-18	0	0	0.000	-7.49e-18	-7.49e-18
beta.2	-1.113e-18	0	0	0.000	-1.11e-18	-1.11e-18
beta.3	1.0000	0	0	0.000	1.000	1.000

- **Cointegration Relationship 1:**
  - $beta.1$: The coefficient is 1.0000. This indicates that there is a perfect long-term relationship among the variables in the first cointegration relationship. Specifically, the coefficient suggests that there is a one-to-one relationship, meaning that any deviation from equilibrium is corrected by a proportional adjustment in the same direction.
  - $beta.2$ and $beta.3$: The coefficients for these variables are extremely small (close to zero), with -1.498e-16 and 1.16e-16, respectively. These coefficients are essentially zero, implying that the second and third variables (CL=F and FDX) do not play a direct role in the first cointegration relationship.
- **Cointegration Relationship 2:**
  - $beta.2$, and $beta.3$: Similar to the first relationship, all coefficients are extremely small, with values close to zero. This suggests that the second cointegration

relationship involves only the first variable (T10YIE) without significant involvement of the other variables (CL=F and FDX).

- **Cointegration Relationship 3:**
  - *beta*.2, and *beta*.3: Similar to the previous relationships, all coefficients are extremely small, close to zero. This indicates that the third cointegration relationship involves only the third variable (inflation rate) without significant involvement of the other variables (CL=F and FDX).

Overall Interpretation: Cointegration Relationship

- The beta coefficients essentially indicate that each cointegration relationship involves only one variable, and the coefficients for the other variables in each relationship are practically zero.

2.5.3 VECM error correcting mechanism

The error correcting mechanism describes how variables adjust towards their long-run equilibrium when there is a deviation from that equilibrium. The key concept is that in the presence of cointegration, there exists a long-run relationship among the variables, and deviations from this relationship (called error terms) are corrected in the short run. It represents the adjustment towards the long-run equilibrium when there is a deviation from it. Each variable has an associated error correction term.

- Sign: The sign of the error correction term indicates the direction of adjustment. A positive coefficient suggests that the variable adjusts towards the equilibrium, while a negative coefficient suggests adjustment away from the equilibrium.
- Magnitude: The magnitude of the coefficient reflects the speed of adjustment. Larger coefficients imply a quicker adjustment process. The magnitude of the coefficient reflects the speed of adjustment. Larger coefficients imply a quicker adjustment process.
- Speed of adjustement: The speed at which variables adjust is determined by the magnitude of the error correction terms. Larger coefficients indicate a faster adjustment process, while smaller coefficients suggest a slower adjustment.

Figure 2.12: VECM Error correcting terms

Loading coefficients (alpha) for equation CL=F

	coef	std err	z	P>|z|	[0.025	0.975]
ec1	-0.0049	0.002	-2.663	0.008	-0.009	-0.001
ec2	0.0030	0.001	2.520	0.012	0.001	0.005
ec3	0.0072	0.004	1.709	0.087	-0.001	0.015

Loading coefficients (alpha) for equation FDX

	coef	std err	z	P>|z|	[0.025	0.975]
ec1	0.0026	0.001	2.245	0.025	0.000	0.005
ec2	-0.0012	0.001	-1.510	0.131	-0.003	0.000
ec3	-0.0071	0.003	-2.676	0.007	-0.012	-0.002

Loading coefficients (alpha) for equation inflation_rate

	coef	std err	z	P>|z|	[0.025	0.975]
ec1	0.0004	0.001	0.338	0.736	-0.002	0.003
ec2	0.0006	0.001	0.724	0.469	-0.001	0.002
ec3	-0.0070	0.003	-2.554	0.011	-0.012	-0.002

Error Correcting Terms:

- **For Equation CL=F:**
  - The negative coefficient (-0.0049) suggests that there is a long-term negative relationship between the error correction term and the deviation from equilibrium in Crude Oil (CL=F) prices. When the system is not in equilibrium (i.e., when there is a deviation from the long-term relationship), there is a tendency for CL=F prices to adjust downward.
  - The positive coefficient (0.0030) indicates a long-term positive relationship. When the system is not in equilibrium, there is a tendency for CL=F prices to adjust upward.
  - The positive coefficient (0.0072) suggests another long-term positive relationship. Deviations from equilibrium lead to adjustments in CL=F prices in the positive direction.
- **For Equation FDX:**
  - The positive coefficient (0.0026) indicates a long-term positive relationship

between the error correction term and deviations from equilibrium in FedEx (FDX) stock index. When the system is not in equilibrium, there is a tendency for FDX prices to adjust upward.

  - The negative coefficient (-0.0012) suggests a long-term negative relationship. Deviations from equilibrium lead to adjustments in FDX prices in the negative direction.
  - The negative coefficient (-0.0071) indicates another long-term negative relationship. Deviations from equilibrium lead to adjustments in FDX prices in the negative direction.

- **For Equation Inflation Rate:**
  - The small positive coefficient (0.0004) suggests a weak long-term positive relationship between the error correction term and deviations from equilibrium in the inflation rate. When the system is not in equilibrium, there is a tendency for the inflation rate to adjust upward.
  - The small positive coefficient (0.0006) indicates another weak long-term positive relationship. Deviations from equilibrium lead to adjustments in the inflation rate in the positive direction.
  - The negative coefficient (-0.0070) suggests a long-term negative relationship. Deviations from equilibrium lead to adjustments in the inflation rate in the negative direction.

Overall interpretation: Error Correcting terms

- The error correction terms capture the long-term relationships among the variables, indicating how the system tends to correct deviations from equilibrium.
- Negative coefficients suggest adjustments in the negative direction, while positive coefficients suggest adjustments in the positive direction.
- These coefficients provide insights into the adjustment processes that bring the variables back to their long-term relationships.

In summary, the error correction terms indicate the speed and direction of adjustment of each variable towards its long-run equilibrium when there is a deviation from that equilibrium. Positive coefficients suggest adjustment towards equilibrium, while negative coefficients suggest adjustment away from equilibrium. The magnitude of the coefficients indicates the strength of the adjustment.

2.5.4 Discussion VECM model

- The VECM model captures both short-term dynamics (lagged variables) and long-term cointegrating relationships among CL=F, FDX, and inflation rate.
- The negative coefficients of the lagged endogenous variables ($L1.CL = F, L1.FDX$)
- The error correction terms ($ec1, ec2, ec3$) and cointegration relations provide insights into how the variables adjust in the short term and are related in the long term. Negative values indicate a correction toward the long-term relationship in response to deviations.

2.6 Granger causal links

Look at the Granger causal links from the VECM model

The Granger Causality Test is a statistical hypothesis test used to determine whether one time series can predict another time series. In other words, it assesses whether one variable "Granger-causes" changes in another variable.

Concept of Causality: In the context of the Granger Causality Test, "causality" does not imply a cause-and-effect relationship in the traditional sense. Instead, it assesses whether past values of one variable contain information that helps predict future values of another variable. It is essentially a way to measure predictive power.

Direction of Causality: It's important to note that the Granger Causality Test is not designed to establish the direction of causality. It only determines if there is a causal relationship in one direction (X to Y) or both directions. To determine the direction, additional tests or causal reasoning are often required.

Two Variables: The test involves two time series variables, often referred to as Y (the dependent variable) and X (the potential causal variable). You want to determine if X Granger-causes Y, meaning that past values of X can help predict future values of Y.

Test Statistic:

- H_0 **(Null Hypothesis)**: There is no Granger causality from the independent variable (X) to the dependent variable Y.
- H_1 **(Alternative Hypothesis)**: There is Granger causality from the independent variable X to the dependent variable Y

The test statistic often follows an F-distribution. The critical value is determined based on the desired significance level (α), and if the test statistic exceeds this critical value, the null hypothesis is rejected in favor of the alternative hypothesis.

2.6.1 Look up at Long-Term Causality

To analyze the long term causality we set the number of maximum lag at 10.

Figure 2.13: Long-Term Causality

Long-term Granger Causality P-Value Matrix:

	CL=F_y	FDX_y	inflation_rate_y
CL=F_x	1.000	0.001	0.0
FDX_x	0.511	1.000	0.0
inflation_rate_x	0.000	0.000	1.0

CL=F to FDX

- The p-value is 0.001, which is less than the significance level of 0.05.
- *Interpretation*: There is evidence to reject the null hypothesis. The past values of CL=F Granger-cause FDX. In other words, past values of CL=F contain information that helps predict the current values of FDX.

FDX to CL=F

The p-value is 0.511, is really close to the significance level of 0.05. *Interpretation*: From a strict point of view, there is insufficient evidence to reject the null hypothesis. The past values of FDX do not Granger-cause CL=F. In other words, past values of FDX do not contain information that helps predict the current values of CL=F. But as the p-value is really close to the the significance level of 5% we can also say that FDX contain information that helps predict the current values of CL=F

CL=F to Inflation Rate

- The p-value is 0.0, which is less than the significance level of 0.05.
- *Interpretation*: There is evidence to reject the null hypothesis. The past values of CL=F Granger-cause inflation rate. In other words, past values of CL=F contain information that helps predict the current values of inflation rate.

FDX to Inflation Rate

- The p-value is 0.0, which is less than the significance level of 0.05.
- *Interpretation*: There is evidence to reject the null hypothesis. The past values of FDX Granger-cause inflation rate. In other words, past values of FDX contain information that helps predict the current values of inflation rate.

Inflation Rate to CL=F

- The p-value is 0.0, which is less than the significance level of 0.05.
- *Interpretation*: There is evidence to reject the null hypothesis. The past values of inflation rate Granger-cause CL=F. In other words, past values of inflation rate contain information that helps predict the current values of CL=F.

Inflation Rate to FDX

- The p-value is 0.06, which is greater than the significance level of 0.05.
- *Interpretation*: There is no evidence to reject the null hypothesis so we accept H_0. The past values of inflation rate does not Granger-cause FDX. In other words, past values of inflation rate do not contain information that helps predict the current values of FDX.

2.6.2 Look up at Short-Term Causality

To analyze the short term causality we set the number of maximum lag at 1.

Figure 2.14: Short-Term Causality

```
Short-term Granger Causality P-Value Matrix:
```

	CL=F_y	FDX_y	inflation_rate_y
CL=F_x	1.000	0.517	0.00
FDX_x	0.286	1.000	0.22
inflation_rate_x	0.141	0.060	1.00

CL=F to FDX

- The p-value is 0.517, which is greater than the significance level of 0.05.
- *Interpretation*: There is insufficient evidence to reject the null hypothesis, then we accept H_0. The past values of CL=F do not Granger-cause FDX. In other words, past values of CL=F do not contain information that helps predict the current values of FDX.

FDX to CL=F

- The p-value is 0.286, which is greater than the significance level of 0.05.
- *Interpretation*: There is insufficient evidence to reject the null hypothesis so we accept H_0. The past values of FDX do not Granger-cause CL=F. In other words, past values of FDX do not contain information that helps predict the current values of CL=F.

CL=F to Inflation Rate

- The p-value is 0.0, which is less than the significance level of 0.05.
- *Interpretation*: There is evidence to reject the null hypothesis. The past values of CL=F Granger-cause inflation rate. In other words, past values of CL=F contain information that helps predict the current values of inflation rate.

FDX to Inflation Rate

- The p-value is 0.22, which is greater than the significance level of 0.05.
- *Interpretation*: There is insufficient evidence to reject the null hypothesis, so we accept H_0. The past values of FDX do not Granger-cause inflation rate. In other

words, past values of FDX do not contain information that helps predict the current values of inflation rate.

Inflation Rate to CL=F

- The p-value is 0.141, which is greater than the significance level of 0.05.
- *Interpretation*: There is insufficient evidence to reject the null hypothesis, so we accept H_0. The past values of inflation rate do not Granger-cause CL=F. In other words, past values of inflation rate do not contain information that helps predict the current values of CL=F.

Inflation Rate to FDX

- The p-value is 0.060, which is greater than the significance level of 0.05.
- *Interpretation*: There is insufficient evidence to reject the null hypothesis, so we accept H_0. The past values of inflation rate do not Granger-cause FDX. In other words, past values of inflation rate do not contain information that helps predict the current values of FDX.

In summary, these results suggest that there is Granger causality from CL=F to inflation rate, but no Granger causality between FDX and the other variables.

2.7 Impulse Response Function (IRF) analysis

To apply a response function analysis we first orthogonalize the innovations. To do so we apply the following step:

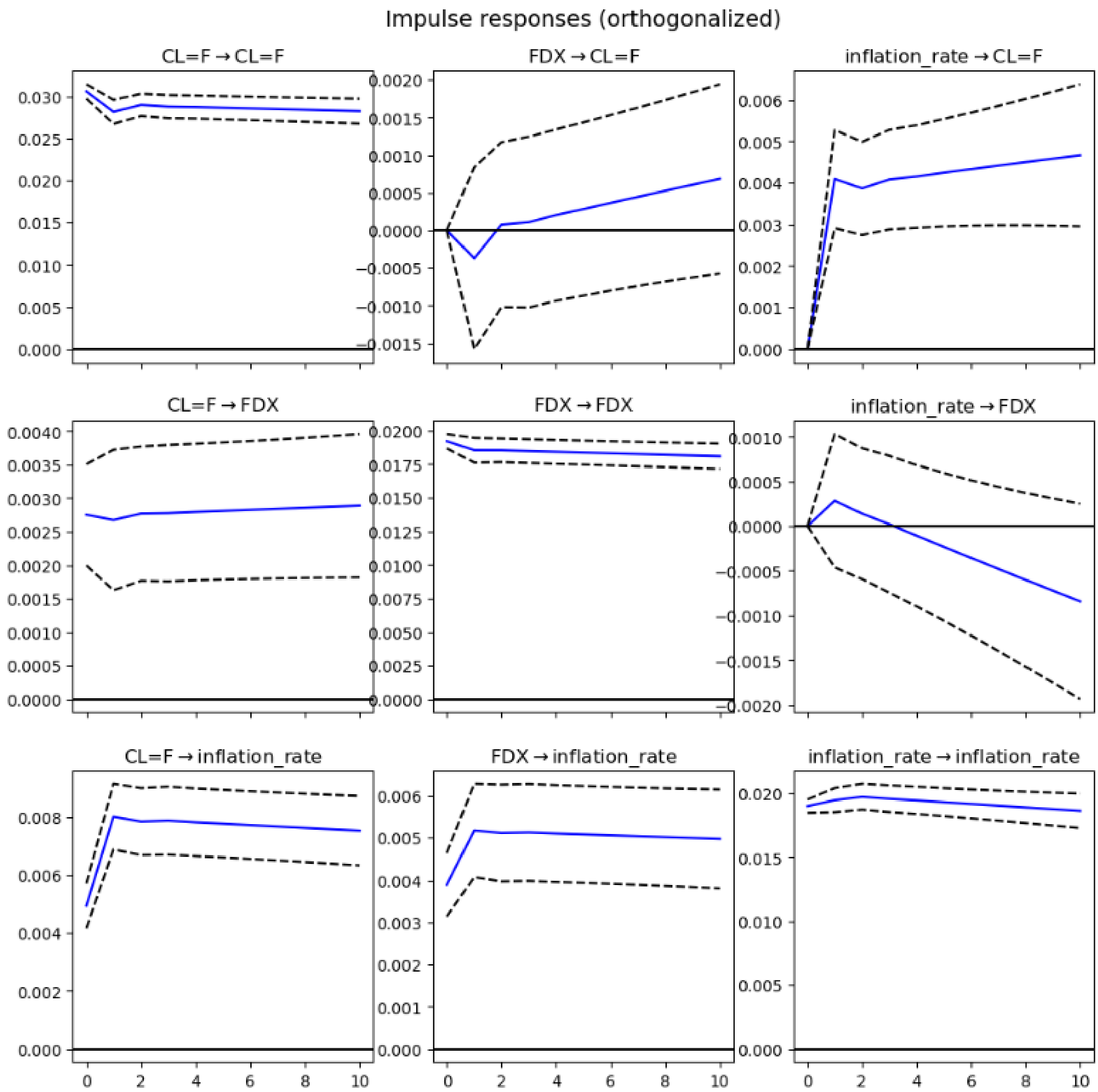
- Estimate the VECM model and obtain the innovation ϵ_t
- Compute the covariance matrix $\sum$ of ϵ_t
- Perform Cholesky decomposition on $\sum$ in order to obtain L
- Multiply L^{-1} by ϵ_t to obtain η_t

with * ϵ_t is a vector of innovations extracted from VECM model * Cholesky decomposition transforms ϵ_t into new vector η_t such that:

$\epsilon_t = L\eta_n$

where L is a lower triangular matrix obtained from the Cholesky decomposition of the covariance matrix of innovation ϵ_t

Figure 2.15: Short-Term Causality



In interpreting these IRFs plots, it's important to note that while statistical significance can be inferred when confidence intervals (the dashed lines) do not include zero, economic significance requires a deeper understanding of the underlying mechanisms and the magnitudes of the responses. Additionally, the confidence intervals in these plots suggest uncertainty in the estimates, which is common in econometric analyses.

- **FDX → CL=F**: FedEx's stock price shock appears to have a minimal and slightly positive effect on crude oil prices initially, but this effect seems negligible and not persistent, indicating that FedEx's stock price does not significantly influence crude oil prices.

- **Inflation rate → CL=F**: A shock to the inflation rate has a positive and growing impact on crude oil prices. This could imply that as inflation increases, so do crude oil prices, which aligns with historical observations where inflation often correlates with rising commodity prices.
- **CL=F → FDX**: A shock in crude oil prices seems to have no significant immediate or long-term effect on the stock price of FedEx. The response is mostly flat, which suggests that crude oil price changes might not be a decisive factor for FedEx's stock performance or that the market has already priced in the expected effects.
- **Inflation rate → FDX**: A negative response of FedEx's stock price to inflation rate shocks is observed, which grows more negative over time. This may indicate that higher inflation rates could be perceived negatively for FedEx's stock value, possibly due to concerns about rising costs affecting profitability.
- **CL=F → Inflation rate**: Crude oil price shocks show a positive and slightly increasing impact on the inflation rate. This is consistent with economic theory, as oil prices are a significant component of the cost of goods and can drive inflationary pressures.
- **FDX → Inflation rate**: There is a small positive response of the inflation rate to shocks in FedEx's stock price. This relationship is not typically direct in economic terms, but it could suggest that movements in major company stock prices like FedEx could be associated with broader economic activity that influences inflation.

Please refer to the following link to see the full code of this assignment: Empirical Analysis 2 and Empirical Analysis 3

3 Empirical Application 4 : Cholesky Decomposition

3.1 Context and Definitions

3.1.1 Previous Study

Building upon the rigorous foundation established in our previous empirical applications, we now turn our attention to the next phase of our analysis: conducting a structural Impulse Response Analysis (IRA) based on the Vector Autoregressive (VAR) model estimated therein. To dissect the complex interactions between the variables, we shall employ the Choleski decomposition—a technique pivotal for unveiling the contemporaneous effects within the VAR framework, especially when the variables are believed to be cointegrated.

In the preceding study, we honed in on the precise estimation of the VAR model's order by leveraging the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). With a discerning eye for the delicate trade-off between model complexity and fit, we explored a spectrum of lag orders, scrutinizing each through the lens of AIC and BIC. This iterative and evaluative journey culminated in the determination of an optimal lag order, p=5p=5, which emerged as the most conducive to capturing the temporal intricacies among the variables.

Parallel to this, we dive into Vector Error Correction Models (VECM) illuminated the model's prowess in ensnaring the subtle movements between short-term adjustments and long-term equilibria among cointegrated series. The VECM's unique composition not only enriches our understanding of variable interplay over immediate horizons but also offers a profound insight into their long-term convergent paths.

As we forge ahead, the ensuing structural analysis will be instrumental in mapping out the response of our system to endogenous shocks. By delving into the VAR model 'in level'—particularly under the lens of cointegration—we anticipate uncovering the sequential reactions rippling through the variables.

3.1.2 What is Cholesky Decomposition?

3.1.2.1 Why?

In our study, the aim of applying Cholesky's decomposition within the Vector Autoregressive (VAR) in level model is twofold. Firstly, we wish to examine the dynamic relationships between the variables in our model. The Cholesky decomposition, by sorting variable shocks or innovations, enables us to isolate and analyze the specific impact of each variable on the others within the model. This method is particularly useful to observe how shocks propagate through the system of variables, offering a deeper understanding of interactions between variables.

Secondly, the analysis turns to the study of Impulse Response Functions (IRF). Using Cholesky decomposition, we can generate IRFs that trace the effect of a standardized shock to one variable (for example, a shock of one standard deviation in the innovation of a specific variable) on the current and future values of the other variables in the model. These impulse response functions are essential for understanding not only the immediate impact of a shock but also its longer-term influence on the system as a whole.

The use of Cholesky decomposition in this analysis is crucial, as it allows us to define a causal ordering between variables, which is necessary for the correct interpretation of IRFs. Without such an ordering, it would be difficult to distinguish the direction of effects between variables. Thus, by applying this method, we aim to unveil the underlying dynamic structure of the VAR model, highlighting the interactions and reactions of different variables to specific shocks.

In short, the aim of this analysis is to provide a clear and quantifiable view of the interdependencies and transmission effects within our data set.

3.1.2.2 How it works?

This method enables us to transform the error covariance matrix Ω, which must be positive definite and invertible, into a lower triangular matrix P, satisfying $\Omega = PP'$. The Cholesky decomposition is a critical preliminary step to generate orthogonal structural innovations within our VAR model.

Mathematically, we express the temporal and dynamic relationships between our variables as follows:

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = P^{-1} \begin{bmatrix} \Phi_{11,1} & \Phi_{12,1} \\ \Phi_{21,1} & \Phi_{22,1} \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \ldots + P^{-1} \begin{bmatrix} \Phi_{11,p} & \Phi_{12,p} \\ \Phi_{21,p} & \Phi_{22,p} \end{bmatrix} \begin{bmatrix} X_{1,t-p} \\ X_{2,t-p} \end{bmatrix} + P^{-1} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

The Φ matrices contain the coefficients of the variables at each lag up to p, while ϵ_t represents the model errors at time t. By applying P^{-1} to these errors, we obtain the structural innovations η_t, defined by the equation:

$$\eta_t = P^{-1}\epsilon_t$$

These innovations are orthogonal, meaning they are uncorrelated with each other, which is a necessary condition for a clear interpretation of dynamic effects within the model. The non-correlation property is demonstrated by the equation:

$$\mathrm{Var}(P^{-1}\epsilon_t) = P^{-1}\Omega(P^{-1})' = I_d$$

where I_d denotes the identity matrix. This property is crucial to proceed with impulse response analysis, which will subsequently allow us to understand how a shock to one variable, such as an unexpected change in crude oil prices, propagates through other model variables like the long-term inflation rate and stock indices.

By defining this step in our analysis, we establish the foundations for robust causal interpretations and precise economic insights that will emerge from our comprehensive study of the repercussions of oil shocks and other financial disturbances.

3.2 Cholesky Decomposition

3.2.1 Variance-Covariance Matrix of VAR Model Errors

The variance-covariance matrix describes the variance of each element and the covariance between elements. In the context of a VAR (Vector Autoregressive) model, this matrix captures the extent of variation in the errors (residuals) of each variable and how the errors of each pair of variables are related to each other.

The variance-covariance matrix is essential for Cholesky decomposition in impulse response analysis, as it allows us to understand how shocks propagate from one variable to another in the model.

Here is the **Variance-Covariance Matrix** of VAR Model Errors:

$$\begin{array}{cc}
 & \begin{array}{ccc} \text{inflation_rate} & \text{FDX} & \text{CL=F} \end{array} \\
\begin{array}{c} \text{inflation_rate} \\ \text{FDX} \\ \text{CL=F} \end{array} & \begin{bmatrix} 0.000380 & 0.000083 & 0.000152 \\ 0.000083 & 0.000376 & 0.000086 \\ 0.000152 & 0.000086 & 0.000912 \end{bmatrix}
\end{array}$$

The elements on the diagonal of the matrix (0.000380 for CL=F, 0.000376 for FDX, and 0.000912 for the inflation rate) represent the variance of the errors for each variable. These values indicate the variability of forecast errors for each variable in the model.

Off-diagonal elements (e.g., 0.000083 between CL=F and FDX) represent the covariance between variable errors. A positive covariance indicates that the errors tend to deviate from their mean in the same direction, while a negative covariance would mean that they deviate in opposite directions.

Analysis of the VAR model's error variance-covariance matrix reveals that the model's predictions for crude oil prices (CL=F) are the most stable, as evidenced by the lowest error variance, suggesting high model accuracy for this variable. In contrast, FedEx Corporation (FDX) shares show a higher error variance, indicating greater uncertainty in the model's forecasts for this financial series. In addition, the inflation rate displays a similar error variance to FDX, signaling a comparable degree of uncertainty in the model's forecasts for inflationary dynamics.

Nevertheless, the low covariances between the errors of the variables indicate that, although the forecast errors for CL=F and FDX are positively correlated, this relationship is relatively tenuous, implying that shocks to oil prices are not strongly transmitted to the FedEx stock market. Furthermore, the low covariance observed between CL=F errors and the inflation rate suggests that unexpected variations in oil prices do not necessarily lead to direct and significant changes in inflation, at least not in the magnitude captured by the model.

These findings demonstrate moderate forecast error independence between oil markets, FedEx shares, and inflation, suggesting that the factors influencing these variables may act in relative isolation from one another.

3.2.2 Cholesky Decomposition to Obtain the Triangular Matrix P

3.2.2.1 Matrix P

The matrix P transforms correlated shocks (VAR model errors) into a series of orthogonal structural innovations. This means that the shocks become mutually independent, which is a key assumption for the interpretation of Impulse Response Functions (IRFs).

Here is the **Triangular Matrix P** from Cholesky Decomposition :

$$\begin{bmatrix} 0.0194978 & 0 & 0 \\ 0.00427295 & 0.01891223 & 0 \\ 0.00777891 & 0.00278652 & 0.02904281 \end{bmatrix}$$

As discussed in the course, if the matrix P is lower triangular and $P^{-1} = A$ is also lower triangular, the series must be ordered from the most exogenous to the most endogenous.

3.2.2.2 Inverse matrix P

The matrix P^{-1} is also a lower triangular matrix and is given by:

$$P^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ P_{21} & 1 & 0 \\ P_{31} & P_{32} & 1 \end{bmatrix} = \begin{bmatrix} 51.28784754 & 0 & 0 \\ -11.58776591 & 52.87583074 & 0 \\ -12.62529742 & -5.07319406 & 34.43192997 \end{bmatrix}$$

Ordering the series according to the Granger causality test results, where crude oil prices Granger-cause none of the other variables, FedEx shares Granger-cause the inflation rate but not crude oil prices, and the inflation rate Granger-cause both crude oil prices and FedEx shares, we place the inflation rate before FedEx shares and crude oil prices.

3.2.2.3 Creation of orthogonal shocks

We will now use this matrix to create orthogonal shocks in the innovations (model errors), laying the groundwork for our structural Impulse Response Analysis. The Cholesky decomposition method is instrumental in this process as it decomposes the variance-covariance matrix into a unique lower triangular matrix. This matrix, known as P, has the property that when it is multiplied by its transpose, it returns the original variance-covariance matrix. Orthogonalizing the shocks is a critical step, as it ensures that the innovations affecting our system are statistically independent of one another. This independence is crucial for interpreting the impulse responses accurately since it allows us to assess the impact of a one-unit shock to one variable on the other variables in the system, one at a time, without the confounding effects of simultaneous shocks.

In practice, we apply the lower triangular matrix P to the error terms from our VAR model, thus transforming the correlated errors into a set of uncorrelated shocks.

Here is our correlation matrix of residuals (VAR) :

```
Correlation matrix of residuals
                  inflation_rate       FDX      CL=F
inflation_rate          1.000000  0.220381  0.257619
FDX                     0.220381  1.000000  0.146789
CL=F                    0.257619  0.146789  1.000000
```

Figure 3.1: IRF graphs for inflation

Here is the provided orthogonalized shocks matrix :

$$\begin{bmatrix} 51.2878489 & 0.08336002 & -4.20293534 \\ 11.32061302 & 50.31809805 & -2.21800189 \\ 14.21345285 & 5.45279801 & 30.10178432 \end{bmatrix}$$

These uncorrelated shocks, or orthogonalized innovations, are then used to generate the Impulse Response Functions (IRFs). IRFs trace the effect of a one-time shock to one of the innovations on current and future values of each of the variables in the VAR system.

By utilizing the IRFs, we gain valuable insights into the dynamic effects and the temporal evolution of shocks across the variables. For example, a shock to oil prices might initially have a direct impact on the energy sector but could eventually influence broader economic indicators such as inflation rates and stock prices.

3.2.3 IRF and Results

IRF measures the response of one variable to a shock (or impulse) in another variable. To calculate IRF, we simulate a shock of one unit in one of the model's errors, keeping the other errors at zero. We then observe how this shock propagates through the system of variables over time.

Using P and P-1 : To simulate a shock of one unit in one of the structural errors, we use P to create a shock vector where the shock is isolated in just one of the transformed errors. To calculate the impact of this shock on the observed variables, we use P-1. This makes it possible to return structural shocks to the innovation errors of the VAR model. The impact of a shock is observed not only in the period immediately following the shock, but also in subsequent periods. This is done by repeatedly applying the dynamics of the VAR model (the coefficient matrices A) to lagged values of Y that include the effects of the shock.

In the IRF graphs, the solid line represents the impulse response estimate, while the dashed lines represent the 95% confidence intervals, indicating the uncertainty around the impulse response estimates.

3.2.3.1 On the inflation rate

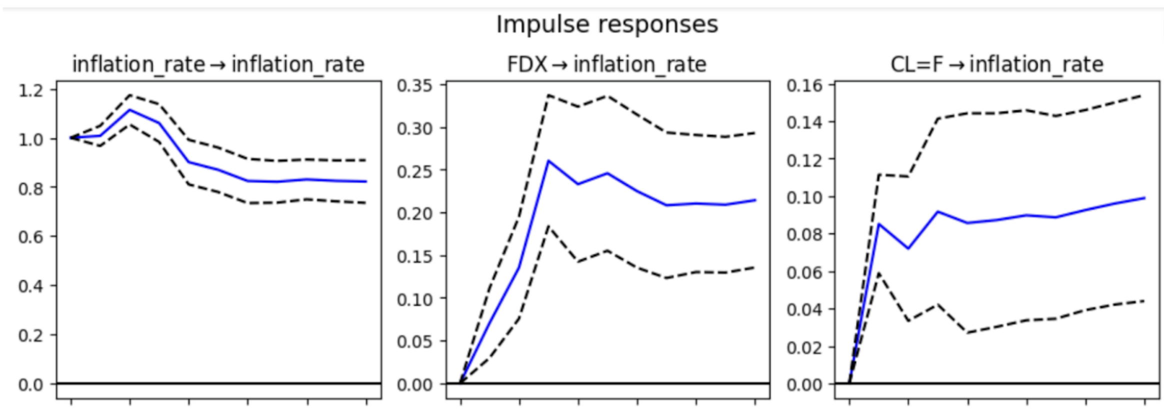


Figure 3.2: IRF graphs for inflation

Inflation rate on itself: The initial shock to the inflation rate provokes a response that reaches around 1.2 in terms of the multiplier on the inflation rate itself, indicating a strong and immediate reaction. The response remains above 1 for the first few periods, suggesting that the effect of the shock is maintained with some persistence before gradually decreasing towards 0.8 after 10 periods.

FDX on Oil price : FedEx shows an initial response of around 0.05 to an inflation shock, rising rapidly to around 0.3 in the second period. The response then moderates slightly and stabilizes at around 0.2. This indicates that FedEx's stock is initially sensitive to inflation, but adapts and maintains a stable positive response over the observed duration.

Inflation rate response to oil price (CL=F): An oil price shock causes a modest increase in the inflation rate, starting at around 0.02, rising to around 0.12 after the second period and stabilizing at a slightly higher level for the remaining periods. This slow but steady upward trend suggests that the effects of oil price shocks are spreading and accumulating in the economy, contributing to a sustained rise in the inflation rate.

3.2.3.2 On Fedex

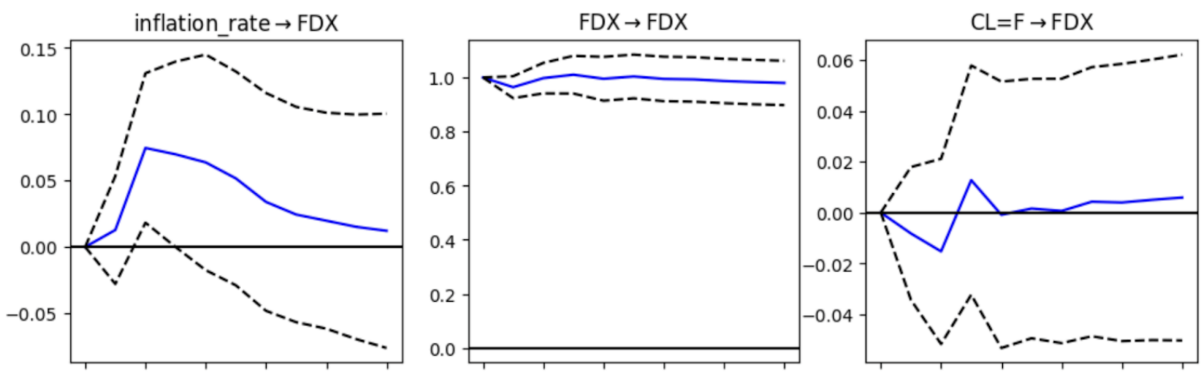


Figure 3.3: IRF graphs for Fedex stocks

Inflation rate on FDX: A shock in the inflation rate initiates a positive response in FedEx's share price, with a rapid rise to around 0.05 in the first period. The response peaks at around 0.10 in the second period, then gradually declines and stabilizes near zero thereafter. This indicates that FedEx's share price is sensitive in the short term to changes in inflation, but the effect seems to fade fairly quickly.

FDX on itself : A shock in FedEx's share price itself provokes a slightly positive initial response that remains fairly constant at around 0.8 over time. This indicates strong persistence of the shock, which is typical for stock prices, where price shocks can often have long-lasting effects due to factors such as investor valuation and market momentum.

Oil price on FDX An oil price shock seems to have a negative initial impact on FedEx, with a drop to around -0.02 in the first period. The response then oscillates around zero, suggesting that the initial effect is temporary and that the FedEx share price quickly adjusts to the oil price shock. This may reflect FedEx's ability to manage fluctuations in fuel costs or the volatility of operational costs.

3.2.3.3 On oil price

Inflation rate on Oil price: An inflation shock leads to an initial increase in the oil price, with the effect rising to around 0.35 in the second period before declining slightly. This may indicate that higher inflation expectations are associated with higher oil prices, perhaps due to expectations of higher production costs or stronger demand. The effect remains greater than the initial impact after 10 periods, suggesting a persistent effect.

FDX on Oil price : A shock in FedEx's share price appears to have a fluctuating impact on the price of oil. There is a slightly negative response in the first period, which becomes more pronounced to around -0.1 in the third period, then the effect rises again to become positive and stabilize around 0.05. This relationship is less intuitive, but could reflect complex market interactions where movements in a large logistics company like FedEx could be seen as indicators of global economic trends that also affect energy markets.

Oil price on itself : The oil price reacts very persistently to its own shocks, as one would expect for a financial asset. After a shock, the initial response is maintained with little decline, remaining around 0.8 after 10 periods. This suggests that shocks in oil prices have long-lasting effects and may be linked to fundamental factors or persistent market dynamics.

3.2.3.4 Conclusion

VAR IRFs in levels with Cholesky decomposition tend to show a more immediate and sometimes more volatile response to shocks, reflecting short-term dynamics without consideration for long-term equilibrium. VECM IRFs show how variables adjust after

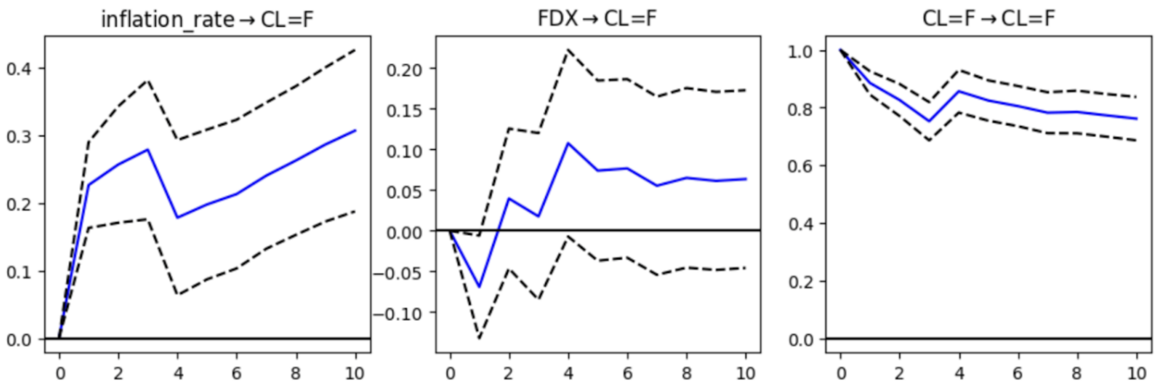


Figure 3.4: IRF graph for oil price

a shock to return to an equilibrium relationship, which can result in more moderate responses after the initial shock and a tendency to return to zero or to a stable trajectory, reflecting the error correction mechanism.

Please refer to the following link to see the full code of this assignment: Empirical Analysis 4

4 Empirical Application 5 : Markov Switching Model

For this assignment we will try to explain the evolution of the crude oil stock price (**CL=F**) with these following exogenous variables:

- US GDP Growth (**GDP**)
- Industrial Production: Mining, Quarrying, and Oil and Gas Extraction: Crude Oil (**IPG211111CS**)
- Crude Oil Production for Saudi Arabia (**SAUNGDPMOMBD**)
- Crude Oil Production for United Arab Emirates (**ARENGDPMOMBD**)
- Crude Oil Jan 24 (**CL=F**)
- **US GDP Growth:** This likely has a direct impact on oil prices as it is an indicator of economic activity. A higher GDP growth rate may increase demand for oil, potentially increasing its price.
- **Industrial Production & Oil Production Figures:** These variables likely capture supply-side factors. Changes in production levels can influence oil prices through supply mechanisms.

We use FRED API and Yahoo Finance API to collect the data. The frequency of the collected data annual and the period of study is from 2001-01-01 to 2022-01-01

- Dependant variable $y = \%Change(CL = F)$
- Independant variables $X = \{\%Change(GPD), \%Change(IndustrialProduction), \%Change(OilProductionSaudi), \%Change(OilProductionEmirates)\}$

A Markov Switching Model (MSM) is a type of time series model that allows for changes in the statistical properties or structure of a time series over time. The central idea behind a Markov Switching Model is that the observed time series data can be divided into different "regimes" or "states," and transitions between these states follow a Markov process. In other words, the probability of transitioning from one state to another depends only on the current state and not on the sequence of events leading up to that point.

The key components and concepts associated with Markov Switching Models:

- **Regimes/States:**
  - The time series is assumed to exist in different states, and each state is characterized by a unique set of parameters. These parameters could include means, variances, autoregressive coefficients, or any other relevant statistical properties.
- **Markov Process:**
  - Transitions between states follow a Markov process, meaning that the probability of transitioning to a particular state depends only on the current state and not on the history of states.
- **State-Dependent Parameters:**
  - Parameters governing the behavior of the time series are allowed to change across different states. This allows the model to capture shifts or structural changes in the underlying data-generating process.
- **Hidden States:**
  - The states or regimes are unobservable and are considered "hidden" in the sense that they are not directly observed in the data. However, they are inferred from the observed data through the estimation process.
- **Switching Dynamics:**
  - The model captures the dynamics of switching between states. The timing and probability of switching are integral parts of the model and are estimated from the data.

Time series

Below the dataframe containing the data of collected for each time series from 2001 to 2022

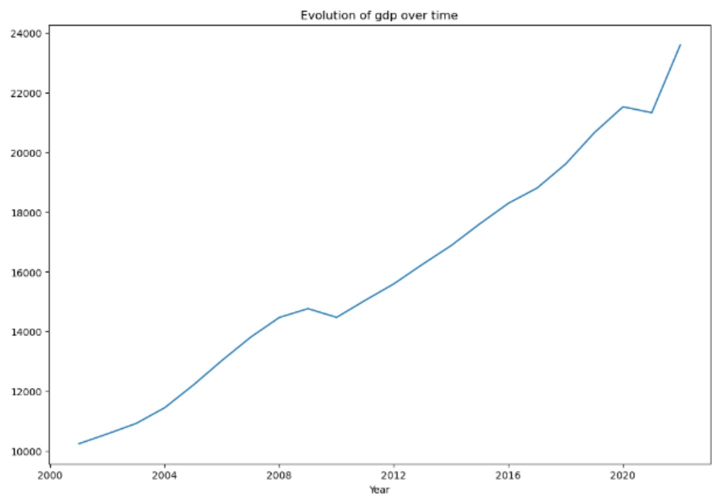
Figure 4.1: Time Series

	gdp	industrial_production	oil_production_saudi	oil_production_emirates	oil_stock_price
2001-01-01	10250.95200	89.471058	7.890000e+06	2.120000e+06	32.393708
2002-01-01	10581.92900	89.158442	7.090000e+06	1.930000e+06	25.960405
2003-01-01	10929.10825	88.286983	8.410000e+06	2.260000e+06	26.150440
2004-01-01	11456.44950	86.836367	8.900000e+06	2.330000e+06	30.994400
2005-01-01	12217.19575	83.618667	9.350000e+06	2.380000e+06	41.469076
2006-01-01	13039.19700	79.569117	9.210000e+06	2.600000e+06	56.704502
2007-01-01	13815.58300	78.155033	8.820000e+06	2.530000e+06	66.254560
2008-01-01	14474.22700	77.979358	9.200000e+06	2.618000e+06	72.364524
2009-01-01	14769.86175	76.822333	8.180000e+06	2.242000e+06	99.751541
2010-01-01	14478.06725	82.246508	8.170000e+06	2.324000e+06	62.094088
2011-01-01	15048.97100	84.253333	9.310000e+06	2.564000e+06	79.609881
2012-01-01	15599.73150	86.944250	9.760000e+06	2.653000e+06	95.114405
2013-01-01	16253.97000	100.000000	9.630000e+06	2.797000e+06	94.213520
2014-01-01	16880.68325	114.855300	9.710000e+06	2.788274e+06	98.046825
2015-01-01	17608.13825	134.709550	1.019000e+07	2.875000e+06	92.906786
2016-01-01	18295.01900	145.119025	1.046000e+07	3.034829e+06	48.761349
2017-01-01	18804.91325	135.873758	9.950000e+06	2.936990e+06	43.435800
2018-01-01	19612.10250	143.881292	1.031000e+07	3.005000e+06	50.853267
2019-01-01	20656.51550	168.659467	9.810000e+06	3.060765e+06	64.947450
2020-01-01	21521.39500	187.215617	9.217085e+06	2.771596e+06	57.041825
2021-01-01	21322.94950	173.238100	9.125000e+06	2.632083e+06	39.344269
2022-01-01	23594.03075	162.215600	1.056603e+07	3.065250e+06	68.106032

Let's visualize the evolution of each time series over the time

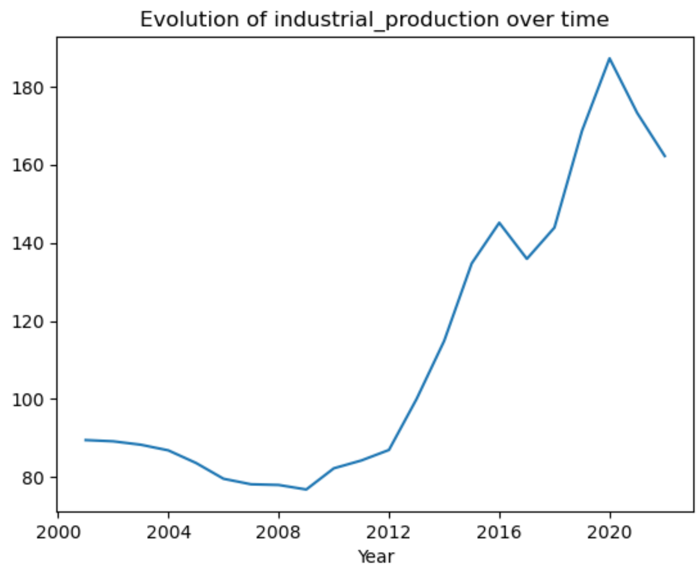
GDP

Figure 4.2: Plot of GDP time series



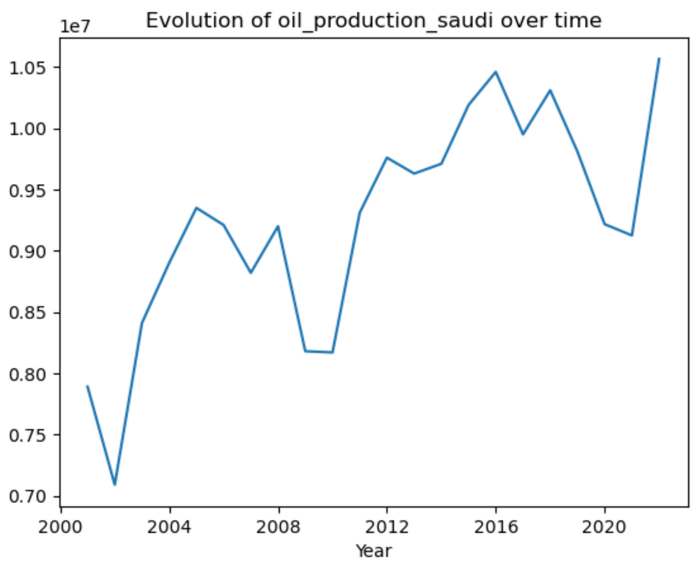
Industrial Production

Figure 4.3: Plot of Industrial Production time series



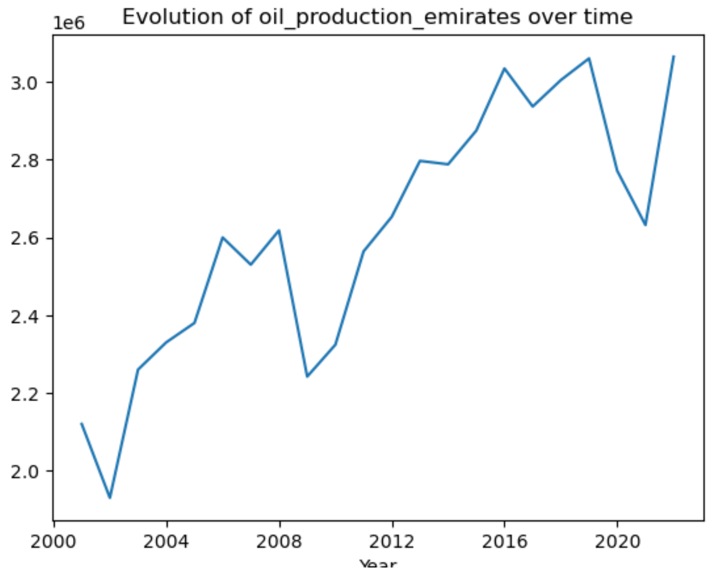
Saudi oil production

Figure 4.4: Plot of Saudi oil production time series



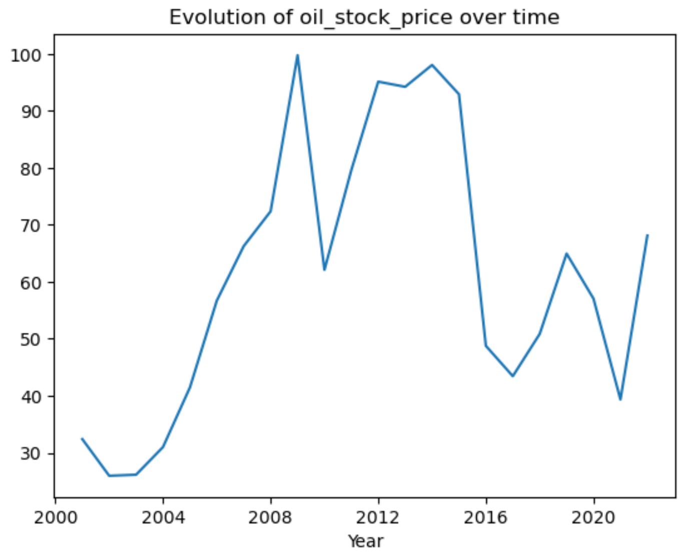
United Arab Emirates oil production

Figure 4.5: Plot of United Arab Emirates oil production time series



Oil Stock Price

Figure 4.6: Plot of oil stock price time series



4.1 Make our time series stationary

4.1.1 Take the fist Difference

To make my time series stationary, we computed the percentage change of each time series as follow:

$$\text{Percentage Change} = \frac{\text{Value}_\text{t} - \text{Value}_\text{t-1}}{\text{Value}_\text{t-1}}$$

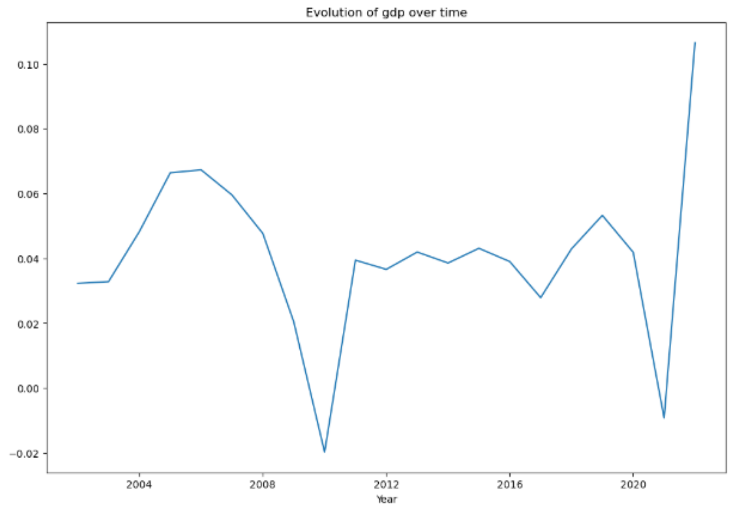
Figure 4.7: Time series after taking %Change

	gdp	industrial_production	oil_production_saudi	oil_production_emirates	oil_stock_price
2002-01-01	0.032287	-0.003494	-0.101394	-0.089623	-0.198597
2003-01-01	0.032809	-0.009774	0.186178	0.170984	0.007320
2004-01-01	0.048251	-0.016431	0.058264	0.030973	0.185234
2005-01-01	0.066403	-0.037055	0.050562	0.021459	0.337954
2006-01-01	0.067282	-0.048429	-0.014973	0.092437	0.367392
2007-01-01	0.059542	-0.017772	-0.042345	-0.026923	0.168418
2008-01-01	0.047674	-0.002248	0.043084	0.034783	0.092220
2009-01-01	0.020425	-0.014838	-0.110870	-0.143621	0.378459
2010-01-01	-0.019756	0.070607	-0.001222	0.036574	-0.377513
2011-01-01	0.039432	0.024400	0.139535	0.103270	0.282085
2012-01-01	0.036598	0.031938	0.048335	0.034711	0.194756
2013-01-01	0.041939	0.150162	-0.013320	0.054278	-0.009472
2014-01-01	0.038558	0.148553	0.008307	-0.003120	0.040687
2015-01-01	0.043094	0.172863	0.049434	0.031104	-0.052424
2016-01-01	0.039009	0.077273	0.026497	0.055593	-0.475158
2017-01-01	0.027871	-0.063708	-0.048757	-0.032239	-0.109217
2018-01-01	0.042924	0.058934	0.036181	0.023156	0.170769
2019-01-01	0.053253	0.172213	-0.048497	0.018557	0.277154
2020-01-01	0.041870	0.110021	-0.060440	-0.094476	-0.121723
2021-01-01	-0.009221	-0.074660	-0.009991	-0.050337	-0.310256
2022-01-01	0.106509	-0.063626	0.157921	0.164572	0.731028

Let's check visually whether after difference our time series, they are stationary

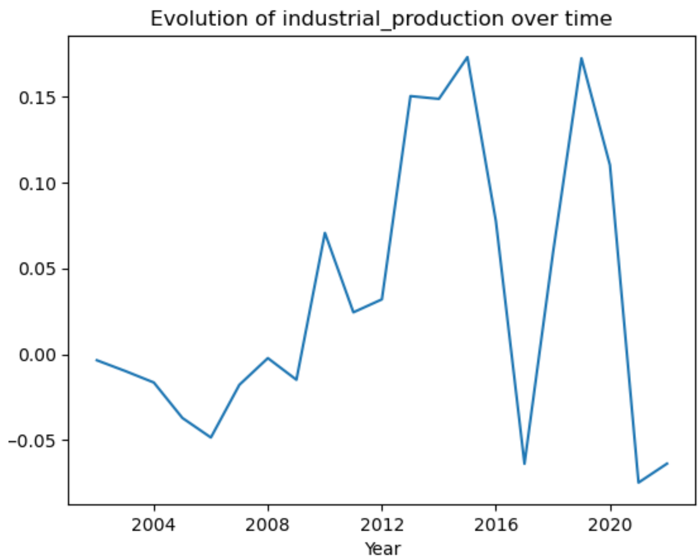
%Change(GDP)

Figure 4.8: Plot of %Change(GDP) time series



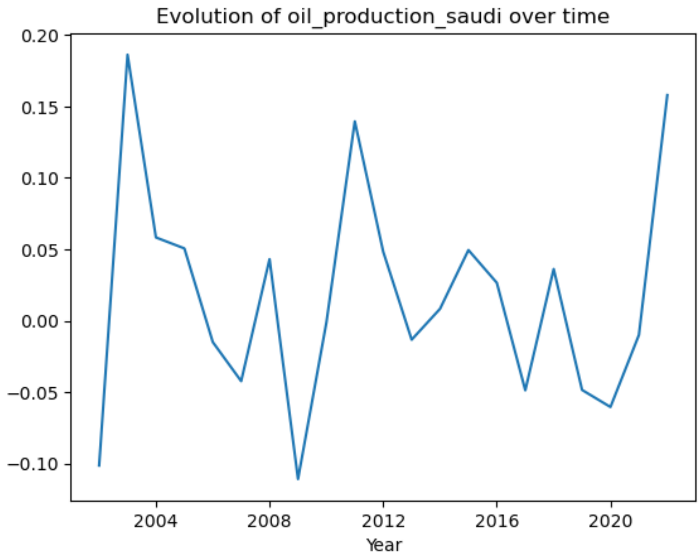
%Change(Industrial Production)

Figure 4.9: Plot of %Change(Industrial Production) time series



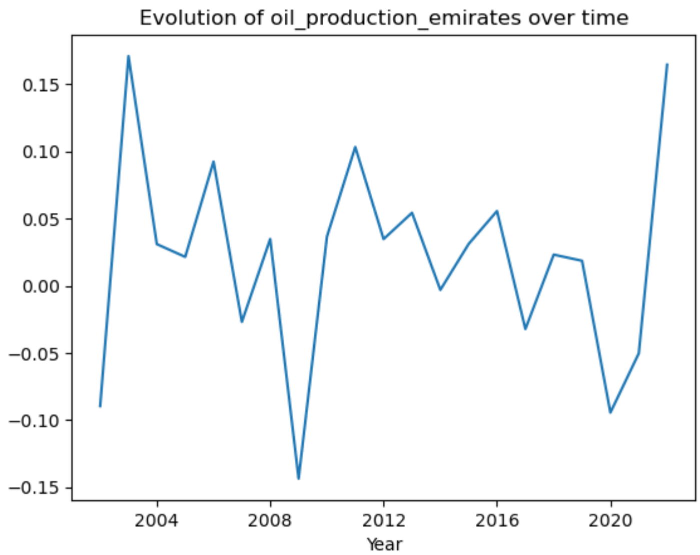
%Change(Saudi oil production)

Figure 4.10: Plot of %Change(Saudi oil production) time series



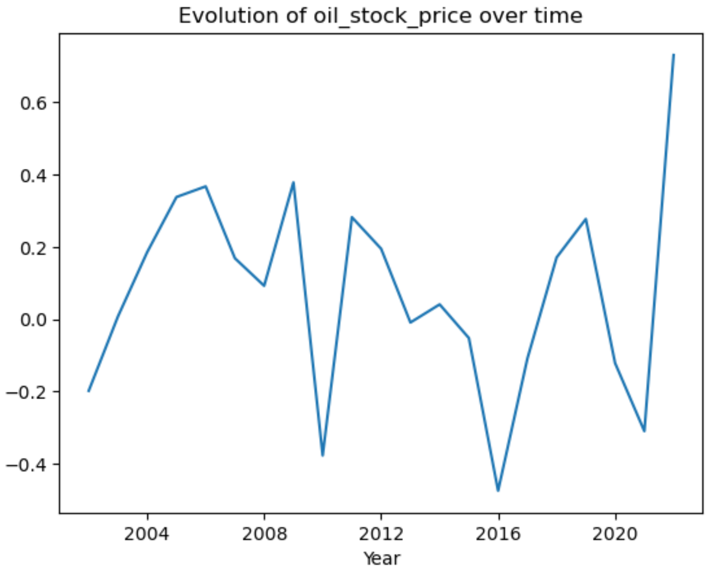
%Change(United Arab Emirates oil production)

Figure 4.11: Plot of %Change(United Arab Emirates oil production) time series



%Change(Oil Stock Price)

Figure 4.12: Plot of %Change(oil stock price) time series



4.1.2 Stationary Test Hypothesis

To the stationairy of each time series we apply the ADF test

Figure 4.13: ADF test results

```
ADF Test for gdp:
ADF Statistic: -3.696314318487574
p-value: 0.00416529623981911
Critical Values: {'1%': -3.8092091249999998, '5%': -3.0216450000000004, '10%':
-2.6507125}
Is Stationary: True
----------------------------------


ADF Test for industrial_production:
ADF Statistic: -2.9795257954645376
p-value: 0.03685117790995719
Critical Values: {'1%': -3.8326031418574136, '5%': -3.0312271701414204, '10%':
-2.655519584487535}
Is Stationary: True
----------------------------------


ADF Test for oil_production_saudi:
ADF Statistic: -4.178094178631945
p-value: 0.0007174737056428482
Critical Values: {'1%': -4.137829282407408, '5%': -3.1549724074074077, '10%': -
2.7144769444444443}
Is Stationary: True
----------------------------------


ADF Test for oil_production_emirates:
ADF Statistic: -3.777356210880336
p-value: 0.003147675434795737
Critical Values: {'1%': -3.8326031418574136, '5%': -3.0312271701414204, '10%':
-2.655519584487535}
Is Stationary: True
----------------------------------


ADF Test for oil_stock_price:
ADF Statistic: -3.9984780898925747
p-value: 0.0014185014202515727
Critical Values: {'1%': -3.8092091249999998, '5%': -3.0216450000000004, '10%':
-2.6507125}
Is Stationary: True
----------------------------------
```

We can see that if we take the return of each time series are stationary.

4.2 Applying simple linear regression

The regression is the following:

$\%Change(oil\_sock\_price) = \beta_0 + \beta_1 \times \%Change(gdp) + \beta_2 \times \%Change(industrial\_production) + \beta_3 \times \%Change(oil\_production\_saudi) + \beta_4 \times \%Change(oil\_production\_emirates) + \varepsilon$

Figure 4.14: Plot of %Change(oil stock price) time series

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        oil_stock_price   R-squared:                       0.588
Model:                            OLS   Adj. R-squared:                  0.485
Method:                 Least Squares   F-statistic:                     5.707
Date:                Sat, 25 Nov 2023   Prob (F-statistic):            0.00474
Time:                        21:49:16   Log-Likelihood:                 6.2974
No. Observations:                  21   AIC:                            -2.595
Df Residuals:                      16   BIC:                             2.628
Df Model:                           4
Covariance Type:            nonrobust
===============================================================================
============
                              coef    std err          t      P>|t|      [0.025
0.975]
-------------------------------------------------------------------------------
------------
const                      -0.2383      0.092     -2.591      0.020      -0.433
-0.043
gdp                         8.2095      1.984      4.137      0.001       4.003
12.416
industrial_production      -0.6011      0.586     -1.026      0.320      -1.843
0.641
oil_production_saudi        0.7317      1.235      0.592      0.562      -1.887
3.351
oil_production_emirates    -0.7138      1.274     -0.560      0.583      -3.415
1.988
==============================================================================
Omnibus:                        4.067   Durbin-Watson:                   1.528
Prob(Omnibus):                  0.131   Jarque-Bera (JB):                2.283
Skew:                          -0.338   Prob(JB):                        0.319
Kurtosis:                       4.467   Cond. No.                         46.6
==============================================================================
```

The R-squared value is 0.588, indicating that approximately 58.8

Coefficients

- **const (Intercept):** The coefficient is -0.2383, with a standard error of 0.092. The intercept represents the estimated value of the dependent variable when all independent variables are zero. In this context, it suggests that when all independent variables are zero, the estimated oil stock price is -0.2383.
- **GDP:** The coefficient is 8.2095, with a standard error of 1.984. Holding other

variables constant, a one-unit increase in GDP is associated with an estimated increase of 8.2095 units in oil stock price. The coefficient for GDP has a p-value of 0.001. With a low p-value, GDP is highly statistically significant. This implies that changes in GDP are associated with a significant impact on oil stock price. Specifically, a one-unit increase in GDP is associated with a statistically significant increase in oil stock price.

- **Industrial production:** The coefficient is -0.6011, with a standard error of 0.586. Holding other variables constant, a one-unit increase in industrial production is associated with an estimated decrease of 0.6011 units in oil stock price. The coefficient for industrial production has a p-value of 0.320. The high p-value indicates that industrial production is not statistically significant at the conventional significance level (0.05). In other words, the data do not provide enough evidence to reject the null hypothesis that the coefficient for industrial production is zero. Changes in industrial production may not be reliably associated with changes in oil stock price in this model.

- **Saudi oil production :** The coefficient is 0.7317, with a standard error of 1.235. Holding other variables constant, a one-unit increase in oil production saudi is associated with an estimated increase of 0.7317 units in oil stock price. The coefficient for oil production saudi has a p-value of 0.562. Similar to industrial production, the high p-value suggests that oil production saudi is not statistically significant at the conventional significance level (i.e. 95%). The data do not provide enough evidence to conclude that changes in oil production saudi are associated with changes in oil stock price in a statistically significant way.

- **United Arab Emirates oil production:** The coefficient is -0.7138, with a standard error of 1.274. Holding other variables constant, a one-unit increase in oil production emirates is associated with an estimated decrease of 0.7138 units in oil stock price. The coefficient for oil production emirates has a p-value of 0.583. Like the previous two variables, oil production emirates is not statistically significant based on the p-value. The data do not provide enough evidence to reject the null hypothesis that the coefficient for oil production emirates is zero in the population.

4.3 Applying ARIMAX model

Let's considering a single linear equation explaining the current value of a stationary financial series (returns for example) by its past values and the current/past values of other explanatory -exogenous series.

To do so, we apply an ARIMAX model to explain the past current/past values of $\%Changeg(oil\_sock\_price)$ by the the current/past values of $\%Changeg(industrial\_production), \%Changeg(oil\_production\_saudi), \%Changeg(oil\_production\_$

$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + X_{1,t}\beta_1 + X_{2,t}\beta_2 + \cdots + X_{k,t}\beta_k + \epsilon_t$

where:

- Y_t: Dependent variable (%Change(oil_stock_price))
- $\phi_1, \phi_2, \ldots, \phi_p$: Autoregressive coefficients
- ϵ_t: White noise error term
- $X_{1,t}, X_{2,t}, \ldots, X_{k,t}$: Exogenous variables at time t
- $\beta_1, \beta_2, \ldots, \beta_k$: Coefficients associated with exogenous variables
- $\theta_1, \theta_2, \ldots, \theta_q$: Moving average coefficients

4.3.1 Find best order for ARIMAX model

To do so we will try to find the best order of the model p and q according to the AIC information criterion.

The best orders are $p = 0$ and $q = 0$. The orders are quite strange because q and p are null. So it equals to do q simple linear regression.

Let's fitting the ARIMAX model with this order.

Figure 4.15: Results of ARIMAX model

SARIMAX Results

Dep. Variable:	oil_stock_price	No. Observations:	21
Model:	ARIMA(0, 1, 0)	Log Likelihood	1.724
Date:	Tue, 12 Dec 2023	AIC	6.553
Time:	11:24:14	BIC	11.531
Sample:	01-01-2002	HQIC	7.525
	- 01-01-2022		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
gdp	9.5520	3.175	3.008	0.003	3.329	15.775
industrial_production	-0.9364	0.756	-1.238	0.216	-2.418	0.546
oil_production_saudi	0.4462	1.004	0.444	0.657	-1.522	2.415
oil_production_emirates	-0.6784	0.874	-0.776	0.438	-2.392	1.035
sigma2	0.0493	0.020	2.521	0.012	0.011	0.088

Ljung-Box (L1) (Q):	1.06	Jarque-Bera (JB):	0.08
Prob(Q):	0.30	Prob(JB):	0.96
Heteroskedasticity (H):	1.74	Skew:	-0.15
Prob(H) (two-sided):	0.48	Kurtosis:	3.04

Coefficients Analysis

- **gdp:** The coefficient for gdp is 9.5520 with a standard error of 3.175. This indicates that a one-unit increase in the percent change of GDP is associated with an estimated 9.5520-unit increase in the percent change of the oil stock price, holding other variables constant.

- **industrial production:** The coefficient is -0.9364 with a standard error of 0.756. A one-unit increase in the percent change of industrial production is associated with an estimated decrease of 0.9364 units in the percent change of the oil stock price, holding other variables constant.

- **oil production saudi, oil production emirates:** The coefficients for these variables represent the estimated impact on the oil stock price for a one-unit increase in the respective exogenous variables.

- sigma2: Represents the variance of the residuals.

The coefficients for the exogenous variables (gdp, industrial production, oil production saudi, oil production emirates) provide estimates of their impact on the oil stock price, holding other factors constant. The Ljung-Box test indicates no significant autocorrelation in the residuals, suggesting that the model captures the temporal dependence adequately. The Jarque-Bera test and Heteroskedasticity test suggest that the residuals are approximately normally distributed and do not exhibit significant heteroskedasticity.

Overall, the positive coefficient for GDP suggests that an increase in GDP is associated with an increase in the oil stock price.

The negative coefficient for industrial production suggests an inverse relationship with the oil stock price.

The coefficients for the oil production variables indicate their impact on the oil stock price.

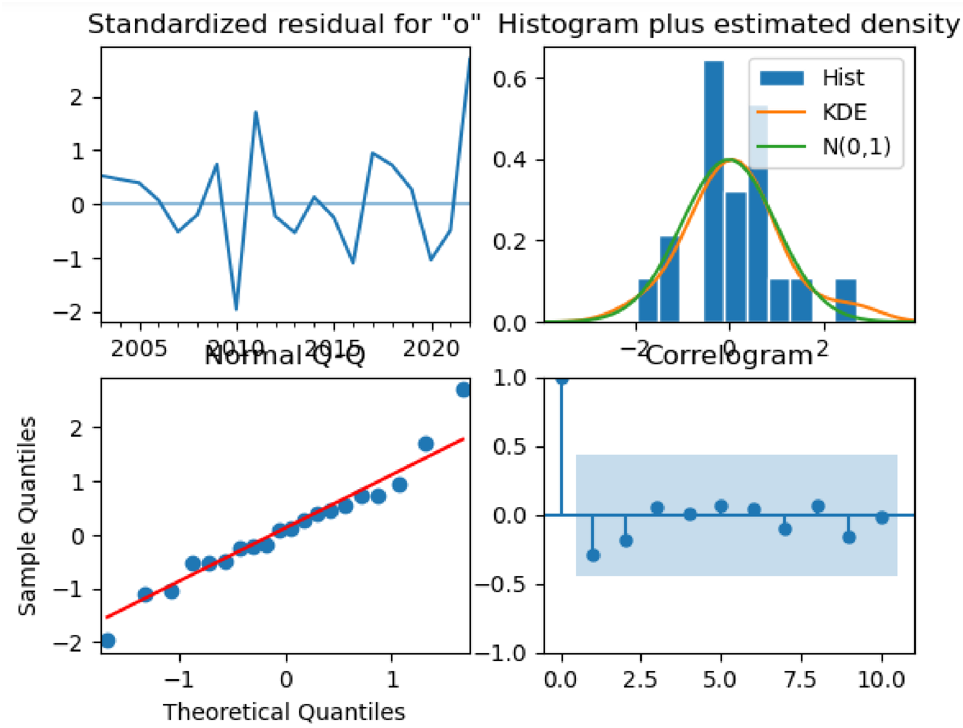
Check on residuals

Figure 4.16: Residuals of ARIMAX

```
###########################################
Check residuals of ARIMA model
###########################################

Mean of Residuals: 0.03481085852748228
We fail to reject H0: The residuals are not independently distributed.
No significant serial correlation in residuals.
```

Figure 4.17: Residuals of ARIMAX



	lb_stat	lb_pvalue
1	2.185158	0.139347
2	3.128572	0.209237
3	3.193281	0.362774

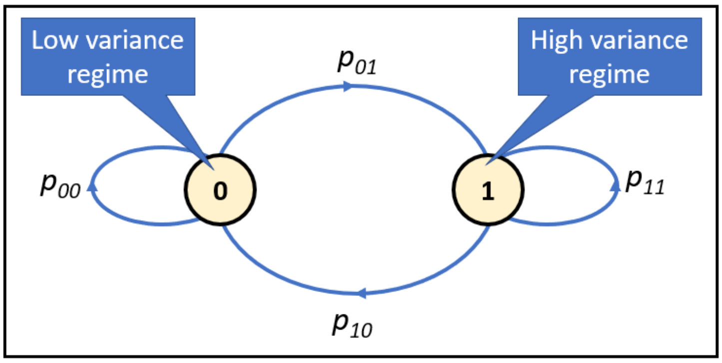
The Q-Q plot of each series looks like a heavy tailed q-q plot. This means that, compared with the normal distribution, there is much more data in the extremities than in the center of the distribution.

Furthermore, according to the density histogram for each series, we can see that green line showing a normal distribution is close from the orange line.

4.4 Apply Markov Switching model

We will use a 2-state Markov Switching Dynamic Regression model to try to model these 'switches' in variance between high and low variance regimes:

Figure 4.18: Markov Switching Model Explained



$$P = \begin{pmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{pmatrix}$$

The transition probabilities are typically represented by a transition matrix P where $P_{i,j}$ is the probability of transitioning from state i to state j

where:

- $\hat{\mu_{tj}} = \hat{\beta_{0j}} + \hat{\beta_{1j}} x_{1t}$
- $$\hat{\beta_S} = \begin{pmatrix} \beta_{00} & \beta_{01} \\ \beta_{10} & \beta_{11} \end{pmatrix}$$
- $\epsilon_{tj} \approx N(0, \sigma_j^2)$

Note that we have also introduced a state-specific variance. We are saying that the residual errors of the model are normally distributed around a zero mean and a variance that switches between two values depending on which state the underlying Markov process is in.

4.4.1 Results Analysis

Figure 4.19: Markov Switching Model results

Markov Switching Model Results

Dep. Variable:	oil_stock_price	**No. Observations:**	21
Model:	MarkovRegression	**Log Likelihood**	17.493
Date:	Sat, 25 Nov 2023	**AIC**	-6.986
Time:	14:16:03	**BIC**	7.637
Sample:	01-01-2002	**HQIC**	-3.813
	- 01-01-2022		
Covariance Type:	approx		

Regime 0 parameters

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	-0.2573	0.023	-11.403	0.000	-0.302	-0.213
x1	6.8127	0.434	15.680	0.000	5.961	7.664
x2	-0.6145	0.129	-4.782	0.000	-0.866	-0.363
x3	0.2329	0.272	0.856	0.392	-0.301	0.766
x4	1.2038	0.245	4.908	0.000	0.723	1.685
sigma2	0.0010	0.000	2.218	0.027	0.000	0.002

Regime 1 parameters

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	-0.3273	0.150	-2.182	0.029	-0.621	-0.033
x1	10.1232	3.339	3.032	0.002	3.579	16.667
x2	0.7474	1.041	0.718	0.473	-1.293	2.788
x3	4.7495	2.612	1.819	0.069	-0.369	9.868
x4	-5.8749	2.768	-2.123	0.034	-11.299	-0.450
sigma2	0.0340	0.014	2.352	0.019	0.006	0.062

Regime transition parameters

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
p[0->0]	1.202e-05	nan	nan	nan	nan	nan
p[1->0]	0.8138	0.122	6.680	0.000	0.575	1.053

A Markov Switching Model has multiple regimes. In our case we have two transitions regime for low volatility and high volatility that can explain different behaviors in the data. The coefficients in each regime show the impact of each variable on the dependent variable (oil stock price) during that regime. Significant coefficients (where $P<0.05$) indicate a strong relationship between the predictor and the dependent variable within that regime.

These regime parameters show the probability of switching from one regime to another. For instance, p[0->1] is the probability of switching from regime 0 to regime 1. These transition probabilities are important for understanding the likelihood of changes between states.

Here please note that:

- **Dep. Variable:** %Change(oil_stock_price)
- **x1:** %Change(GDP)
- **x2:** %Change(Industrial_production)
- **x3:** %Change(Saudi_oil_production)
- **x4:** %Change(United_Arab Emirates_oil_production)

By looking at the fitted coefficients' values in the results, we can write the regime-specific model equations as follows:

For regime=0 (low-variance regime) $oil\_sock\_price = -0.2573 + 6.8127 \times gdp - 0.6145 \times industrial\_production + 0.2329 \times oil\_production\_saudi + 1.2038 \times oil\_production\_emirates + \epsilon_t$

where:

$\epsilon_t = 0.0010$

For regime=1 (high-variance regime) $oil\_sock\_price = -0.3273 + 10.1232 \times gdp + 10.1232 \times industrial\_production + 0.7474 \times oil\_production\_saudi - 5.8749 \times oil\_production\_emirates + \epsilon_t$

where $\epsilon_t = 0.0340$

- In our case Regime 0 is interpreted as a "low volatility" or "stable" market period, the coefficients in this regime would explain how the exogenous variables impact oil stock prices when the market is stable.

- Regime 1 might represent a "high volatility" or "unstable" market period, where the coefficients show the relationship during these more turbulent times.

Regime Transition Parameters:

- p[0->0]: 1.202e-05 (Transition probability from regime 0 to regime 0)
- p[1->0]: 0.8138 (Transition probability from regime 1 to regime 0)

We can deduce :

- p[1->1]: 1-p[0->0] = 1 - 1.202e-05 = 0.99998798
- p[0->1]: 1-p[1->0] = 1 - 0.8138 = 0.186

Hence,

$$P = \begin{pmatrix} 1.202^{-5} & 0.186 \\ 0.8138 & 0.99998798 \end{pmatrix}$$

p[0->0] (Transition probability from Regime 0 to Regime 0):

- Value: 1.202e-05
- This is the probability of staying in Regime 0. The extremely small value indicates a very low likelihood of remaining in the same regime.
- The very low probability of staying in Regime 0 (p[0->0]) indicates that the system is unlikely to remain in the same state if it is currently in a more stable economic condition (Regime 0).

p[1->0] (Transition probability from Regime 1 to Regime 0):

- Value: 0.8138
- This is the probability of transitioning from Regime 1 to Regime 0. The value of 0.8138 suggests a high likelihood of moving from the regime associated with different economic conditions (Regime 1) to a more stable state (Regime 0).

p[1->1] (Transition probability from Regime 1 to Regime 1):

- Value: 1 - p[0->0] = 1 - 1.202e-05 = 0.99998798
- This is the probability of staying in Regime 1. The very close-to-1 value (0.99998798) indicates a very high likelihood of remaining in the regime characterized by specific economic conditions (Regime 1).

p[0->1] (Transition probability from Regime 0 to Regime 1):

- Value: 1 - p[1->0] = 1 - 0.8138 = 0.186
- This is the probability of transitioning from Regime 0 to Regime 1. The value of 0.186 suggests a moderate likelihood of moving from a more stable economic state (Regime 0) to a regime associated with different economic conditions (Regime 1).

Overall, the transition probability from "Regime 0" to "Regime 0" is very close to zero (1.202e-05), suggesting a low probability of staying in the same regime. The transition probability from "Regime 1" to "Regime 0" is 0.8138, indicating a relatively high probability of transitioning from "Regime 1" to "Regime 0.". The statistically significant result (low p-value) indicates that this transition probability is different from zero. Indeed, the high transition probability from Regime 1 to Regime 0 suggests that after a period of economic volatility (Regime 1), there's a high likelihood of returning to a more stable state (Regime 0). This could be interpreted as a cyclical behavior in the crude oil market.

More precisely, we can see that when stock oil price are in a low variance regime, they tend to switch to a high variance regime less than 99% of the time, while if the stock oil price is in a high variance regime, it tends to switch back to a low variance regime with roughly 81% probability.

Regime-Specific Coefficients:

There are two regimes, each with its set of coefficients for the constant and predictors. The regimes are denoted as "Regime 0" and "Regime 1." For example, in "Regime 0," the coefficient for the first predictor (x1) is 6.8127, while in "Regime 1," the coefficient for the same predictor is 10.1232.

In more details:

> **Regime 0:**The constant term (-0.2573) represents the baseline level of crude oil stock prices in Regime 0. Positive coefficient for US GDP Growth (x1 = 6.8127) suggests that in Regime 0, an increase in US GDP Growth is associated with an increase in crude oil stock prices. Negative coefficient for Industrial Production (x2 = -0.6145) implies that higher industrial production in mining, quarrying, and oil and gas extraction is associated with lower crude oil stock prices in Regime 0. The other exogenous variables (Crude Oil Production Saudi Arabia, Crude Oil Production UAE) also have specific effects on crude oil stock prices in this regime.

Regime 1:

- The constant term (-0.3273) represents the baseline level of crude oil stock prices in Regime 1.
- A substantial positive coefficient for US GDP Growth (x1 = 10.1232) indicates a stronger positive relationship between US GDP Growth and crude oil stock prices in Regime 1 compared to Regime 0.
- Industrial Production (x2), Crude Oil Production Saudi Arabia (x3), and Crude Oil Production UAE (x4) also have specific effects on crude oil stock prices in Regime 1.
- The negative coefficient for Crude Oil Production UAE (x4) suggests a negative impact on crude oil stock prices in this regime.

Variance of the Error Term (σ_2):

The variance of the error term differs between regimes, being 0.0010 in "Regime 0" and 0.0340 in "Regime 1."

Significance of Parameters:

The statistical significance of each coefficient and transition parameter is indicated by the z-test and associated p-values. Low p-values (typically < 0.05) suggest significance.

- **Regime 0**
  - Regime 0 appears to have significant coefficients for const, gdp, industrial production, and oil production emirates, suggesting that these variables are important predictors
  - The coefficient for oil production saudi is not statistically significant in Regime 0 (p-value > 0.05), indicating that it may not be a significant predictor in this particular regime.
  - The variance of the error term (sigma2) is statistically significant, indicating that the estimated volatility of the residuals in Regime 0 is different from zero.
- **Regime 1**
  - In Regime 1, the intercept, coefficients for gdp, and oil production saudi are statistically significant, suggesting that these variables are important predictors of the dependent variable in this regime.
  - The coefficients for industrial production and oil production emirates are not statistically significant in Regime 1 (p-value > 0.05), indicating that they may not be significant predictors in this particular regime.
  - The variance of the error term (sigma2) is statistically significant, suggesting that the estimated volatility of the residuals in Regime 1 is different from zero.

Regime transition paramater interpretation:

The two regimes capture different states or patterns in the data. The transition parameters provide insights into the likelihood of transitioning between these states.

In our case, the statistically significant coefficient for p[1->0] implies that there is a substantial probability of transitioning from Regime 1 to Regime 0. This has economic implications, suggesting that there are periods (Regime 1) with distinct characteristics or conditions that are likely to be followed by a shift to a different set of conditions (Regime

0). This information can be valuable for decision-makers, investors, and policymakers who want to understand and anticipate changes in the economic environment affecting crude oil stock prices. The statistical significance provides confidence in the reliability of this transition probability estimate.

Discussion

In summary, the Markov Switching Model identifies two regimes with distinct economic characteristics, and the transition probability provides insights into the likelihood of moving between these regimes over time. This can be valuable for understanding and managing risks in the crude oil market.

This type of model is useful for capturing and modeling changes in the behavior of the variable "oil stock price" over time, allowing for different regimes with distinct sets of coefficients. Interpretation should involve examining the regime-specific coefficients, transition probabilities, and variance components to understand the dynamics of the system. Additionally, assessing the goodness of fit and conducting diagnostic checks is important for validating the model's

4.5 Transition Probabilities overt the time

Figure 4.20: Transition Probabilities over the time

	0	1
2002-01-01	9.728765e-01	0.027123
2003-01-01	1.032378e-11	1.000000
2004-01-01	9.722675e-01	0.027732
2005-01-01	3.569336e-06	0.999996
2006-01-01	9.999735e-01	0.000026
2007-01-01	8.398264e-06	0.999992
2008-01-01	9.897269e-01	0.010273
2009-01-01	6.225666e-107	1.000000
2010-01-01	9.981182e-01	0.001882
2011-01-01	7.077952e-05	0.999929
2012-01-01	3.842172e-07	1.000000
2013-01-01	9.949692e-01	0.005031
2014-01-01	5.826603e-07	0.999999
2015-01-01	9.978735e-01	0.002126
2016-01-01	2.271569e-59	1.000000
2017-01-01	9.883835e-01	0.011616
2018-01-01	1.406115e-04	0.999859
2019-01-01	4.704781e-17	1.000000
2020-01-01	9.994963e-01	0.000504
2021-01-01	3.278222e-04	0.999672
2022-01-01	9.827332e-01	0.017267

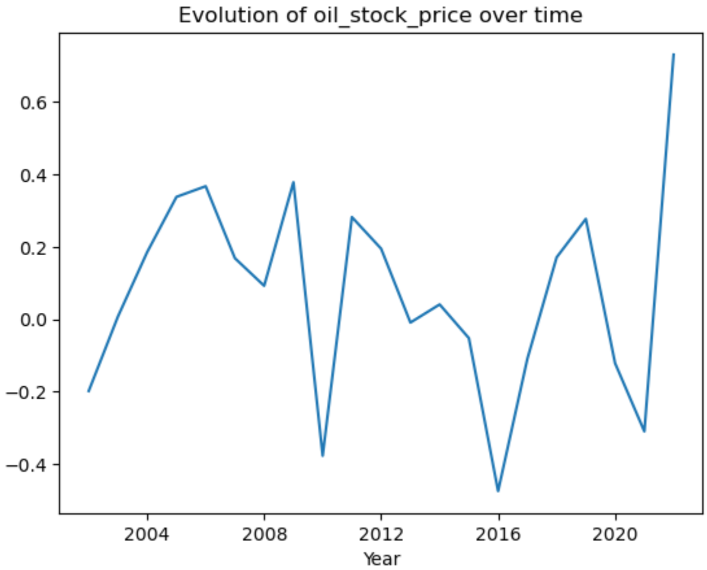
- The probability of transitioning from Regime 0 to Regime 1 varies across time.
- In the early years (e.g., 2002, 2004, 2006), the probabilities are relatively low for Regime 1, suggesting a high likelihood of staying in Regime 0.
- In 2008, the probability of from Regime 0 increases a lot from 8.32e-6 to 9.89e-1, indicating a moderate likelihood of entering a different state.
- The transition probability sharply drops in 2009 for Remige 0 (6.225e-107), suggesting a high likelihood of staying in Regime 1 (unstable economy conditions, with high volatility) during the financial crisis.
- Subsequent years (2010, 2013, 2015, 2017, 2020) show varying probabilities, with a trend of higher probabilities in recent years (e.g., 2017, 2020), indicating an increased likelihood of transitioning to Regime 0 during those periods.

- In some years (e.g., 2002, 2004, 2006), the probability is close to zero for both regime.
- In other years (e.g., 2005, 2007, 2010, 2013, 2015), the probability of transitioning to Regime 1 is higher, suggesting a greater likelihood of moving to a more unstable state.
- Notably, in 2016, the probability is extremely low (2.271569e-59) for Regime 0, indicating a very high likelihood of staying in Regime 1 during that year.

4.5.1 Plot Markov Switching Model Results

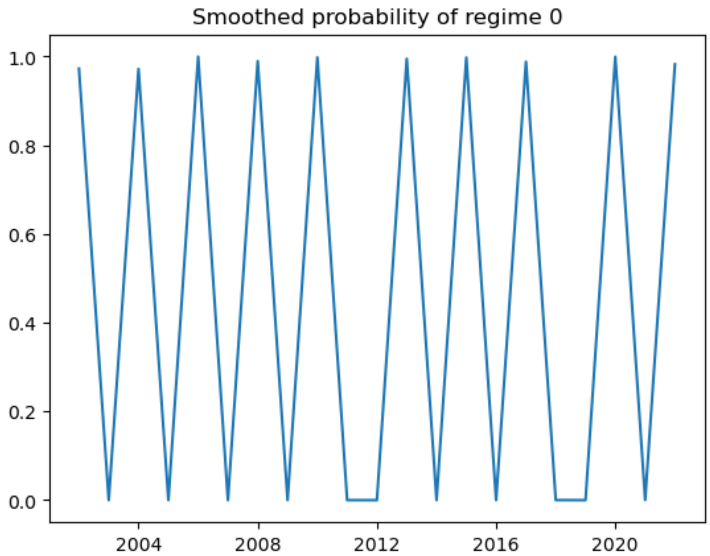
%Change(Oil stock price)

Figure 4.21: Evolution of %Change(Oil stock price) over the time



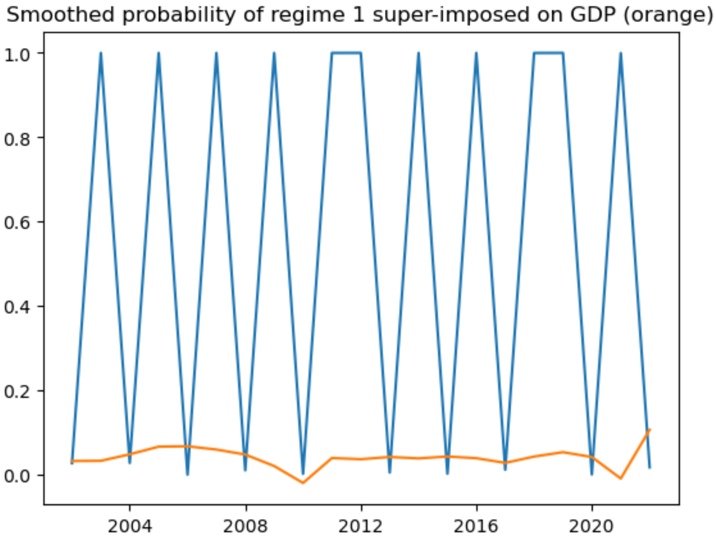
Smoothed probability of regime 0

Figure 4.22: Smoothed probability of regime 0



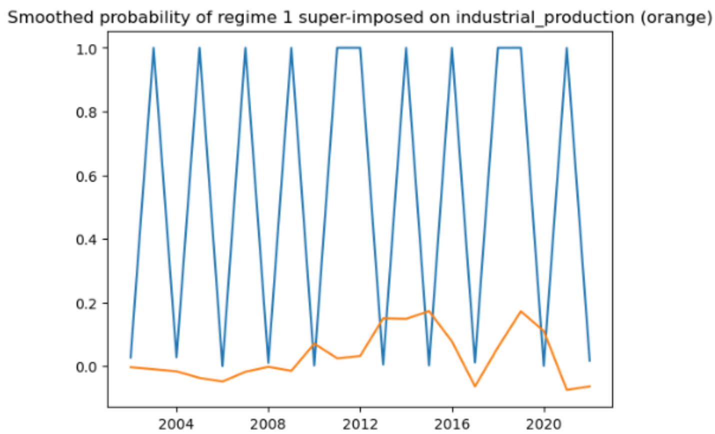
Smoothed probability of regime 1 super imposed on GDP

Figure 4.23: Smoothed probability of regime 1 super imposed on GDP (orange)



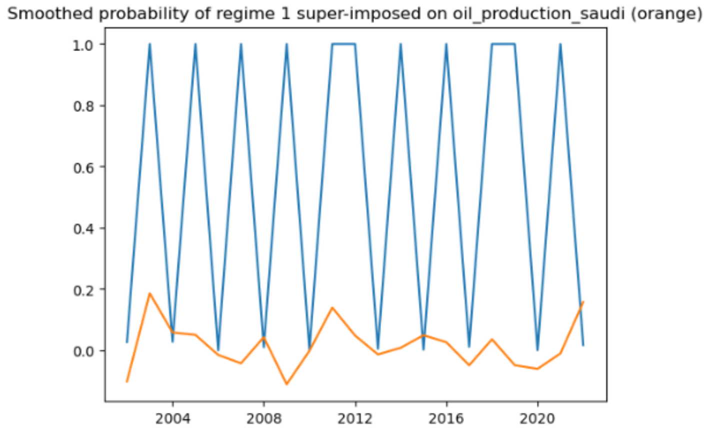
Smoothed probability of regime 1 super imposed on Industrial Production

Figure 4.24: Smoothed probability of regime 1 super imported on Industrial Production (orange)



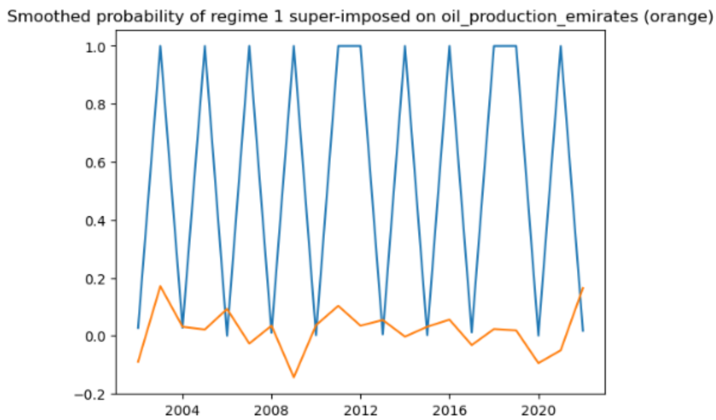
Smoothed probability of regime 1 super imposed on Saudi oil production

Figure 4.25: Smoothed probability of regime 1 super imposed on Saudi oil production (orange)



Smoothed probability of regime 1 super imposed on United Arab Emirates

Figure 4.26: Smoothed probability of regime 1 super imposed on United Arab Emirates (orange)



Regime Dynamics:

- The varying transition probabilities indicate shifts in the economic conditions or market states over time.
- Periods with low transition probabilities (staying in the same regime) might represent relatively stable economic conditions or market environments.
- Years with higher transition probabilities, especially from Regime 0 to Regime 1, might coincide with periods of economic uncertainty, financial crises, or other events impacting the crude oil market.

4.6 Comparison between ARIMAX and Markov-Switching model:

Figure 4.27: Actual VS predicted Values with Regression Line (ARIMAX)

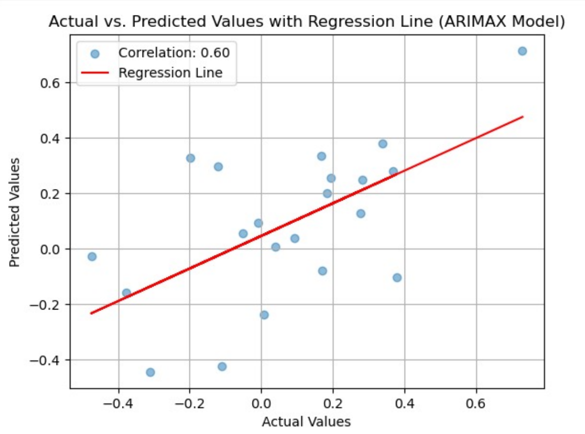
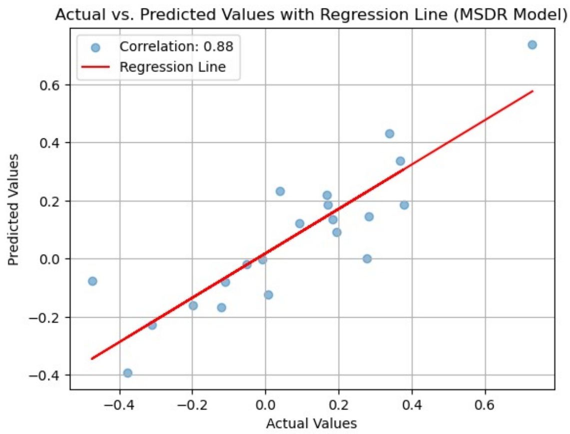


Figure 4.28: Actual VS predicted Values with Regression Line (Markov-Switching Model



In the provided scatter plots, we observe that the Markov Switching Dynamic Regression (MSDR) model has a markedly higher correlation coefficient of 0.88 compared to the ARIMAX model's coefficient of 0.60. This substantial difference in correlation suggests that the MSDR model has a stronger linear relationship between the predicted and actual values.

The data points in the MSDR model's scatter plot are more tightly clustered around the regression line, which implies a higher data concentration and, consequently, a more precise prediction. In contrast, the ARIMAX model shows a looser fit, with data points more spread out from the regression line, indicating a larger variance and potentially less accurate predictions.

The difference between the two models can be particularly impactful in financial analysis, where even small improvements in predictive accuracy can lead to significantly different investment decisions and risk assessments. The MSDR model's tighter clustering and higher correlation suggest it may be better at capturing the volatility and regime shifts often observed in economic and financial data. This ability to detect and adapt to different states in the data can be especially valuable in forecasting and interpreting complex market dynamics.

Furthermore, the higher correlation in the MSDR model implies that it might be better suited for tasks that require understanding the impact of various economic factors under different market conditions. For instance, during a financial crisis, the MSDR model may better account for the abrupt changes in market behavior, whereas the ARIMAX model might underestimate or fail to quickly adjust to such shifts.

Overall, when comparing the two models, it is evident that the MSDR model not only fits the current dataset better but also offers a more dynamic and responsive approach to modeling financial time series data.

4.7 Conclusion

- Monitoring the transition probabilities can be useful for investors and policymakers to anticipate potential shifts in the crude oil market.
- High transition probabilities could signal periods of increased market volatility or structural changes in the underlying economic factors affecting crude oil prices.
- Understanding the timing and likelihood of regime transitions can inform decision-making processes, such as portfolio adjustments or risk management strategies.

Please refer to the following link to see the full code of this assignment: Empirical Analysis 5

5 Empirical Application 6 : Smooth transition model

In the realm of financial analysis, accurately modeling and predicting economic indicators is important. Traditional linear regression models have long been a staple in this field, offering a straightforward approach to understanding relationships between variables. However, the complexity and dynamic nature of financial markets often exceed the capabilities of linear models. This gap necessitates the shift towards non-linear models, specifically Smooth Transition Regression (STR) models.

The primary objective of this study is to enhance an existing linear model by incorporating non-linearity through the estimation of an STR model. By transitioning from a linear to a non-linear framework, we aim to capture the intricate patterns and relationships often present in economic data, such as Gross Domestic Product (GDP) and industrial production figures, that are missed by traditional linear models.

Adopting non-linear models like STR in financial modeling is crucial. Linear models often assume a constant relationship between variables over time, an assumption that falls short in the volatile and shifting landscape of financial markets. STR models, with their ability to model different behaviors in different regimes or periods, provide a more accurate and realistic representation of financial dynamics.

In the next sections, we will explore the theoretical underpinnings of STR models, justify the choice of the transition variable in the context of our financial data, and detail the process of model estimation and validation. Our analysis will culminate in a comparison of the performance of the STR model against the original linear model, highlighting the improvements and insights gained through this non-linear approach.

5.1 Theoretical Background

5.1.1 Markov Switching Dynamic Regression Models

Basics of Markov Switching Dynamic Regression

Markov Switching Dynamic Regression (MSDR) are an extension of linear regression models that allow for regime changes in the data. These models are formulated as:

$$Y_t = \beta_{0,S_t} + \beta_{1,S_t} X_t + \epsilon_t \tag{5.1}$$

where Y_t represents the dependent variable at time t, X_t is the independent variable at time t, β_{0,S_t} and β_{1,S_t} are the regime-dependent coefficients, S_t is the state or regime at time t, and ϵ_t is the error term. MSDR models incorporate Markov chains to model the transition between different states or regimes, capturing dynamic changes in the relationship between variables over time.

Common Use Cases in Finance

In finance, MSDR models are particularly useful for capturing the dynamics of financial markets that exhibit regime changes. They are applied in analyzing stock market behaviors during different market conditions, estimating the impact of macroeconomic changes on securities, and understanding structural shifts in economic relationships. Their ability to switch between different states based on the data makes them well-suited for financial series that demonstrate non-linear patterns, such as those observed during financial crises or periods of rapid economic change.

5.1.2 ARIMAX Models

Basics of ARIMAX

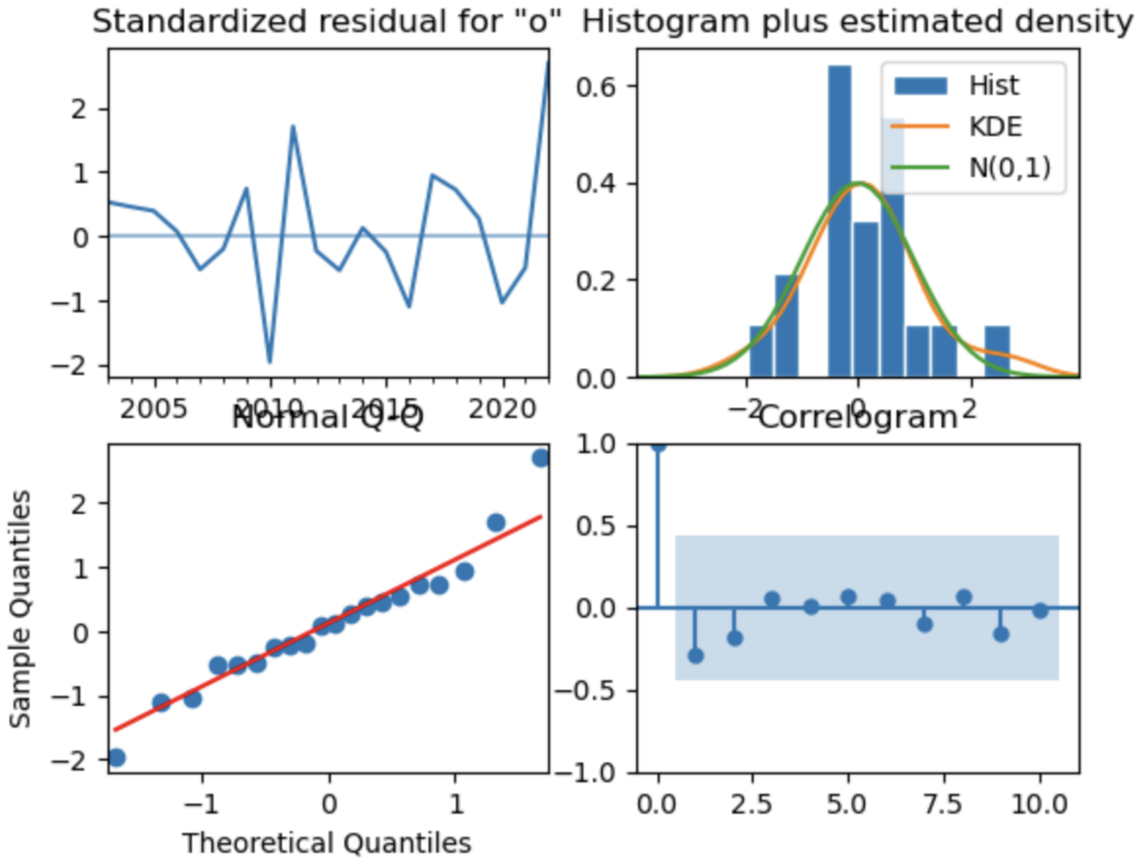
ARIMAX (AutoRegressive Integrated Moving Average with eXogenous variables) models extend the ARIMA (AutoRegressive Integrated Moving Average) framework by incorporating external or exogenous variables. The model can be expressed as:

$$Y_t = \beta_0 + \beta_1 X_t + AR(p) + MA(q) + I(d) + \epsilon_t \tag{5.2}$$

where Y_t is the dependent time-series variable (e.g., stock prices), X_t represents the exogenous variable(s) affecting Y_t, $AR(p)$ denotes the autoregressive part with p lags, $MA(q)$ is the moving average part with q lags, $I(d)$ is the differencing order to make the series stationary, β_0 and β_1 are coefficients for the exogenous variables, and ϵ_t is the error term at time t. The inclusion of X_t allows the model to capture the impact of external factors on the time series.

Common Use Cases in Finance

In finance, ARIMAX models are commonly used for incorporating external information or indicators that are believed to influence the financial time series being modeled. For example, an ARIMAX model might include interest rates, inflation figures, or economic indicators as exogenous variables to predict stock market indices or currency exchange rates. The model can also be used for forecasting economic indicators themselves, like GDP, by including related financial variables as predictors. The preprocessing of variables for an ARIMAX model include stationarity checks, differencing if needed, and inclusion of relevant exogenous variables in the analysis (see empirical application 5).



5.1.3 Introduction to STR Models

Definition and Characteristics of STR Models

Economic theories often posit that economies exhibit different behaviors under various regimes, depending on specific variable values. The initial modeling approach for this was through discrete switching models, which assume a finite number of regimes. These models' core is the switching variable, either observable or unobservable.

Transitioning to a more nuanced view, a generalized form of these models is represented as:

$$y_t = x_t'\phi + (x_t'\theta) \cdot G(\gamma, c; s_t) + u_t \quad \text{for } t = 1, 2, ..., T \tag{5.3}$$

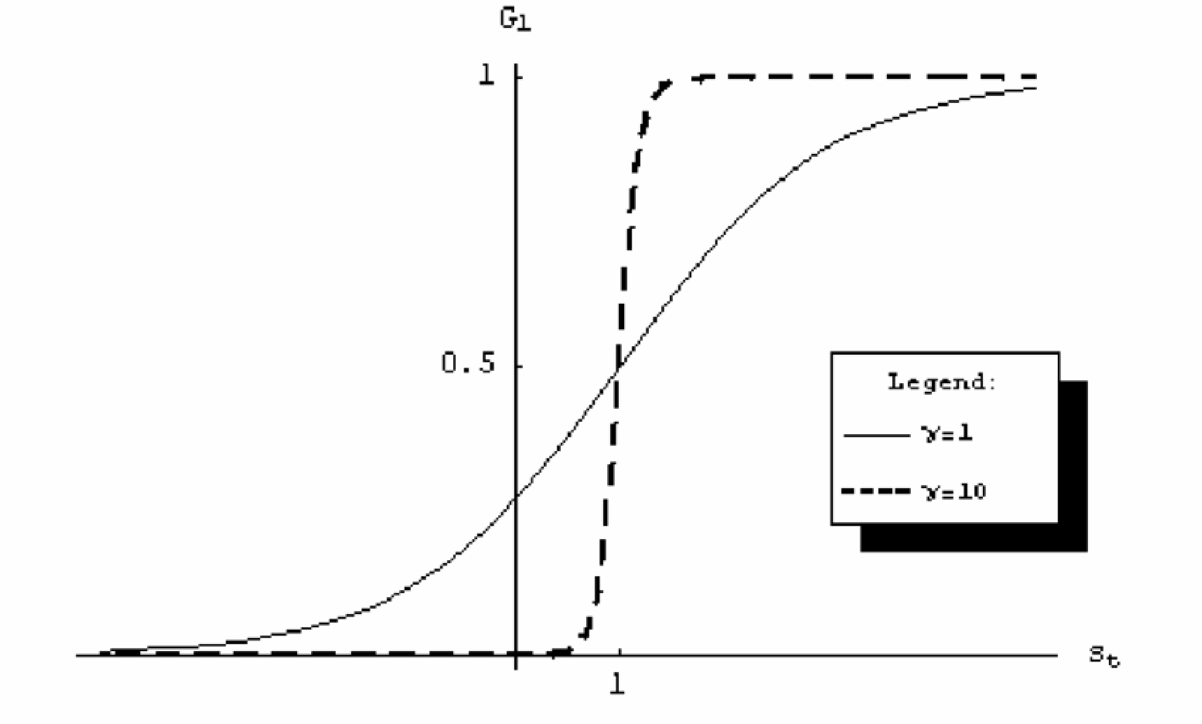
Here, ϕ and θ are parameter vectors, x_t is the vector of explanatory variables, and u_t represents a sequence of independent, identically distributed errors. The function G, typically bounded between 0 and 1, facilitates modeling not only the extreme states but also intermediate states. The parameters $\gamma > 0$ and c denote the speed and location of the transition. The transition variable s_t is often an explanatory variable or a time trend.

One of the popular functional forms of the transition function is the LSTR1 Model:

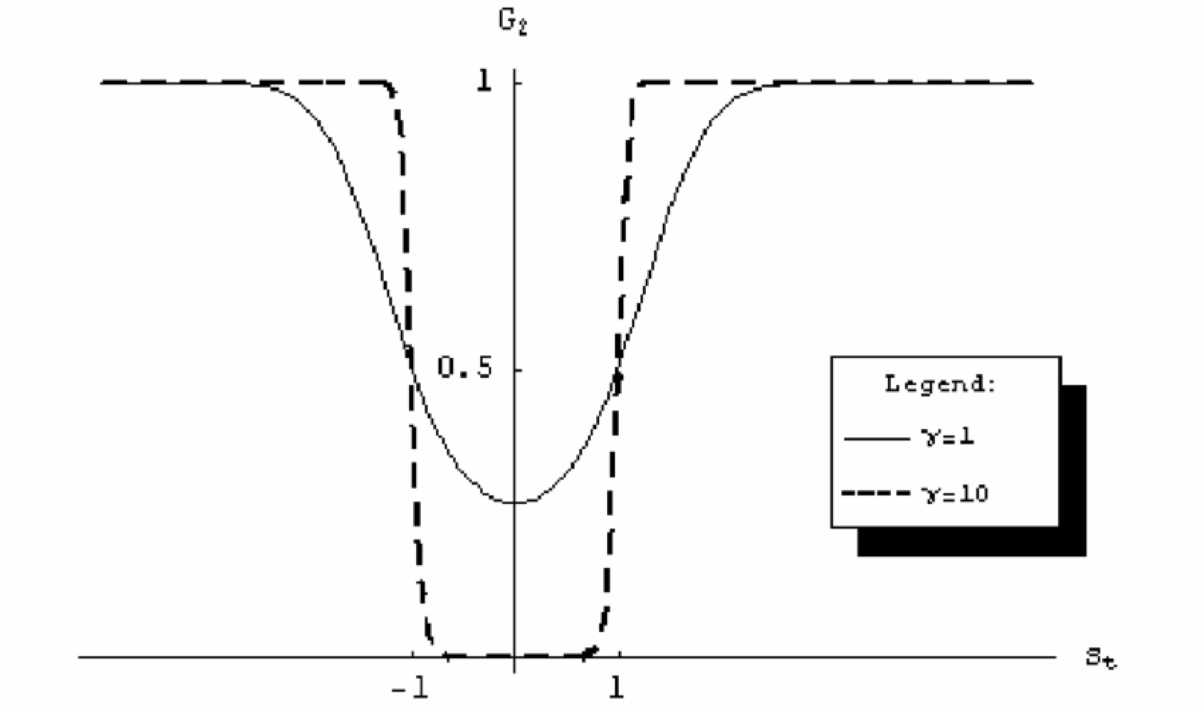
$$G(\gamma, c; s_t) = \frac{1}{1 + e^{-\gamma(s_t - c)}} \tag{5.4}$$

In this model, $G1$ increases monotonically with s_t, ranging between 0 and 1, and $G1(\gamma, c; c) = 0.5$, signifying c as the transition midpoint. The slope parameter γ influences the transition's rapidity, with higher values indicating faster changes.

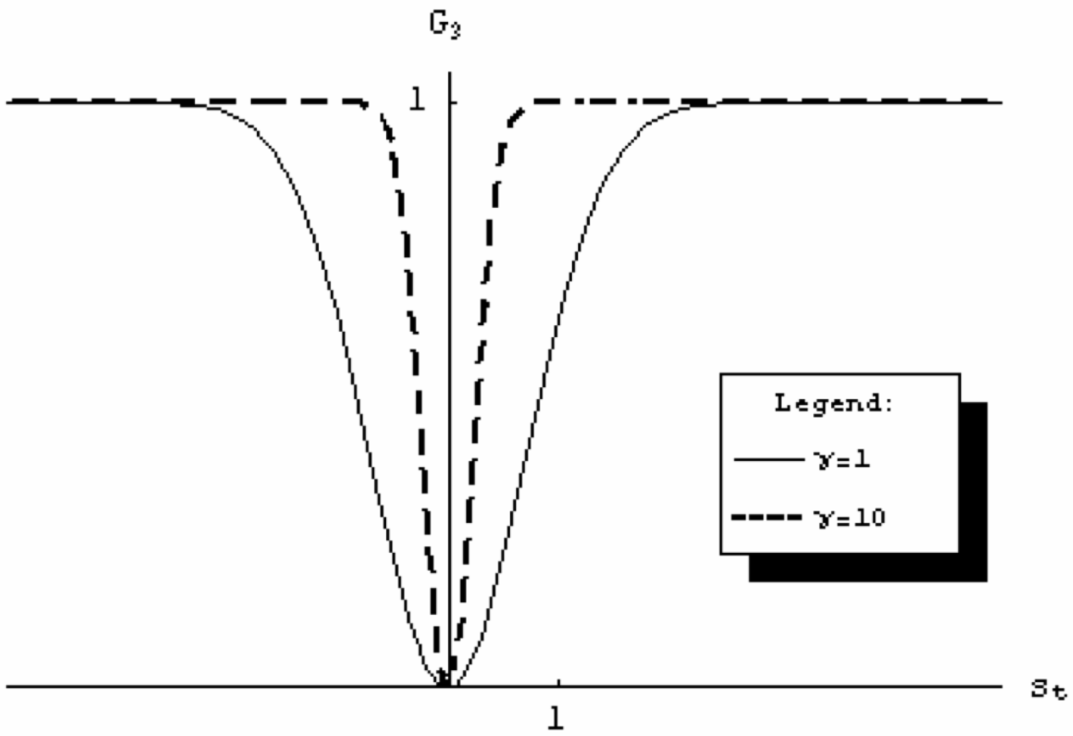
When the parameter γ approaches infinity in the definition of $G1$, model (1) evolves into a switching regression model, oscillating between the extreme regimes $y = x'\phi + u$ and $y = x'(\phi + \theta) + u$. For $\gamma = 0$, the function G remains constant at 0.5, thus reducing model (1) to a linear regression model.



- LSTR2 Model: $G(\gamma, c_1, c_2; s_t) = \frac{1}{1+e^{-\gamma(s_t-c_1)(s_t-c_2)}}$. This model is advantageous for situations requiring non-monotonous transitions, such as in cases of reswitching. The function $G2$ is symmetrical around $c_1 + c_2$ and never reaches 0, with its minimum value falling between 0 and 0.5.



- ESTR Model: $G3(\gamma, c; s_t) = 1 - e^{-\gamma(s_t-c)^2}$. This model is ideal when small absolute values of the transition variable lead to small values of the transition function. The function $G3$ is non-monotonous and symmetric about c.



Both the LSTR2 and ESTR models allow for reswitching, but they differ in the speed of this transition. The ESTR model shows a much faster transition from one state to another compared to the LSTR2 model, especially when the gap between c_1 and c_2 is large.

Advantages over Linear Models in Financial Contexts

Unlike discrete switching models such as those explored by Hansen (1999), Smooth Transition Regression (STR) models characterize the process of transition as a continuous one, which is influenced by a specific transition variable. This attribute of STR models is particularly beneficial in situations where the precise timing of regime shifts is ambiguous or when transitions into new regimes occur over brief periods. Consequently, STR models offer a nuanced understanding of the dynamic behavior of variables, even during periods of transition, providing valuable insights.

The ability of STR models to capture non-linearities and regime shifts renders them highly effective for analyzing a wide range of economic variables. Firstly, these models are adept at modeling institutional structural changes, making them especially pertinent for studying economies in transition that often undergo numerous structural shifts. Secondly, there is substantial evidence of asymmetrical dynamics in economic variables in established market economies, as indicated by the work of Johansen (2002) on asymmetric wage reactions in different sectors, and Legrenzi (2005), who examined the asymmetric adjustment of real exchange rates. Thirdly, STR methodology has gained prominence in the study of exchange rates and has recently found applications in examining Okun's Law and the Phillips curve. Lastly, the extension of STR methodology to Vector Autoregression (VAR) models and panel data opens up a broad spectrum of new possibilities, allowing for the modeling of multiple variables and the inclusion of heterogeneity in disaggregated data.

5.2 The STR Model

5.2.1 Methodology and choice of transition variable

The methodology we have adopted focuses on the analysis of economic time series covering the period from 2000 to 2022. The dataset includes key indicators such as GDP, industrial production, as well as oil production from Saudi Arabia and the Emirates, along with oil stock prices. These data, sourced from reliable sources like FRED and Yahoo Finance, have been meticulously imported, processed, and merged into a coherent DataFrame, highlighting a structured and unified approach. The data preprocessing includes essential steps such as annual resampling and first differencing of oil stock prices to ensure stationarity, a crucial step for time series analysis. Regarding the selection of the transition variable for the Smooth Transition Regression (STR) model, we use oil-related variables, given their significant economic impact and correlation with overall economic indicators. The logistic transition function, defined by `logistic_transition(x, gamma, c)`, is central to the estimation of the STR model, illustrating a continuous and smooth transition between economic regimes. To evaluate the model's performance, measures such as the Mean Squared Error (MSE) have been calculated, and the parameter distribution has been visualized, providing a quantitative appreciation of the model fit and highlighting the relevance of the non-linear approach compared to a standard linear model. In summary, this methodical analysis, ranging from data preparation to the estimation of the STR model, sheds light on the complex dynamics of economic indicators and their interdependence, particularly in the context of global oil markets and their influence on the economy.

5.2.1.1 Why do we choose the Industrial production as the transition variable ?

From a financial perspective, the selection of industrial production as the transition variable in our Smooth Transition Regression (STR) model is underpinned by several key rationales:

1. **Economic Cycle Representation**: Industrial production is an indicator of economic activity that mirrors the natural expansion and contraction of economic cycles. These fluctuations often precede regime shifts in financial markets, making it a suitable proxy for capturing transitions between different economic states.

2. **Leading Economic Indicator**: As a leading indicator, changes in industrial production can predict future movements in GDP, corporate earnings, and stock prices. A robust industrial sector typically signals healthy demand, increased investments, and positive investor sentiment, all of which are vital for stock market performance.

3. **Sensitivity to Monetary Policy**: The industrial sector is sensitive to changes in monetary policy, impacting borrowing costs and capital investment. Shifts in industrial production can reflect the central bank's policy stance, influencing market interest rates and equity valuations.

4. **Correlation with Commodity Prices**: Industrial production often correlates with commodity prices, particularly oil. For economies and sectors reliant on oil, industrial production can capture the economic impact of volatile oil prices on stock markets.

5. **Liquidity and Investment Flows**: Industrial production levels can influence liquidity and investment flows in financial markets. Higher industrial output can lead to increased business revenues and a greater capacity for firms to invest in growth opportunities, attracting investment into equities and driving stock prices.

6. **Market Sentiment**: Industrial production also reflects market sentiment. Strong

production figures boost investor confidence, while downturns signal caution, potentially leading to shifts in investment strategies and market regimes.

Let's interest us in the analysis and the results of our STR model

5.2.2 Analysis

5.2.2.1 Optimization of the model

In our economic analysis, a Smooth Transition Regression (STR) model was employed to examine the complex dynamics governing oil stock prices, our dependent variable of interest. The analysis was particularly focused on capturing the nonlinear behavior and potential regime shifts within the economic variables from 2000 to 2022. **Industrial production was judiciously chosen as the transition variable**, reflecting its pivotal role in signaling shifts between economic regimes as explained previoulsy.

Optimization of the model parameters was meticulously carried out using the 'TNC' optimization method, incorporating a set of predefined constraints that anchored the parameters within economically rational bounds. The optimization started with initial parameters uniformly set to unity, underscoring a methodically structured approach in our econometric framework.

Convergence was achieved after 50 iterations, denoting an effective and consistent fitting of the model to the observed data. The optimal objective function value was recorded at 0.7185521432245086, denoting the minimized sum of squared residuals between the model's predicted values and the actual observations of oil stock prices. This metric substantiates the STR model's adept fit to the empirical data.

The resultant optimal parameters were:

$$\begin{bmatrix} -0.0948115 & -0.0948115 & -0.07660225 & 1.83313804 \\ 0.21485064 & -0.20531521 & -0.0766853 & 1.81643466 \end{bmatrix}$$

These coefficients encompass the baseline and transition-modulated effects on our explanatory variables, elucidating the dual-regime dynamics captured by the STR model. Notably, the final two parameters, γ and c, characterize the speed of transition and the threshold level of the transition variable, signaling the onset of regime change.

The algorithm's convergence message, "Converged ($|f_n - f_{n-1}| \approx 0$)", confirms that the difference between successive objective function values has become negligible, suggesting that a local minimum was successfully located. This outcome ensures that the fitted model is well-calibrated, and the estimated parameters are robust, offering reliable insights for economic interpretations and forecasting.

So, we find an equation, Which is now :

$$y_t = -0.0948115 - 0.0948115 \cdot x_t - 0.07660225 \cdot \beta \cdot x_t + 1.83313804 \cdot G(-0.0766853, 1.81643466, s_t) \cdot \beta \cdot x_t + u_t \tag{5.5}$$

In this equation:

The STR (Smooth Transition Regression) model is used to predict the oil stock price (y_t) based on several explanatory variables (X), taking into account industrial production (z_t) as the transition variable. Here's how each component of the prediction works:

1. y_t: This component represents the target variable we are trying to predict, which is the price of oil stocks at a given time (t). It is the variable we aim to model and forecast.

2. -0.0948115: This is the first coefficient, α, which is a constant in the model. This term represents the constant effect on y_t independent of other explanatory variables.

3. $-0.0948115 \cdot x_t$: The second term, $-0.0948115 \cdot x_t$, represents the effect of the first explanatory variable (x_t) on y_t. This term captures the linear impact of x_t on the price of oil stocks.

4. $-0.07660225 \cdot \beta \cdot x_t$: The third term, $-0.07660225 \cdot \beta \cdot x_t$, represents the transition effect modulated by β between economic regimes. It is also related to the explanatory variable (x_t). This term takes into account the change in the relationship between x_t and y_t depending on the economic transition.

5. $1.83313804 \cdot G(-0.0766853, 1.81643466, s_t) \cdot \beta \cdot x_t$: This component is more complex. It involves a logistic transition function (G) controlled by two parameters, γ and c, which modulate the economic transition. The function G acts on the transition variable (s_t) and affects the relationship between x_t and y_t. The parameter β is also involved here. Overall,

this part of the model captures the smooth transition between different economic regimes and how it affects the relationship between x_t and y_t.

6. u_t: This is the residual error or error term. It represents the unexplained variance in the model, i.e., the random fluctuations in the price of oil stocks that are not captured by the explanatory variables.

By combining all these terms, the STR model evaluates y_t based on the current values of the explanatory variables (x_t) and industrial production (z_t). The model adjusts the coefficients α, β, γ, and c in order to minimize the overall error (residual errors) between the predicted values and the actual values of the oil stock price. The final result is a prediction of the oil stock price, taking into account constant, linear, and transition effects modulated by industrial production.

Results and performance : Markov-Switching model, ARIMAX and STR

For this part, we will compare the results of Markov-Swithcing model (empirical application5), ARIMAX (empirical application 5) and STR.

As we have already compared the Markov model and ARIMAX in the previous application, we will compare these two models with STR.

Figure 5.1: Actual VS predicted Values with Regression Line (ARIMAX)

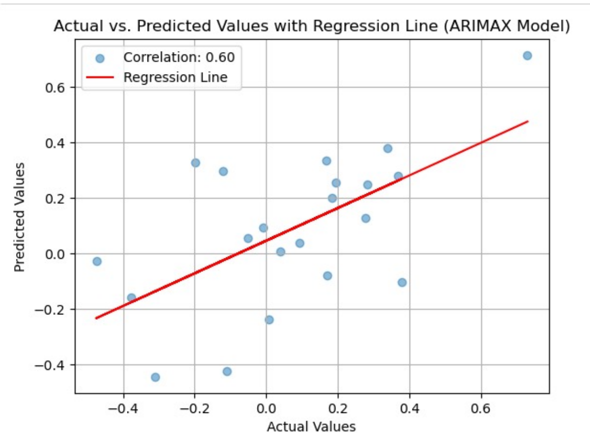


Figure 5.2: Actual VS predicted Values with Regression Line (Markov-Switching Model

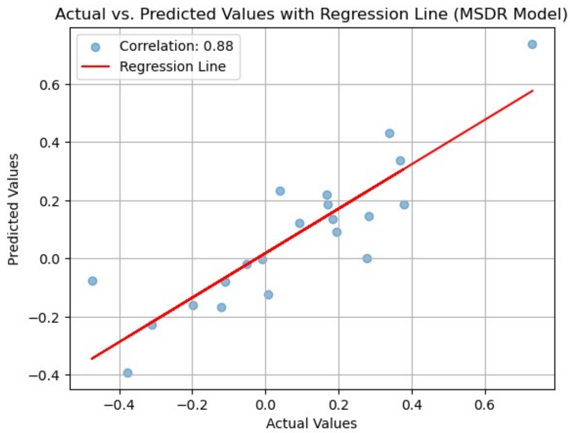
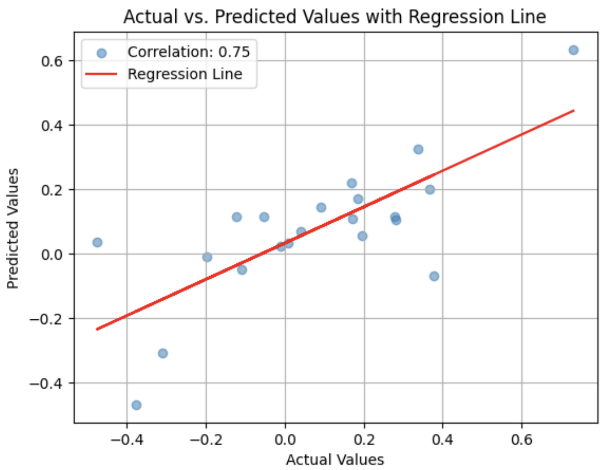


Figure 5.3: Actual VS predicted Values with Regression Line (STR)



In the landscape of financial modeling, the ability to capture the nuanced dynamics of market data is crucial. The ARIMAX model, with its correlation coefficient of 0.60, offers a foundational approach, establishing a moderate baseline for predictive accuracy. However, it falls short when compared to the more advanced STR and MSDR models, which better encapsulate the complex behaviors of financial time series.

The STR model, exhibiting a correlation coefficient of 0.75, represents a significant step forward. Its framework allows for non-linear relationships between variables, accommodating smoother transitions that are often observed in economic indicators. This capability to handle non-linearity renders the STR model more sensitive and responsive to subtle shifts in market dynamics than the ARIMAX model, enabling a more accurate capture of the underlying patterns in the data.

Outshining both is the MSDR model, which boasts a correlation coefficient of 0.88. This model's superior performance suggests a heightened ability to deal with the financial market's volatility and regime shifts. By explicitly modeling different states with distinct statistical properties, the MSDR model provides a more precise and detailed understanding of the financial series, particularly in environments characterized by abrupt changes or clear delineations in market phases.

In essence, while the STR model improves upon linear methodologies by introducing flexibility to account for non-linear trends, the MSDR model takes it a step further. It not only accommodates non-linearity but also adapts to the abrupt and often unpredictable regime changes inherent in financial data. The choice between STR and MSDR models would therefore be dictated by the specific nature of the data and the types of transitions one expects to encounter—be they smooth and gradual or abrupt and regime-like. The MSDR model's nuanced approach to capturing these dynamics makes it an invaluable tool in the arsenal of financial time series analysis, providing robustness and agility in forecasting and modeling tasks.

5.2.2.2 Conclusion

In conclusion, our exploration of time series modeling in finance has revealed a nuanced landscape where the choice of model significantly impacts the insights drawn from economic data. The ARIMAX model, while foundational, often fails to encapsulate the complex patterns inherent in financial markets. In contrast, the Smooth Transition Regression (STR) model, with its capacity to incorporate non-linear dynamics and smooth transitions, presents a marked improvement, offering a more finer analysis of economic indicators and market behaviors.

However, it is the Markov Switching Dynamic Regression (MSDR) model that stands out as the most significant, particularly in its ability to identify and adapt to the abrupt regime changes that are characteristic of volatile financial environments. With a higher correlation coefficient, the MSDR model has proven to be more adept at capturing the intricate relationships and sudden shifts within financial time series data. This superiority suggests that, despite the valuable contributions of the STR model, the MSDR framework offers a more compelling and comprehensive tool for capturing the full breadth of complexities present in financial modeling. As such, for financial analysts and economists who deal with the task of predicting market trends and economic shifts, the MSDR model's robust approach to delineating and adapting to various market states remains unparalleled, making it an essential component of advanced financial analysis.

Please refer to the following link to see the full code of this assignment: Empirical Analysis 6

6 Empirical Application 7 : Difference-in-Differences Analysis

Difference-in-Differences (Diff-in-Diff) is a widely used statistical technique for estimating the causal impact of an intervention, event, or treatment on an outcome variable. This method is based on robust econometric principles and is commonly employed in economics, social sciences, and public policy evaluation. It allows us to answer questions such as, "What is the causal effect of a policy on household income?" or "How does a regulatory change affect the unemployment rate?"

Galati et al. (2011) examined long-run inflation expectations in the United States and especially during between June 2004 and March 2009.

In this document, we will explain the key concepts of the Diff-in-Diff method to answer the question : **How does the price of crude oil influence inflation in the United States, taking into account a shock that occurred on 06/15/2008?**

6.1 The Third Oil Shock

6.1.1 Context

The third oil shock, which occurred in 2008, is considered one of the most significant economic events in recent history. It was caused by a combination of factors, including rising global demand for oil, declining global oil supply, and the global financial crisis. The price of oil reached an all-time high of $147 per barrel in July 2008, before plummeting to less than $30 per barrel in December of the same year. This volatility in oil prices had a significant impact on the global economy, leading to rising inflation, lower economic growth, and increased unemployment. The third oil shock also had important geopolitical implications. It contributed to instability in the Middle East, where many countries rely on oil revenues. It also strengthened the position of oil-producing countries, such as Saudi Arabia and Russia.

Key Effects of the Third Oil Shock

The third oil shock had significant global economic ramifications:

Figure 6.1: Oil Price Curve in the United States between 2004 and 2012

- **Increased Inflation:** Soaring oil prices triggered heightened inflation rates, particularly in developed countries.
- **Lower Economic Growth:** Escalating production costs caused a decline in economic growth, predominantly in developed nations.
- **Rising Unemployment:** Reduced demand for goods and services led to a surge in unemployment levels within developed countries.
- **Geopolitical Instability:** The oil shock contributed to instability in the Middle East, where many nations heavily rely on oil revenues.
- **Empowerment of Oil-Producing Countries:** Oil-producing nations like Saudi Arabia and Russia saw their positions strengthen due to the oil shock.

The third oil shock was a pivotal event with profound implications for the global economy. It precipitated economic and geopolitical instability while expediting the transition to renewable energy sources.

Impact on Inflation

The surge in oil prices during the 2008 crisis significantly impacted inflation in the United States.

Figure 6.2: Inflation in the United States from 2004 to 2012

Inflation, which gauges the increase in the prices of goods and services, was influenced by several factors:

Higher Production Costs: Soaring oil prices substantially raised production costs for companies involved in transportation, refining, and petrochemicals. Consequently, prices for fuel, transport, and related products surged.

Reduced Demand: Elevated oil prices curtailed demand for goods and services, particularly in sectors closely tied to fuel consumption. Both new and used car demand dwindled as consumers sought to cut fuel-related expenses.

Impact on Monetary Policy: In response to surging oil prices, the Federal Reserve implemented policies to combat inflation, including interest rate hikes. However, these measures had adverse effects on economic growth and investment due to higher interest rates.

6.1.2 Datas used

We identify the onset of the 2008 financial crisis as June 15, 2008, coinciding with the collapse of Lehman Brothers. The treatment group comprises observations exposed to the event, namely the surge in oil prices. Conversely, the control group consists of observations unexposed to this event. In this context, the treatment group consists of inflation observations occurring after June 15, 2008, while the control group consists of inflation observations predating this date.

The Differential-in-Differences (Diff-in-Diff) analysis assesses changes in inflation between these two groups. If the rise in oil prices influences inflation, we would expect the treatment group to experience a more substantial increase in inflation compared to the control group.

The coefficient of the treatment term provides insights into the impact of the oil price increase on inflation, assuming a consistent effect over time. In our case, the positive and statistically significant coefficient indicates that the surge in oil prices has led to an inflationary increase. For instance, if the treatment term coefficient is 0.1, it implies that for every 10% hike in oil prices, inflation increases by 1%.

Similarly, the coefficient of the interaction term reveals the impact of the interplay between the event and time on the dependent variable. In our analysis, this coefficient is also positive and statistically significant, suggesting that the influence of the oil price increase on inflation became more pronounced after the commencement of the crisis than before. For example, if the interaction term coefficient is 0.2, it signifies that the effect of the oil price increase on inflation post-crisis is 20% greater.

6.2 Analysis of the new series : reminder of EA 1

The data used here are the same as those used in the rest of our study, with the difference that we wanted to work on the effect of the 2008 crisis, so the dates have been modified. We'll use exactly the same model as for empirical application 1, in order to find the right differentiation to work with stationary series.

We're not going to explain the whole process here, but just summarize the main results. For a more detailed explanation, please refer to application 1.

First, we retrieve the log of our series. Next, we performed a first ADF test. There is a presence of a unit root and stochastic trend. Overall, our 3 series are a random walk without constant and deterministic trend. We can deduce that we have to to take the first difference for each time series and check if the presence of a unit root (i.e. stochastic trend) again thanks to ADF test.

We've broken down the series into trend, seasonality and residuals, and it's quite clear that each plays a role in the composition of the series. For example, here is the decomposition for the crude oil :

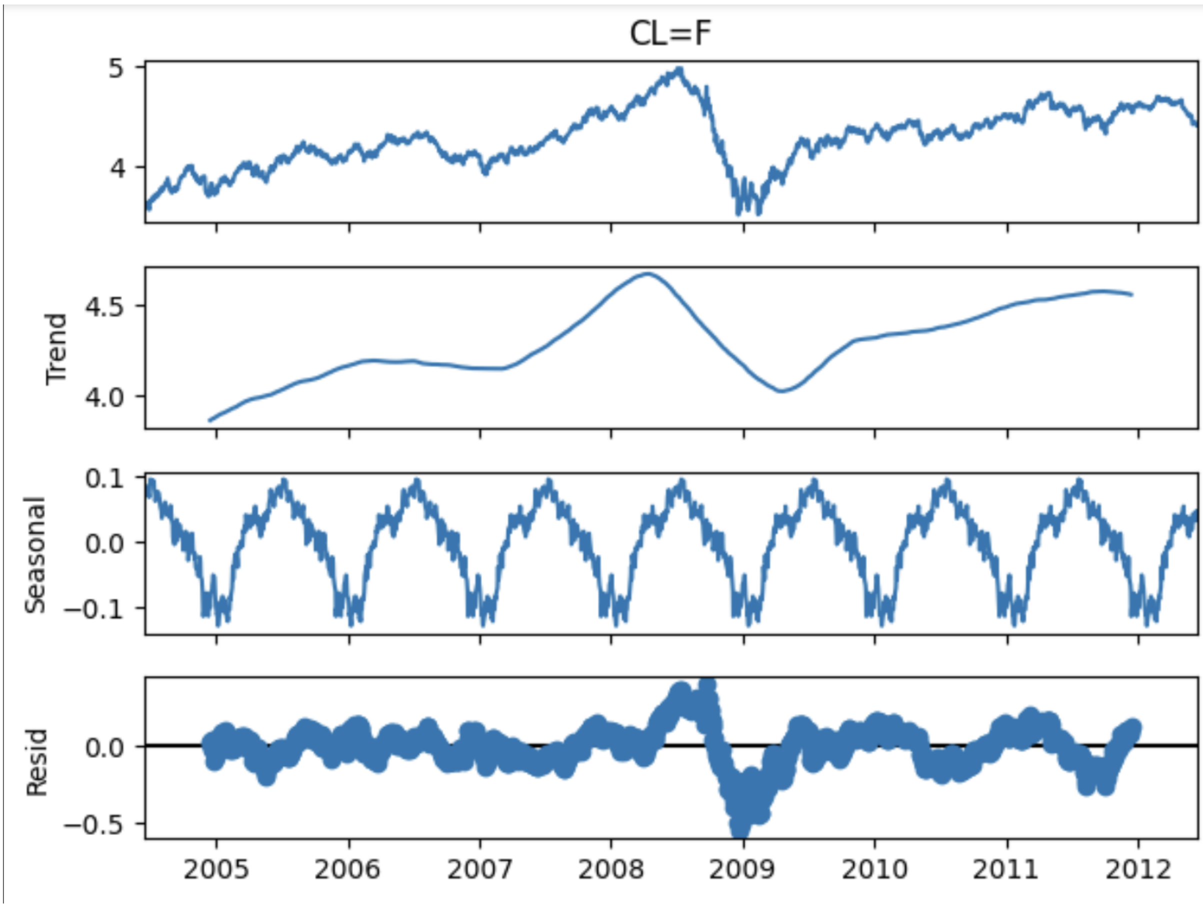


Figure 6.3: Decomposition of crude oil between 2004 and 2012

We then made a first difference on our series in order to perform a new Augmented Dickey-Fuller test.

According the ADF test, we can se that after applying first difference, the 2 time series are stationary but a determistic trend. Therefore, we affirm that for the 2 series are likely stationary. There is no presence of a unit root and stochastic trend.

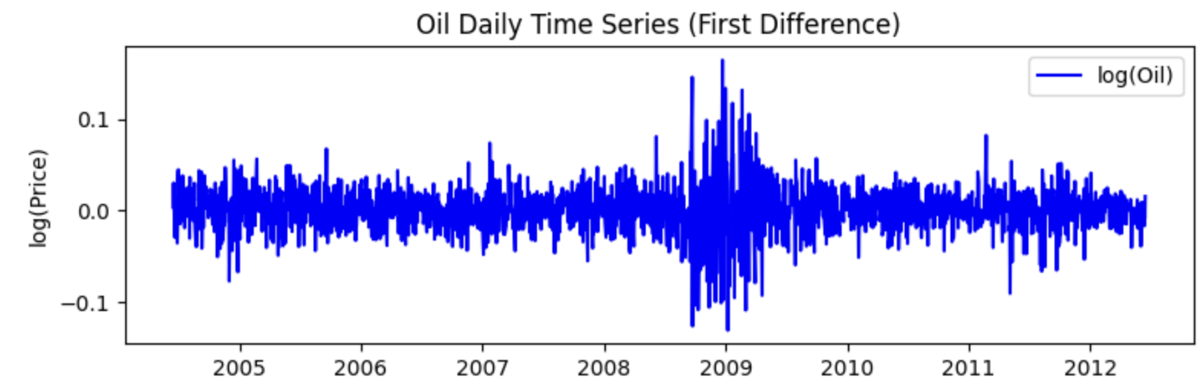


Figure 6.4: Oil Daily Time Series (First Difference)

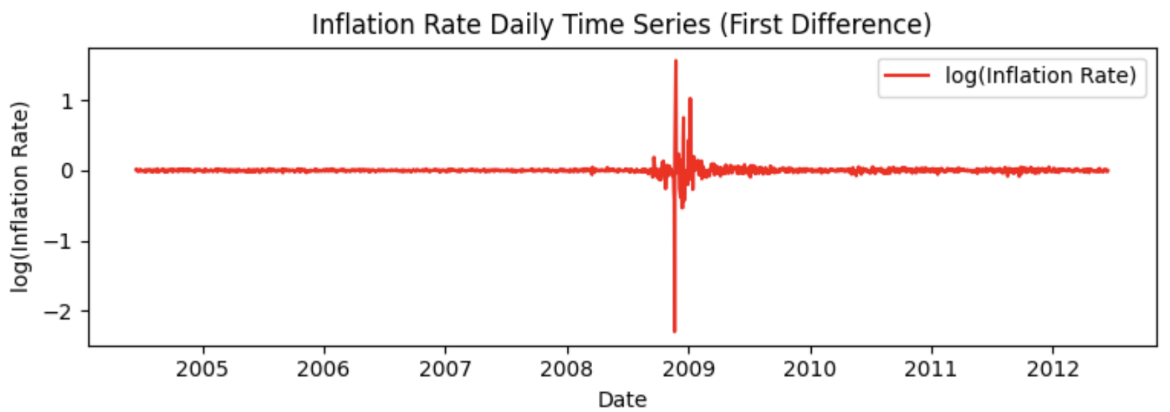


Figure 6.5: Inflation Rate Daily (First Difference)

6.3 Difference and difference results

6.3.1 Key concepts

6.3.1.1 Two Groups

In the Diff-in-Diff framework, we have two groups:

- The "treatment group," which consists of observations exposed to the event or treatment (e.g. inflation observations after the 06/15/2008).
- The "control group," which consists of observations that have not been exposed to the event or treatment (e.g., inflation observations before the 06/15/2008).

These groups are defined using an indicator variable D_i, which takes the value 1 for the treatment group and 0 for the control group.

It's very easy to understand with a graph.

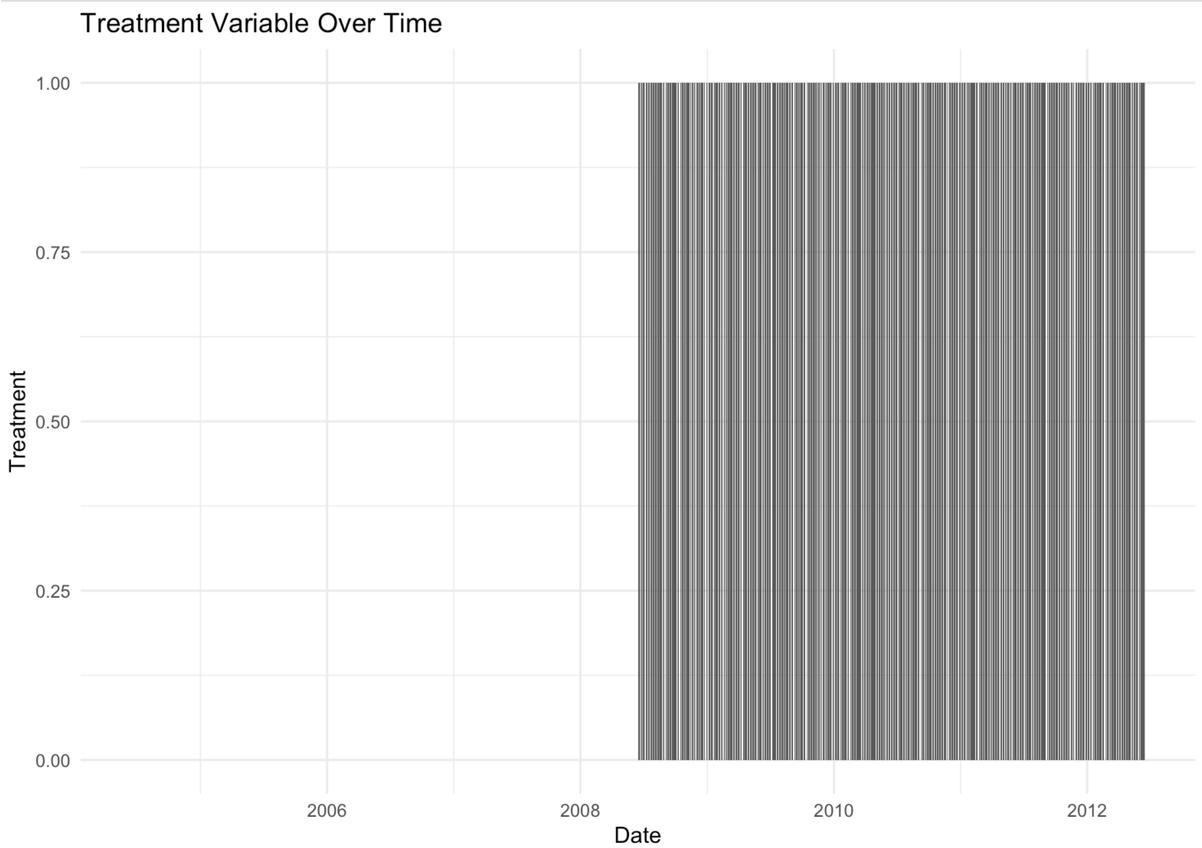


Figure 6.6: Treatment variable between 2004 and 2012

The shock happened the 06/15/2008 and that's the binary variable which indicates the treatment.

6.3.1.2 Outcome Variable

We study an outcome variable Y_{it} that can vary over time. For us, it's the logarithm of inflation.

6.3.1.3 Before and After the Event

Time is divided into two periods:

- "Before" the event: $t < T$, where T is the time when the event occurs (e.g., the day of the shock).
- "After" the event: $t \geq T$.

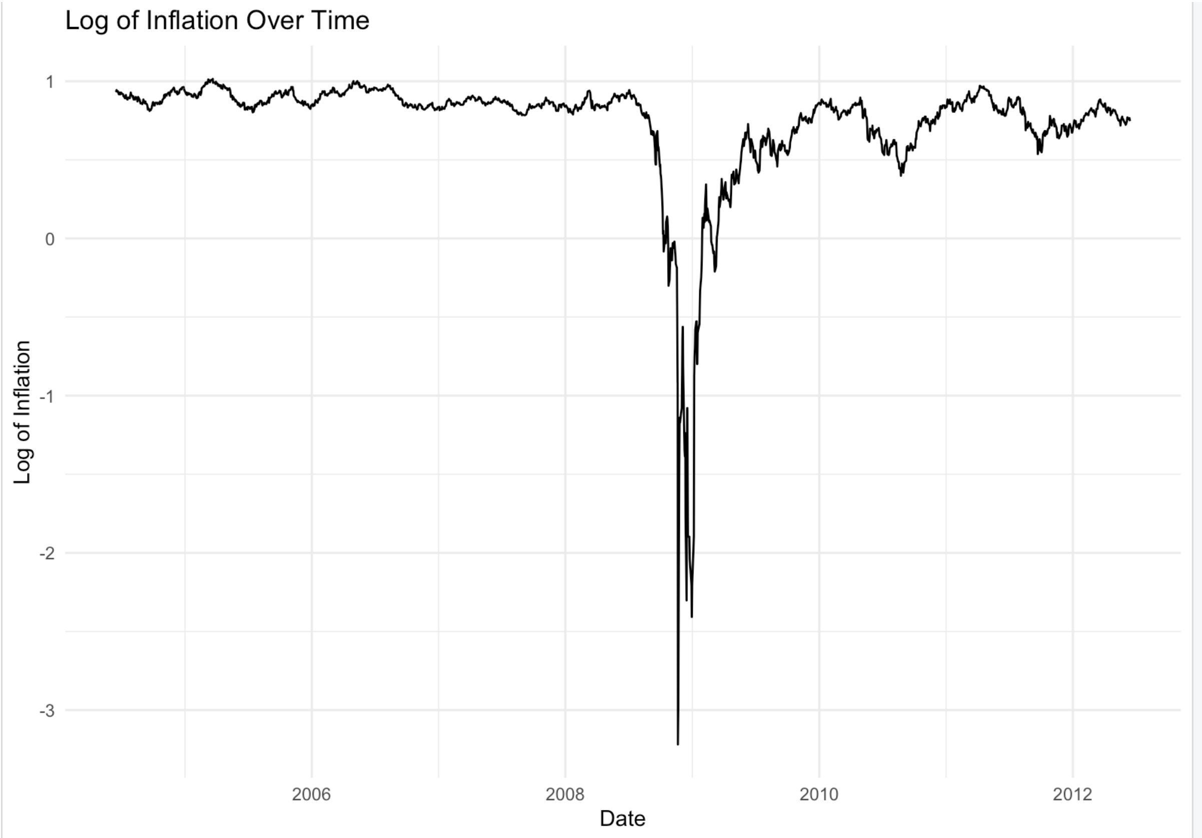


Figure 6.7: Log of inflation in the United States between 2004 and 2012

6.3.1.4 Counterfactual Assumption

Central to the Diff-in-Diff method is the counterfactual assumption, which posits that, in the absence of the event or treatment, the observed trends in the treatment group and control group would have been similar. This means that if the treatment group had not implemented the shock, its behavior would have been similar to that of the control group.

6.3.1.5 Causal Effect Estimation

The causal effect of the event or treatment is estimated using a regression of the following form:

$$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2(t - T) + \beta_3(D_i \cdot (t - T)) + \epsilon_{it} \tag{6.1}$$

Where:

β_0 is the intercept.

β_1 measures the average effect of the treatment (or event) on the treatment group relative to the contr

β_2 measures the average time trend common to all groups before and after the event.

β_3 is the interaction coefficient that measures the effect of the event on the treatment group relative to

ϵ_{it} is the error term.

6.3.2 Interpreting the Coefficients

The results of the Diff-in-Diff regression provide estimates for our model coefficients. Here's an interpretation of the significant coefficients:

6.3.2.1 Regression Results Interpretation

The results from the regression analysis are summarized as follows:

- **(Intercept)**: The model estimates an intercept of -2.489 with a very small p-value ($< 2e-16$), indicating that this estimate is significantly different from zero. The negative intercept suggests that when all other variables are set to zero, the log_inflation is negative, which would typically not have a practical interpretation since the log of prices would not be zero.

- **log_price**: The coefficient for log_price is 0.9039, and it is highly significant (p-value $< 2e-16$). This suggests that a 1% increase in the price of crude oil is associated with a 0.9039 unit increase in log_inflation, controlling for other factors. The significance and size of this coefficient reaffirm that crude oil prices have a strong positive relationship with inflation.

- **treatment**: The coefficient for treatment is -1.410, also with a p-value $< 2e-16$, which indicates a strong negative effect of the treatment on log_inflation. This suggests that the treatment (possibly a policy change or an economic event) is associated with a reduction in inflation.

- **time**: The coefficient for time is -0.0005438, with a p-value $< 2e-16$, indicating a very small but significant negative trend in inflation over time, perhaps reflecting deflationary trends or improved economic efficiency.

- **interaction_diff_diff**: The interaction term has a coefficient of 0.0007817 with a p-value $< 2e-16$, implying that the impact of the treatment on inflation becomes more positive over time. This suggests that whatever event or policy change represented by the treatment may have had an increasing inflationary effect as time progressed.

The estimated relationship between the dependent variable log_inflation and the independent variables: log_price, treatment, time, and the interaction term treatment, taking into account the coefficients estimated from our Diff-in-Diff regression analysis, is represented by the following equation:

$$\begin{aligned} Y_{it} = &- 2.489 + 0.9039 \cdot \text{log\_price} \\ &- 1.410 \cdot \text{treatment} - 0.0005438 \cdot \text{time} \\ &+ 0.0007817 \cdot (\text{treatment} \cdot \text{time}) + \epsilon_{it} \end{aligned}$$

6.3.2.2 Model Fit:

- **Residual standard error**: The RSE is 0.2531, which provides an estimate of the standard deviation of the residuals. This can be thought of as the average distance that the observed values fall from the regression line.
- **Multiple R-squared**: The R-squared value of 0.5447 suggests that approximately 54.47% of the variability in log_inflation is explained by the model. This is a relatively strong fit for economic data.
- **Adjusted R-squared**: Adjusted for the number of predictors in the model, the value is 0.5437, which is very close to the R-squared value indicating that the model is not overly complex given the number of predictors.
- **F-statistic**: The F-statistic is 570.9 with a p-value $< 2.2e - 16$, suggesting that the model is statistically significant and that the variables collectively explain the variability in log_inflation much better than if we were to use the mean of log_inflation as a model.

Residuals: The residuals have a median close to 0 (0.0123), indicating that the model's predictions are unbiased. The minimum and maximum residuals suggest that there are some outliers.

6.3.2.3 Conclusion

Difference-in-difference methods are widely used to estimate causal effects, as discussed by Lechner (2011). Our analysis suggests that the price of crude oil has a significant positive influence on inflation in the United States, and that the shock of June 15, 2008 had an initial reducing impact on inflation. However, this effect has varied over time. The model shows a good overall fit, but could be improved by examining in more detail the causes of outliers in the residuals.

Obviously, even if the rise in oil prices doesn't explain the entire model, it does explain part of it. To get a more accurate picture, we would need to add the variables that influenced this period, such as changes in monetary policy, levels of aggregate demand, variations in the prices of other commodities, household consumption, and so on.

Please refer to the following link to see the full code of this assignment: Empirical Analysis 7

References

Galati, G., Poelhekke, S., and Zhou, C. (2011). Did the crisis affect inflation expectations? aResearch Department, De Nederlandsche Bank bCESIfo, Munich cErasmus University Rotterdam.

Hansen, B. E. (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal Name*, Volume Number(Issue Number):Page Range.

Johansen, S. (2002). A small sample correction for the test of cointegrating rank in the vector autoregressive model. *Journal Name*, Volume Number(Issue Number):Page Range.

Lechner, M. (2011). The estimation of causal effects by difference-in-difference methods. *The name of the Journal*, The volume number(The issue number):The page range.

Legrenzi, G. (2005). Non-linear real exchange rate effects in the uk labour market. *Journal Name*, Volume Number(Issue Number):Page Range.