# Work done in Data Processing

Given on sheet 2 were the prices of 8 indices, the first and foremost job was to clean the data and only consider those dates for which we have prices for all eight indices. We have prices for IJA Comdty (R2) and PKR1 Comdty (L2) from 6/16/2023 to 1/29/2021. So for analysis, I've not taken these two price indexes. The cleaned dataset of six commodity price indexes from 2023-06-16 to 2013-01-04 is here [Link].
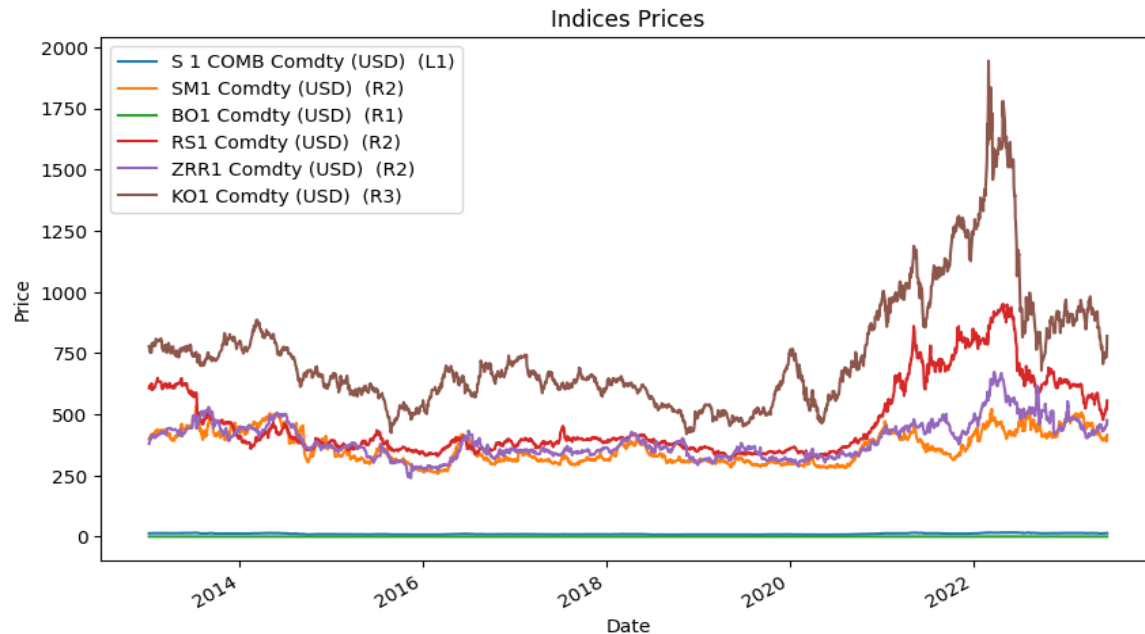
The returns of all these six indexes can be found here[Link]
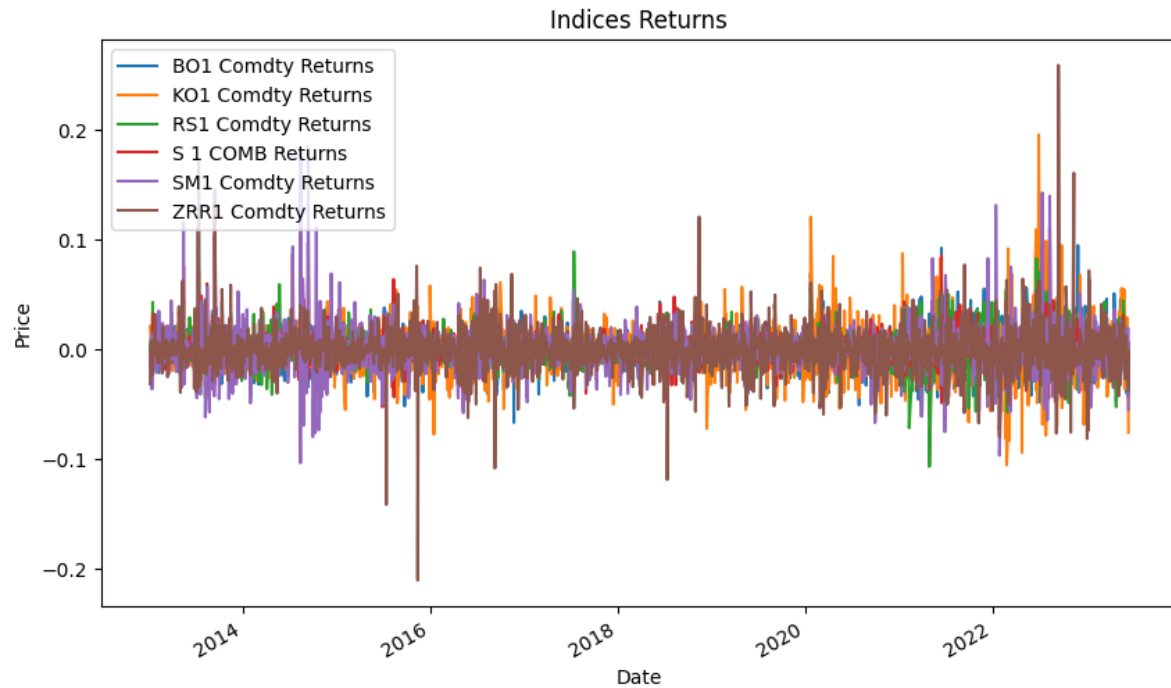
Further descriptive statistics of all the vegetable oil containing the following Obs. needs to be done.

- **Mean**
- **Maximum**
- **Minimum**
- **Standard deviation**
- **Skewness Kurtosis**
- **Jarque-Bera**
- **Q (5)**
- **Q (10)**
- **Q (20)**
- **ARCH (5)**
- **ARCH (10)**
- **ARCH (20)**
- **ADF**
- **P-P**

Also, the TVP-VAR analysis needs to be done with the external effects of the US Economic Policy Uncertainty Index, Dollar Index Spot Currency and Generic 1st 'CL.' Future Commodity

The analysis of US Economic Policy Uncertainty Index is [Link to Google Colab] Summary of ARIMA(2,1,2) Model

```
                              SARIMAX Results
==============================================================================
Dep. Variable:                Mid Price   No. Observations:                1825
Model:                  ARIMA(2, 1, 2)   Log Likelihood              -10224.515
Date:                Sun, 25 Jun 2023   AIC                          20459.030
Time:                        12:38:36   BIC                          20486.574
Sample:                             0   HQIC                         20469.191
                               - 1825
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.2603      0.593     -0.439      0.661      -1.423       0.903
ar.L2          0.1023      0.096      1.065      0.287      -0.086       0.290
ma.L1         -0.4196      0.595     -0.706      0.480      -1.585       0.746
ma.L2         -0.3706      0.503     -0.737      0.461      -1.356       0.615
sigma2      4327.4563     80.918     53.480      0.000    4168.861    4486.052
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):              1953.49
Prob(Q):                              0.94   Prob(JB):                         0.00
Heteroskedasticity (H):               0.45   Skew:                             1.16
Prob(H) (two-sided):                  0.00   Kurtosis:                         7.51
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```
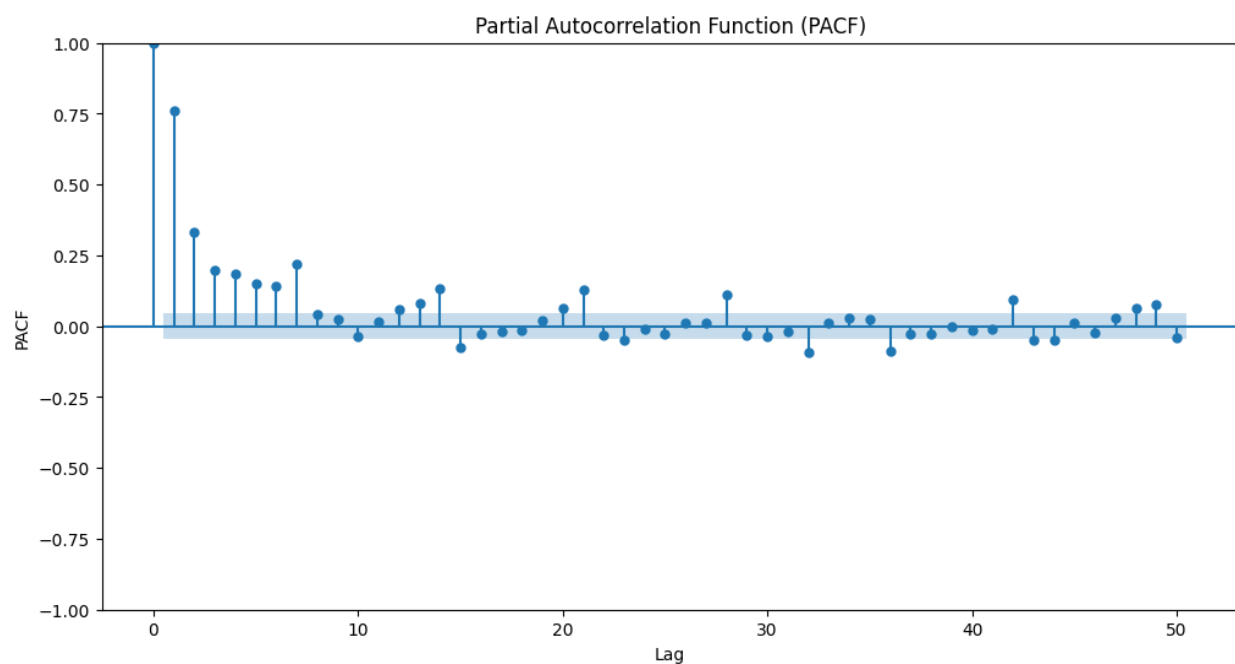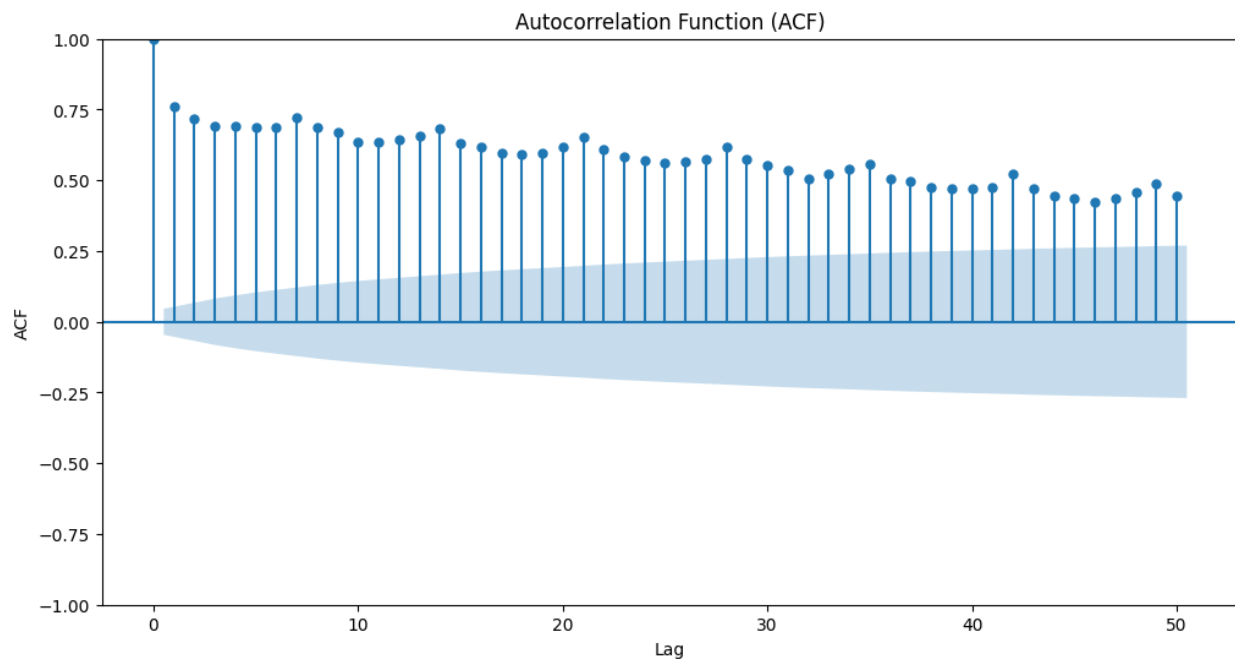

ARMA Model Predictions

Autocorrelation Function (ACF)

Partial Autocorrelation Function (PACF)
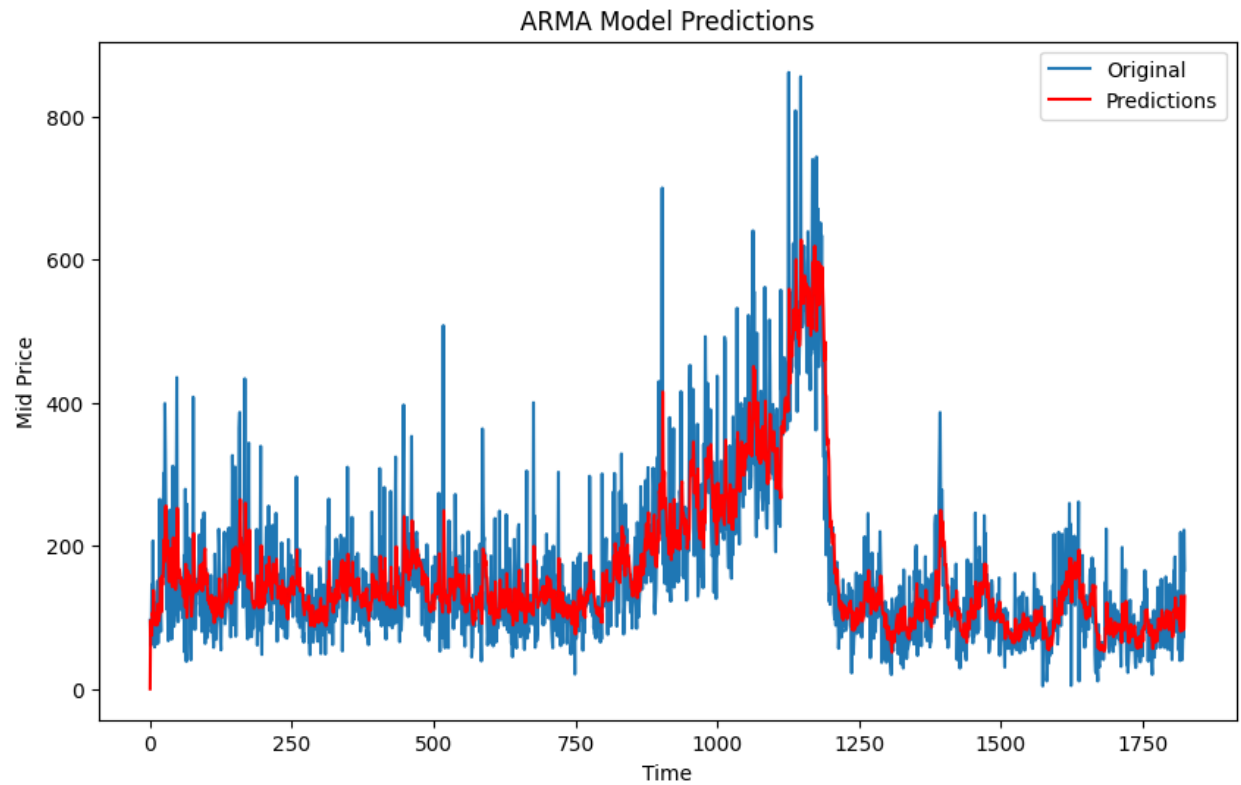
ADF Statistics

ADF Statistic: -2.7129590059669355
p-value: 0.07182251516198512
Critical Values:
   1%: -3.4339840952648695

5%: -2.8631452508003057
　　10%: -2.567624583142913

Based on the ADF test results, we do not have sufficient evidence to conclude that the time series is stationary. The data may exhibit some degree of non-stationarity. However, the results are not definitive, and further analysis or exploration of the data may be necessary to determine its stationarity.

On Applying the Moving Average Smoothing technique, the new ADF Statistics we get are:

ADF Statistic: -2.9056001212886695
p-value: 0.044700644401516075
Critical Values:
　　1%: -3.434002362895342
　　5%: -2.863153315036348
　　10%: -2.5676288771043367
Now the series is stationary.

Later using the AIC criterion to get the best-fitting model, we get the following table

	(p, d, q)	AIC
0	(7, 1, 7)	20287.032907
1	(7, 1, 5)	20295.519637
2	(7, 1, 6)	20295.774217
3	(6, 1, 5)	20297.747802
4	(4, 1, 5)	20311.139954
...	...	...
59	(4, 1, 0)	20576.307078
60	(3, 1, 0)	20627.737892
61	(2, 1, 0)	20711.444622
62	(1, 1, 0)	20819.961076
63	(0, 1, 0)	21156.712295

64 rows × 2 columns

Therefore, this suggests are ARIMA model with an AR(7) process and a MA(7) process.

Now, we can print a summary of the best model, which an ARIMA (7,1,7).

```
                               SARIMAX Results
==========================================================================
Dep. Variable:             Mid Price   No. Observations:            1825
Model:               ARIMA(7, 1, 7)    Log Likelihood          -10128.516
Date:              Sun, 25 Jun 2023    AIC                      20287.033
Time:                      12:42:38    BIC                      20369.665
Sample:                           0    HQIC                     20317.515
                             - 1825
Covariance Type:                opg
==========================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------
ar.L1         -0.6162      0.128     -4.827      0.000      -0.866      -0.366
ar.L2         -0.5855      0.126     -4.654      0.000      -0.832      -0.339
ar.L3         -0.8299      0.048    -17.291      0.000      -0.924      -0.736
ar.L4         -0.7639      0.086     -8.898      0.000      -0.932      -0.596
ar.L5         -0.5464      0.139     -3.944      0.000      -0.818      -0.275
ar.L6         -0.6966      0.092     -7.578      0.000      -0.877      -0.516
ar.L7          0.1236      0.036      3.431      0.001       0.053       0.194
ma.L1         -0.0776      0.125     -0.621      0.535      -0.323       0.168
ma.L2          0.0672      0.096      0.697      0.486      -0.122       0.256
ma.L3          0.3732      0.083      4.520      0.000       0.211       0.535
ma.L4          0.1693      0.098      1.721      0.085      -0.023       0.362
ma.L5         -0.0945      0.111     -0.855      0.393      -0.311       0.122
ma.L6          0.2253      0.086      2.628      0.009       0.057       0.393
ma.L7         -0.5778      0.081     -7.139      0.000      -0.736      -0.419
sigma2      4036.7515     91.244     44.241      0.000    3857.916    4215.586
==========================================================================
Ljung-Box (L1) (Q):                0.03   Jarque-Bera (JB):         1687.57
Prob(Q):                           0.87   Prob(JB):                    0.00
Heteroskedasticity (H):            0.44   Skew:                        1.01
Prob(H) (two-sided):               0.00   Kurtosis:                    7.26
==========================================================================
```
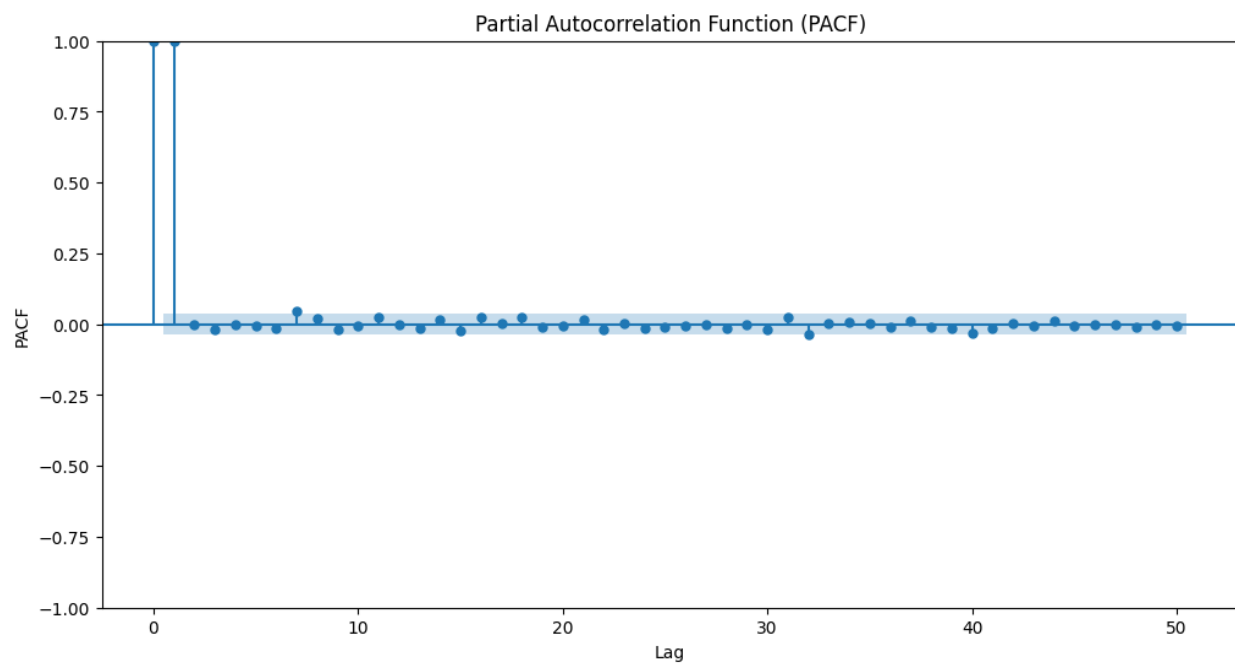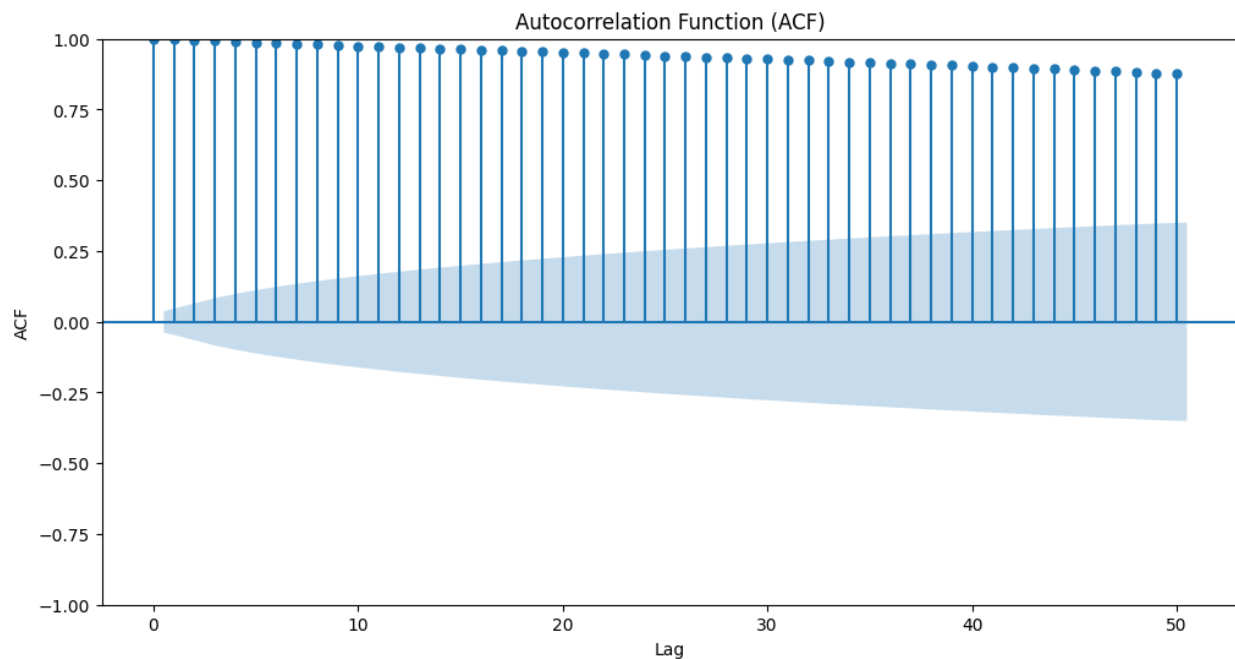
The analysis of Dollar Index Spot Currency [Link to Google Colab]
Summary of ARIMA(2,1,2) model

```
                              SARIMAX Results
==============================================================================
Dep. Variable:               Mid Price   No. Observations:                1825
Model:               ARIMA(7, 1, 7)   Log Likelihood              -10128.516
Date:               Sun, 25 Jun 2023   AIC                          20287.033
Time:                     12:42:38   BIC                          20369.665
Sample:                            0   HQIC                         20317.515
                             - 1825
Covariance Type:                 opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.6162      0.128     -4.827      0.000      -0.866      -0.366
ar.L2         -0.5855      0.126     -4.654      0.000      -0.832      -0.339
ar.L3         -0.8299      0.048    -17.291      0.000      -0.924      -0.736
ar.L4         -0.7639      0.086     -8.898      0.000      -0.932      -0.596
ar.L5         -0.5464      0.139     -3.944      0.000      -0.818      -0.275
ar.L6         -0.6966      0.092     -7.578      0.000      -0.877      -0.516
ar.L7          0.1236      0.036      3.431      0.001       0.053       0.194
ma.L1         -0.0776      0.125     -0.621      0.535      -0.323       0.168
ma.L2          0.0672      0.096      0.697      0.486      -0.122       0.256
ma.L3          0.3732      0.083      4.520      0.000       0.211       0.535
ma.L4          0.1693      0.098      1.721      0.085      -0.023       0.362
ma.L5         -0.0945      0.111     -0.855      0.393      -0.311       0.122
ma.L6          0.2253      0.086      2.628      0.009       0.057       0.393
ma.L7         -0.5778      0.081     -7.139      0.000      -0.736      -0.419
sigma2      4036.7515     91.244     44.241      0.000    3857.916    4215.586
===================================================================================
Ljung-Box (L1) (Q):                   0.03   Jarque-Bera (JB):              1687.57
Prob(Q):                              0.87   Prob(JB):                         0.00
Heteroskedasticity (H):               0.44   Skew:                             1.01
Prob(H) (two-sided):                  0.00   Kurtosis:                         7.26
```

The ADF Statistics of the given time series

ADF Statistic: -1.0493752063996695
p-value: 0.7348328267904379
Critical Values:
 1%: -3.43274059216515
 5%: -2.8625961912802333

10%: -2.567332244679086

The critical values at different significance levels (1%, 5%, and 10%) are provided for comparison with the ADF statistic. If the ADF statistic is more negative (smaller in magnitude) than these critical values, it provides stronger evidence in favor of stationarity. In this case, the ADF statistic does not exceed the critical values significantly. Based on these results, we can infer that the time series is likely non-stationary and may contain a unit root. Further analysis or transformation may be required to achieve stationarity in the data.

Applying log difference for stationarity
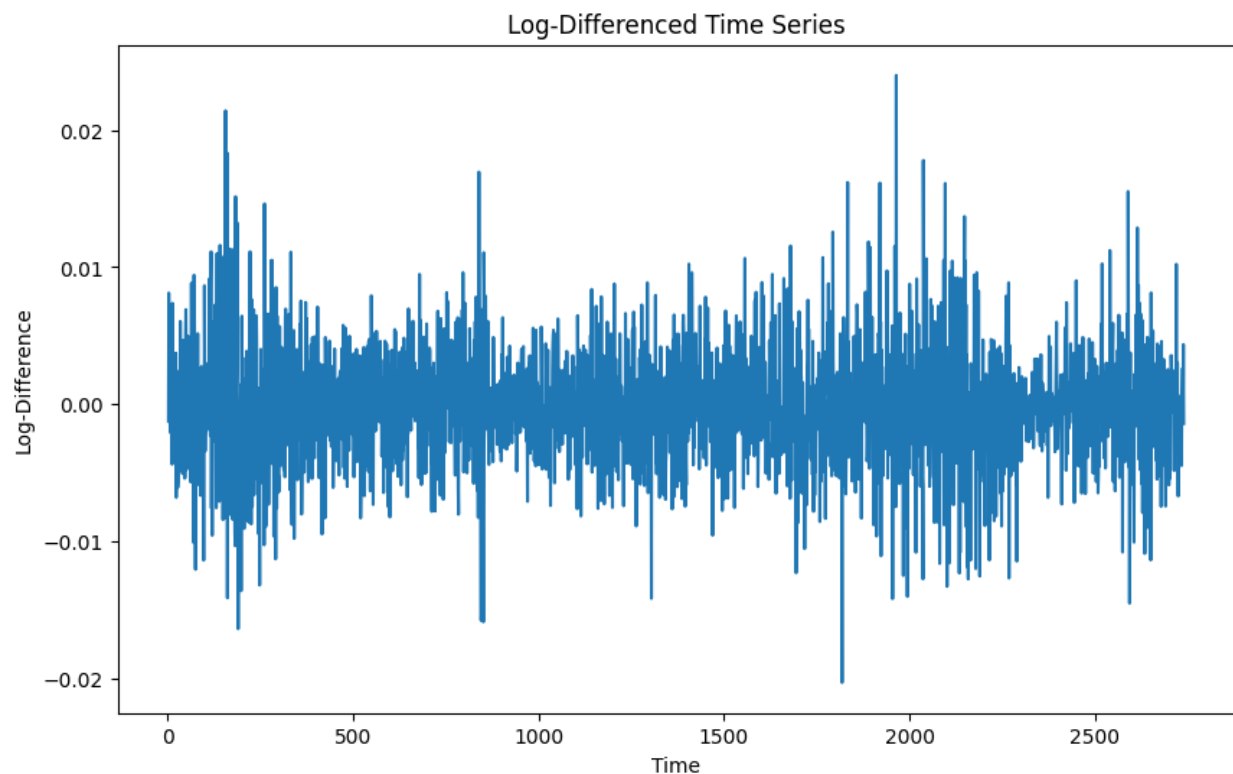ADF Statistic: -21.24440353107157
p-value: 0.0
Critical Values:
    1%: -3.4327467254180486
    5%: -2.8625988999327463
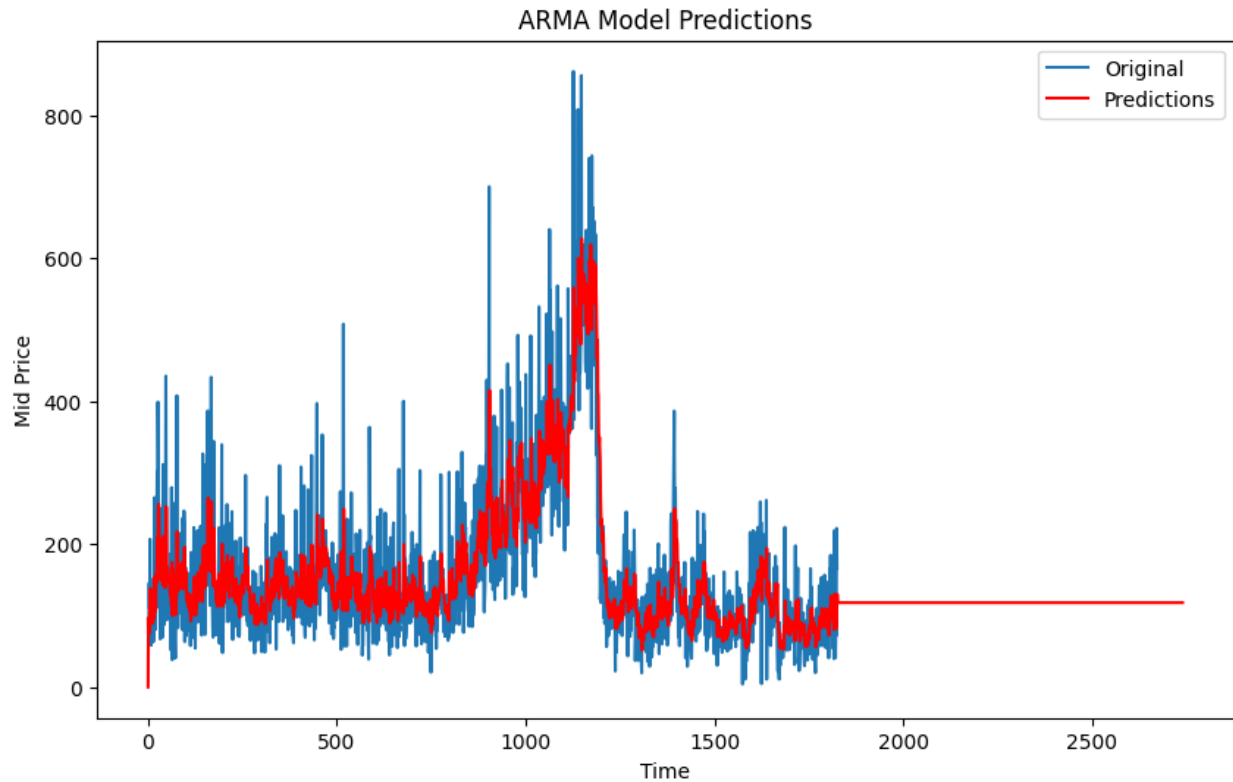   10%: -2.567333686767106

Log-Differenced Time Series

Since this dataset was small, AIC criterion was not applied.

The analysis of Generic 1st 'CL' Futures [Link to Google Colab]
Summary of ARIMA(2,1,2) Model

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                 Mid Price   No. Observations:           1825
Model:                 ARIMA(7, 1, 7)   Log Likelihood          -10128.516
Date:                 Sun, 25 Jun 2023   AIC                      20287.033
Time:                       12:42:38   BIC                      20369.665
Sample:                            0   HQIC                     20317.515
                              - 1825
Covariance Type:                 opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.6162      0.128     -4.827      0.000      -0.866      -0.366
ar.L2         -0.5855      0.126     -4.654      0.000      -0.832      -0.339
ar.L3         -0.8299      0.048    -17.291      0.000      -0.924      -0.736
ar.L4         -0.7639      0.086     -8.898      0.000      -0.932      -0.596
ar.L5         -0.5464      0.139     -3.944      0.000      -0.818      -0.275
ar.L6         -0.6966      0.092     -7.578      0.000      -0.877      -0.516
ar.L7          0.1236      0.036      3.431      0.001       0.053       0.194
ma.L1         -0.0776      0.125     -0.621      0.535      -0.323       0.168
ma.L2          0.0672      0.096      0.697      0.486      -0.122       0.256
ma.L3          0.3732      0.083      4.520      0.000       0.211       0.535
ma.L4          0.1693      0.098      1.721      0.085      -0.023       0.362
ma.L5         -0.0945      0.111     -0.855      0.393      -0.311       0.122
ma.L6          0.2253      0.086      2.628      0.009       0.057       0.393
ma.L7         -0.5778      0.081     -7.139      0.000      -0.736      -0.419
sigma2      4036.7515     91.244     44.241      0.000    3857.916    4215.586
===================================================================================
Ljung-Box (L1) (Q):                0.03   Jarque-Bera (JB):              1687.57
Prob(Q):                           0.87   Prob(JB):                         0.00
Heteroskedasticity (H):            0.44   Skew:                             1.01
Prob(H) (two-sided):               0.00   Kurtosis:                         7.26
```

ARMA Model Predictions

The ADF Statistics of the time series:
ADF Statistic: -3.113222352376439
p-value: 0.025590024112739685
Critical Values:
   1%: -3.432764423688327
   5%: -2.8626067160580204
   10%: -2.567337848081528

In summary, based on the ADF test results, we can conclude that the time series is likely stationary. This implies that the series does not exhibit a unit root and has a stable mean and covariance structure, making it suitable for certain time series analysis techniques and modeling approaches.

Applying AIC criterion to find the best model, we get

	(p, d, q)	AIC
0	(2, 1, 5)	11325.908264
1	(1, 1, 1)	11329.617687
2	(0, 1, 3)	11330.406411
3	(0, 1, 2)	11330.435204
4	(2, 1, 1)	11330.668329
...	...	...
59	(4, 1, 7)	11345.860219
60	(1, 1, 0)	11347.079730
61	(5, 1, 7)	11347.837227
62	(7, 1, 5)	11347.844436
63	(0, 1, 0)	11467.309902

64 rows × 2 columns

Once the function is done running, you should see that the order associated with the lowest AIC is (2,1,5). Therefore, this suggests are ARIMA model with an AR(2) process and a MA(5) process.

Now, we can print a summary of the best model, which an ARIMA (2,1,5).

```
                              SARIMAX Results
==============================================================================
Dep. Variable:                Mid Price   No. Observations:                2740
Model:                  ARIMA(2, 1, 5)   Log Likelihood             -10221.620
Date:                Tue, 27 Jun 2023   AIC                         20459.241
Time:                        12:09:06   BIC                         20506.563
Sample:                             0   HQIC                        20476.341
                               - 2740
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.4305      0.390     -1.105      0.269      -1.194       0.333
ar.L2          0.3023      0.408      0.741      0.459      -0.498       1.102
ma.L1         -0.2469      0.391     -0.631      0.528      -1.014       0.520
ma.L2         -0.6972      0.292     -2.386      0.017      -1.270      -0.124
ma.L3          0.0974      0.311      0.313      0.754      -0.512       0.706
ma.L4          0.0064      0.055      0.116      0.908      -0.102       0.115
ma.L5          0.0607      0.023      2.619      0.009       0.015       0.106
sigma2      4318.2955     81.730     52.836      0.000    4158.109    4478.482
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):             8617.48
Prob(Q):                              0.97   Prob(JB):                        0.00
Heteroskedasticity (H):               0.00   Skew:                            1.40
Prob(H) (two-sided):                  0.00   Kurtosis:                       11.23
===================================================================================
```