FERM534 - Applied Financial Econometrics II - Assigment I

Project by: N. Emir Eğilli

```
import yfinance as yf
import matplotlib.pyplot as plt
```

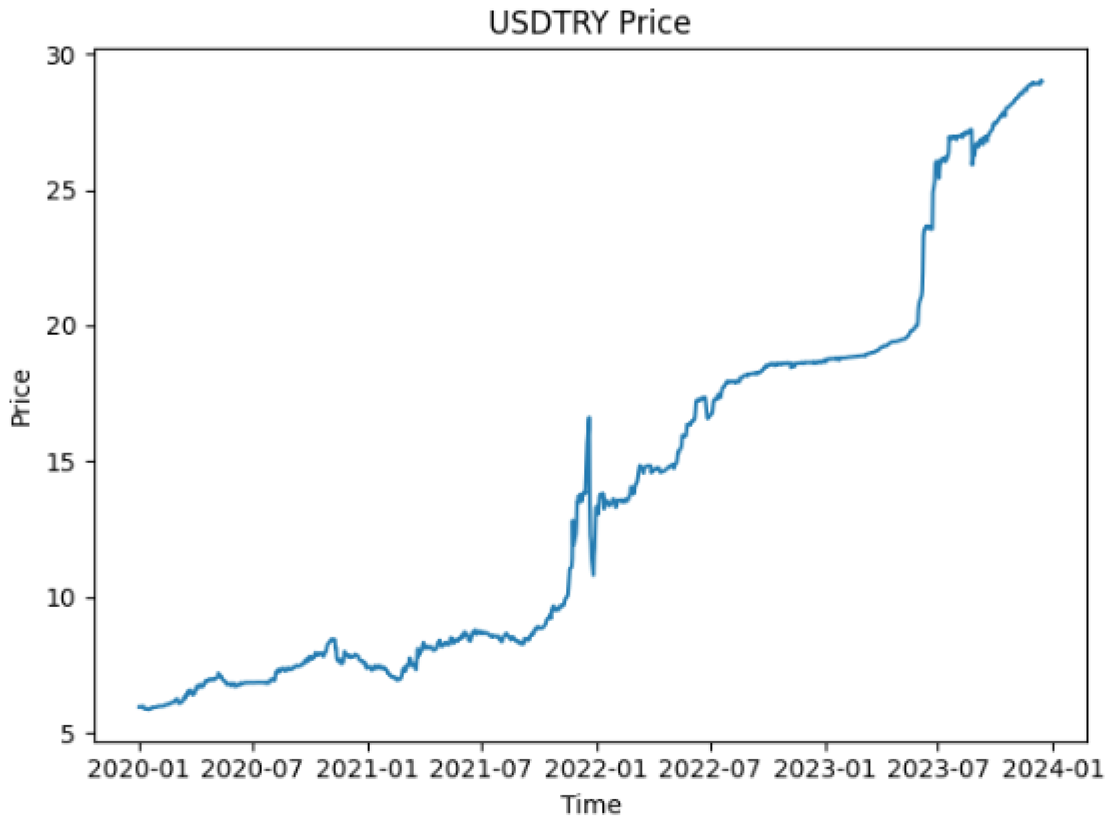
In [8]:

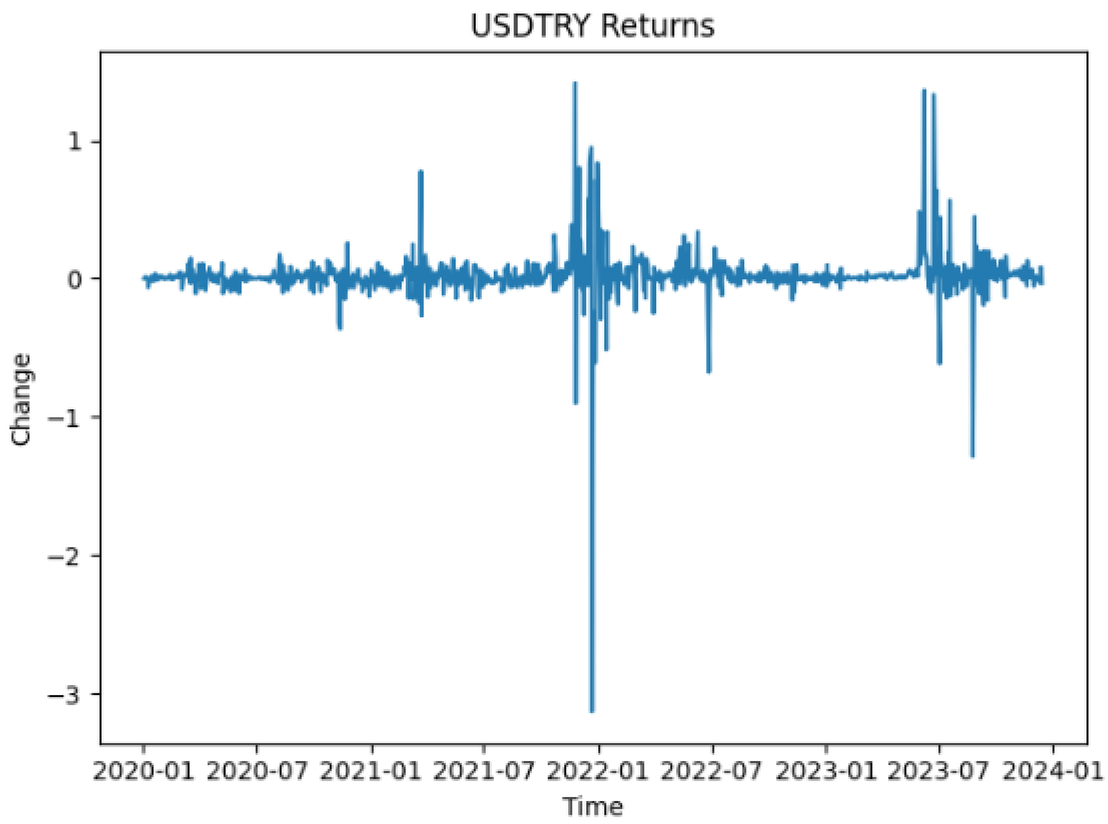
```
import warnings
warnings.filterwarnings("ignore")

DTRY = yf.download("TRY=X", start="2020-01-01", end = "2023-12-15")['Ad
j Close']
rDTRY = DTRY.diff().dropna()
```

```
plt.plot(DTRY)
plt.xlabel('Time')
plt.ylabel('Price')
plt.title('USDTRY Price')
plt.tight_layout()
plt.show()

plt.plot(rDTRY)
plt.xlabel('Time')
plt.ylabel('Change')
plt.title('USDTRY Returns')
plt.tight_layout()
plt.show()
```


USDTRY Price
Price
Time
30
25
20
15
10
5
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07
2024-01


USDTRY Returns
Change
Time
1
0
−1
−2
−3
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07
2024-01

1) Perform KPSS and Augmented Dickey-Fuller tests on DTRY and rDTRY series. Comment. Conclude.

In [6]:

```
from statsmodels.tsa.stattools import kpss
```

In [9]:

```
kpss1 = kpss(DTRY, regression='ct', nlags='auto', store=False)
print(kpss1)
```

```
(0.859378076711003, 0.01, 19, {'10%': 0.119, '5%': 0.146, '2.5%': 0.176, '1%': 0.216})
```

In [10]:

```
kpss1 = kpss(DTRY, regression='ct', nlags='legacy', store=False)
print(kpss1)
```

```
(0.7528188962469078, 0.01, 22, {'10%': 0.119, '5%': 0.146, '2.5%': 0.176, '1%': 0.216})
```

Since there's a trend in the price, we have used 'ct' in regression. For the P values, using both "auto" and "legacy" prompts less than 5 percent significance, H0 hypothesis can be rejected and HAlternative that data is not stationary is accepted, which means in their current form, they may not be suitable for certain statistical or predictive modeling purposes.

```
kpss2 = kpss(rDTRY, regression='c', nlags='legacy', store=False)
print(kpss2)
```

```
(0.3974160409056736, 0.07826894788548551, 22, {'10%': 0.347, '5%': 0.463, '2.5%': 0.574, '1%': 0.739})
```

In [12]:

```
kpss2 = kpss(rDTRY, regression='c', nlags='auto', store=False)
print(kpss2)
```

```
(0.5192309744043105, 0.037335366125155284, 0, {'10%': 0.347, '5%': 0.463, '2.5%': 0.574, '1%': 0.739})
```

Since there is no trend in returns, we used 'c' in regression. As the P value of rDTRY is higher than 5% in legacy. Null hypothesis cannot be rejected. As the data is stationary, it cannot be used in statistical and predictive modeling. However, when we use 'auto' this time the null hypothesis can be rejected and HAlternative is accepted. Data/series is not stationary and once again it cannot be used.

```python
from statsmodels.tsa.stattools import adfuller
```

```python
adf1 = adfuller(DTRY, regression='ct', autolag='AIC')
print(adf1)
```

```
(-1.7636247258288458, 0.7220281335007377, 21, 1010, {'1%': -3.967761463
4121485, '5%': -3.414845832687729, '10%': -3.1296138693838014}, -614.35
87515900551)
```

P value higher than 5%, data series is not stationary. We cannot use the data as is.

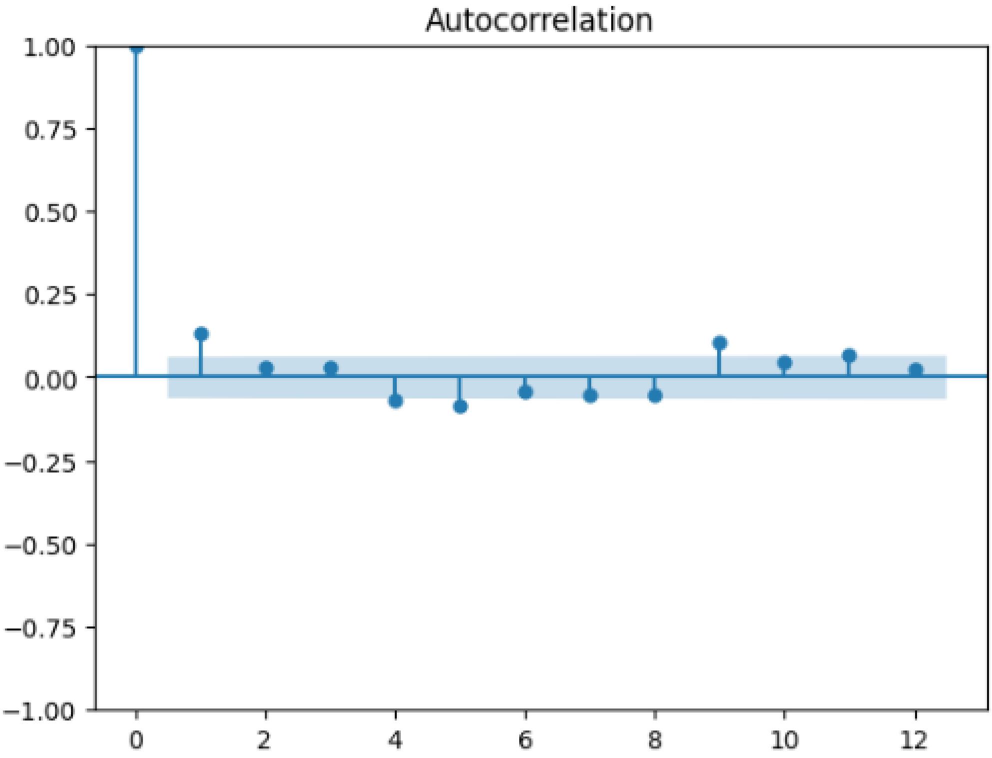
```python
adf2 = adfuller(rDTRY, regression='c', autolag='AIC')
print(adf2)
```

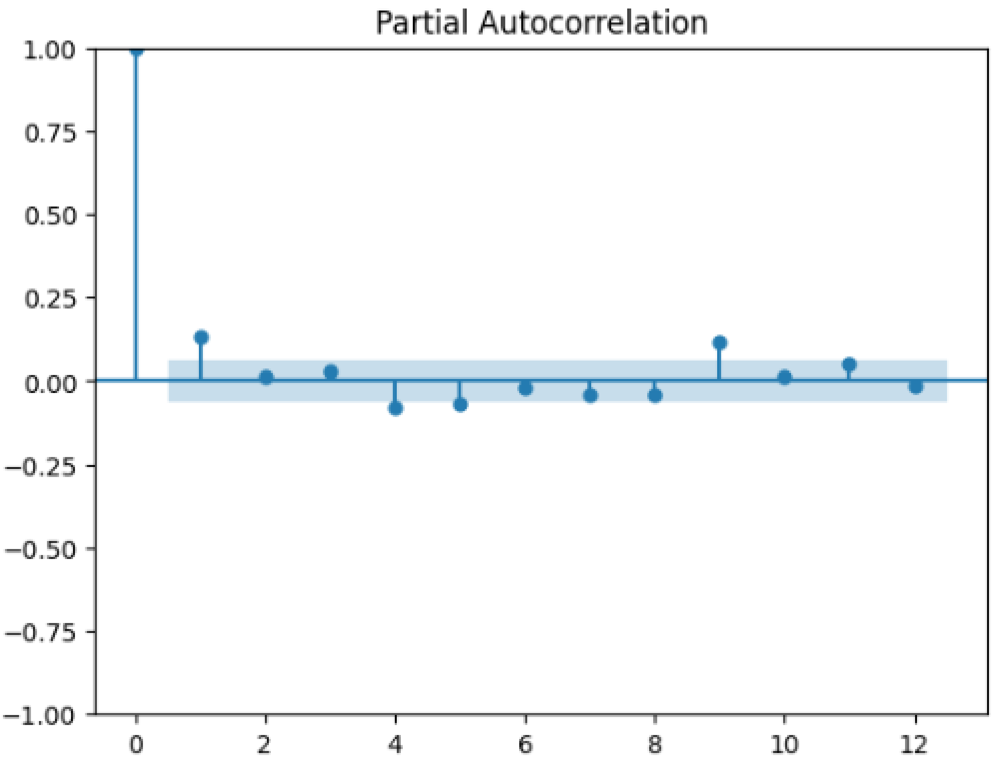
```
(-5.88621420046895, 2.997377187502574e-07, 22, 1008, {'1%': -3.43685399
89632336, '5%': -2.864411567265667, '10%': -2.568298955065823}, -614.63
11724018676)
```

Since there is no trend in returns, we used 'c' in regression. As the P value of rDTRY is less than 5%, null hypothesis can be rejected and HAlternative hypothesis that the series/data is stationary is accepted. We can use it in statistical and predictive modeling.

2) Plot the ACF and PACF for the rDTRY series setting the maximum lag to 12.

```
from statsmodels.graphics.tsaplots import plot_pacf, plot_acf
plot_acf(rDTRY, alpha=0.05, lags=12)
plot_pacf(rDTRY, alpha=0.05, lags=12)
```





Based on the partial autocorrelation plot showing a significant spike at lag 1 and the autocorrelation plot indicating a gradual decline for rDTRY, an AR(1) model is the most appropriate to try first. The absence of other significant spikes in the partial autocorrelation plot suggests that higher-order AR models are likely unnecessary.

3) For an AR model, which model orders would you try based on the plots you created above. Comment.

```
from statsmodels.tsa.ar_model import ar_select_order
mod = ar_select_order(rDTRY, maxlag=12, ic='aic', glob=True)
mod.ar_lags
[1, 4, 5, 8, 9, 11]
```

Based on the partial autocorrelation plot, the lags inferred above seems to be in line.

4) Estimate each AR model that you determined in question 3. Compare the models based on the Akaike's information criterion. Which model would you prefer. Justify your choice. Call your preferred model m1.

```
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.graphics.tsaplots import plot_predict
# AR(1)
ar1arima = ARIMA(rDTRY, order=(1, 0, 0)).fit()
print(ar1arima.summary())
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:              Adj Close   No. Observations:                 1031
Model:                 ARIMA(1, 0, 0)   Log Likelihood                 308.263
Date:                Mon, 25 Dec 2023   AIC                           -610.525
Time:                        09:23:17   BIC                           -595.710
Sample:                    01-02-2020   HQIC                          -604.903
                         - 12-14-2023
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0223      0.007      3.014      0.003       0.008       0.037
ar.L1          0.1331      0.009     15.232      0.000       0.116       0.150
sigma2         0.0322      0.000    140.905      0.000       0.032       0.033
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):            638976.10
Prob(Q):                              0.96   Prob(JB):                         0.00
Heteroskedasticity (H):               3.76   Skew:                            -4.95
Prob(H) (two-sided):                  0.00   Kurtosis:                       124.56
===================================================================================

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

```python
# AR(4)
ar4arima = ARIMA(rDTRY, order=(4, 0, 0)).fit()
print(ar4arima.summary())
```

```
                               SARIMAX Results                                
==============================================================================
Dep. Variable:              Adj Close   No. Observations:                 1031
Model:                 ARIMA(4, 0, 0)   Log Likelihood                 311.931
Date:                Mon, 25 Dec 2023   AIC                           -611.862
Time:                        09:23:19   BIC                           -582.233
Sample:                    01-02-2020   HQIC                          -600.618
                         - 12-14-2023                                         
Covariance Type:                  opg                                         
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0223      0.008      2.807      0.005       0.007       0.038
ar.L1          0.1333      0.009     15.109      0.000       0.116       0.151
ar.L2          0.0092      0.021      0.444      0.657      -0.031       0.050
ar.L3          0.0375      0.018      2.115      0.034       0.003       0.072
ar.L4         -0.0790      0.012     -6.549      0.000      -0.103      -0.055
sigma2         0.0320      0.000    106.148      0.000       0.031       0.033
===================================================================================
Ljung-Box (L1) (Q):                   0.03   Jarque-Bera (JB):            627086.13
Prob(Q):                              0.86   Prob(JB):                         0.00
Heteroskedasticity (H):               3.69   Skew:                            -4.94
Prob(H) (two-sided):                  0.00   Kurtosis:                       123.42
===================================================================================

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

```python
# AR(5)
ar5arima = ARIMA(rDTRY, order=(5, 0, 0)).fit()
print(ar5arima.summary())
```

```
                               SARIMAX Results                                
==============================================================================
Dep. Variable:              Adj Close   No. Observations:                 1031
Model:                 ARIMA(5, 0, 0)   Log Likelihood                 314.323
Date:                Mon, 25 Dec 2023   AIC                           -614.645
Time:                        09:23:20   BIC                           -580.077
Sample:                    01-02-2020   HQIC                          -601.526
                         - 12-14-2023                                         
Covariance Type:                  opg                                         
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0223      0.008      2.890      0.004       0.007       0.038
ar.L1          0.1278      0.009     14.312      0.000       0.110       0.145
ar.L2          0.0117      0.021      0.552      0.581      -0.030       0.053
ar.L3          0.0381      0.019      2.020      0.043       0.001       0.075
ar.L4         -0.0696      0.013     -5.340      0.000      -0.095      -0.044
ar.L5         -0.0679      0.015     -4.661      0.000      -0.096      -0.039
sigma2         0.0318      0.000    103.214      0.000       0.031       0.032
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):            649194.27
Prob(Q):                              0.97   Prob(JB):                         0.00
Heteroskedasticity (H):               3.68   Skew:                            -4.92
Prob(H) (two-sided):                  0.00   Kurtosis:                       125.54
===================================================================================

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

```python
# AR(8)
ar8arima = ARIMA(rDTRY, order=(8, 0, 0)).fit()
print(ar8arima.summary())
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:              Adj Close   No. Observations:                 1031
Model:                 ARIMA(8, 0, 0)   Log Likelihood                 316.133
Date:                Mon, 25 Dec 2023   AIC                           -612.266
Time:                        09:23:22   BIC                           -562.883
Sample:                    01-02-2020   HQIC                          -593.525
                         - 12-14-2023
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0224      0.007      3.019      0.003       0.008       0.037
ar.L1          0.1243      0.009     13.815      0.000       0.107       0.142
ar.L2          0.0074      0.023      0.314      0.753      -0.039       0.053
ar.L3          0.0334      0.020      1.666      0.096      -0.006       0.073
ar.L4         -0.0708      0.014     -5.149      0.000      -0.098      -0.044
ar.L5         -0.0636      0.014     -4.589      0.000      -0.091      -0.036
ar.L6         -0.0133      0.023     -0.567      0.571      -0.059       0.033
ar.L7         -0.0324      0.016     -1.966      0.049      -0.065   -9.28e-05
ar.L8         -0.0417      0.027     -1.543      0.123      -0.095       0.011
sigma2         0.0317      0.000     97.361      0.000       0.031       0.032
===================================================================================
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):            633059.81
Prob(Q):                              0.88   Prob(JB):                         0.00
Heteroskedasticity (H):               3.74   Skew:                            -4.83
Prob(H) (two-sided):                  0.00   Kurtosis:                       124.01
===================================================================================

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

```python
# AR(9)
ar9arima = ARIMA(rDTRY, order=(9, 0, 0)).fit()
print(ar9arima.summary())
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:              Adj Close   No. Observations:                 1031
Model:                 ARIMA(9, 0, 0)   Log Likelihood                 322.767
Date:                Mon, 25 Dec 2023   AIC                           -623.534
Time:                        09:23:24   BIC                           -569.213
Sample:                    01-02-2020   HQIC                          -602.919
                         - 12-14-2023
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0223      0.008      2.628      0.009       0.006       0.039
ar.L1          0.1290      0.009     14.454      0.000       0.112       0.147
ar.L2          0.0111      0.024      0.470      0.638      -0.035       0.057
ar.L3          0.0350      0.020      1.755      0.079      -0.004       0.074
ar.L4         -0.0636      0.015     -4.354      0.000      -0.092      -0.035
ar.L5         -0.0556      0.014     -3.947      0.000      -0.083      -0.028
ar.L6         -0.0170      0.024     -0.722      0.470      -0.063       0.029
ar.L7         -0.0332      0.018     -1.847      0.065      -0.068       0.002
ar.L8         -0.0556      0.035     -1.589      0.112      -0.124       0.013
ar.L9          0.1126      0.028      3.970      0.000       0.057       0.168
sigma2         0.0313      0.000     91.488      0.000       0.031       0.032
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):            630594.66
Prob(Q):                              0.96   Prob(JB):                         0.00
Heteroskedasticity (H):               3.70   Skew:                            -4.90
Prob(H) (two-sided):                  0.00   Kurtosis:                       123.76
===================================================================================

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

```python
# AR(11)
ar11arima = ARIMA(rDTRY, order=(11, 0, 0)).fit()
print(ar11arima.summary())
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:              Adj Close   No. Observations:                 1031
Model:                ARIMA(11, 0, 0)   Log Likelihood                 324.309
Date:                Mon, 25 Dec 2023   AIC                           -622.617
Time:                        09:23:25   BIC                           -558.419
Sample:                    01-02-2020   HQIC                          -598.254
                         - 12-14-2023
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0223      0.009      2.376      0.017       0.004       0.041
ar.L1          0.1266      0.009     13.800      0.000       0.109       0.145
ar.L2          0.0061      0.023      0.259      0.796      -0.040       0.052
ar.L3          0.0384      0.020      1.937      0.053      -0.000       0.077
ar.L4         -0.0616      0.015     -4.187      0.000      -0.090      -0.033
ar.L5         -0.0539      0.014     -3.810      0.000      -0.082      -0.026
ar.L6         -0.0132      0.024     -0.540      0.589      -0.061       0.035
ar.L7         -0.0304      0.018     -1.695      0.090      -0.066       0.005
ar.L8         -0.0577      0.036     -1.587      0.112      -0.129       0.014
ar.L9          0.1101      0.031      3.605      0.000       0.050       0.170
ar.L10         0.0080      0.027      0.299      0.765      -0.045       0.061
ar.L11         0.0523      0.015      3.448      0.001       0.023       0.082
sigma2         0.0312      0.000     87.760      0.000       0.031       0.032
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):             622824.91
Prob(Q):                              0.98   Prob(JB):                          0.00
Heteroskedasticity (H):               3.61   Skew:                             -4.96
Prob(H) (two-sided):                  0.00   Kurtosis:                        123.00
===================================================================================

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

Preferred Model: Based on AIC criterion, AR9's AIC value is shows the least value and AR9 is statistically significant. So we select "M1:AR9".

5) Perform Ljung-Box test to check the validity of the m1 model. Comment. Conclude.

```
from statsmodels.stats.diagnostic import acorr_ljungbox
ljungbox_results = acorr_ljungbox(ar9arima.resid, lags=[10], model_df=9
)
print(ljungbox_results)
     lb_stat  lb_pvalue
10  0.656851   0.417674
```

We conducted the Ljung-Box test on the residuals, which are essentially what remains after applying our model to the dataset. The purpose of this test is to verify if the residuals are devoid of any discernible structure, essentially resembling white noise. Such a result would indicate that our model has successfully captured the existing patterns in the dataset, demonstrating its effectiveness and efficiency.

In this context, the null hypothesis (H0) posits that the residuals are indeed white noise, while the HAlternative suggests the contrary. Given that the p-value from our test exceeds the 5% significance threshold, we cannot reject the null hypothesis. Consequently, we infer that the residuals display characteristics of white noise, affirming that our model comprehensively accounts for the structural attributes within the dataset. Therefore, we can conclude that our model is both effective and efficient in its design and application.

6) Refine M1 model by excluding all the insignificant variables. Call the refined model M2. Write down the refined model explicitly.

```
ar9arimaR = ARIMA(rDTRY, order=((1,0,0,1,1,0,0,0,1), 0, 0)).fit()
print(ar9arimaR.summary())
```

```
                               SARIMAX Results
==========================================================================================
Dep. Variable:                   Adj Close   No. Observations:                   1031
Model:               ARIMA([1, 4, 5, 9], 0, 0)   Log Likelihood                   318.990
Date:                     Mon, 25 Dec 2023   AIC                              -625.979
Time:                             09:23:33   BIC                              -596.349
Sample:                         01-02-2020   HQIC                             -614.734
                              - 12-14-2023
Covariance Type:                       opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0223      0.008      2.912      0.004       0.007       0.037
ar.L1          0.1360      0.009     14.943      0.000       0.118       0.154
ar.L4         -0.0568      0.012     -4.674      0.000      -0.081      -0.033
ar.L5         -0.0603      0.015     -4.076      0.000      -0.089      -0.031
ar.L9          0.1024      0.023      4.394      0.000       0.057       0.148
sigma2         0.0315      0.000    119.543      0.000       0.031       0.032
===================================================================================
Ljung-Box (L1) (Q):                   0.03   Jarque-Bera (JB):            614415.91
Prob(Q):                              0.87   Prob(JB):                         0.00
Heteroskedasticity (H):               3.66   Skew:                            -4.85
Prob(H) (two-sided):                  0.00   Kurtosis:                       122.20
===================================================================================

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

Using Auto ARIMA up to lag 9, we allocated value 1 to significant lags (lags of 1, 4, 5 and 9) and 0 to insignificant lags. Significant lags will be M2 model. We have our formula as below:

$$Y_t - u = Constant + \beta_1 * (Y_{t-1} - u) + \beta_4 * (Y_{t-4} - u) + \beta_5 * (Y_{t-5} - u) + \beta_9 * (Y_{t-9} - u)$$

In this model, 'u' or 'Constant' is 0.0223, and the coeffiecients of betas are 0.1360, -0.0568, -0.0603 and 0.1024 respectively. Each coefficient corresponds to the influence of its respective lag on the current value of the series. This model structure captures the significant autocorrelations at the identified lags, providing a tailored approach to forecasting the time series.

7) Perform Ljung-Box test to check the validity of m2 model. Comment.Conclude

```
lb_test = acorr_ljungbox(ar9arimaR.resid, lags=[10], model_df=4)
print(lb_test)
     lb_stat   lb_pvalue
10  7.29119    0.294756
```

In this scenario, where the p-value is 0.294, exceeding the 5% significance threshold, we are unable to reject the null hypothesis (H0) asserting that the data represents white noise. This outcome implies that our model has sufficiently captured the underlying structures within the dataset. Consequently, we can confidently conclude that the model is both effective and efficient in its representation and analysis of the data. This conclusion is reinforced by the fact that the p-value significantly surpasses the benchmark for statistical significance, underscoring the model's robustness in accounting for the data's inherent patterns and characteristics.

---

8) Identify an alternative ARIMA model by using the auto ARIMA module. Call the model m3. Estimate the model.

```
from pmdarima.arima import auto_arima
auto_arima(rDTRY)
```

ARIMA

```
 ARIMA(3,0,3)(0,0,0)[0] intercept
```

We will call ARIMA(3, 0, 3) for M3. Since there's an intercept, there is an upwards drift.

```
ar33arima = ARIMA(rDTRY, order=(3, 0, 3)).fit()
print(ar33arima.summary())
```

```
                               SARIMAX Results                                
==============================================================================
Dep. Variable:              Adj Close   No. Observations:                 1031
Model:                 ARIMA(3, 0, 3)   Log Likelihood                 318.873
Date:                Mon, 25 Dec 2023   AIC                           -621.746
Time:                        09:23:57   BIC                           -582.240
Sample:                    01-02-2020   HQIC                          -606.754
                         - 12-14-2023                                         
Covariance Type:                  opg                                         
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0223      0.007      3.078      0.002       0.008       0.037
ar.L1          0.6299      0.088      7.158      0.000       0.457       0.802
ar.L2          0.3612      0.124      2.918      0.004       0.119       0.604
ar.L3         -0.7342      0.057    -12.854      0.000      -0.846      -0.622
ma.L1         -0.5051      0.088     -5.727      0.000      -0.678      -0.332
ma.L2         -0.4133      0.110     -3.774      0.000      -0.628      -0.199
ma.L3          0.7256      0.049     14.738      0.000       0.629       0.822
sigma2         0.0315      0.000    100.890      0.000       0.031       0.032
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):            617894.06
Prob(Q):                              0.96   Prob(JB):                         0.00
Heteroskedasticity (H):               3.64   Skew:                            -4.87
Prob(H) (two-sided):                  0.00   Kurtosis:                       122.53
===================================================================================

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

All lags show statistical significance.

---

9) Perform Ljung-Box test to check the validity of model m3.Comment.Conclude.

```
lb_test = acorr_ljungbox(ar33arima.resid, lags=[10], model_df=6)
print(lb_test)
      lb_stat  lb_pvalue
10  11.613589   0.020468
```

As the p value (0.02) is less than 5% significance level, we reject the null hypothesis and accept HAlternative that the data has structure and is not a white noise. Therefore, we can conclude that the model does not adequately cover the structures in the data and is not an efficient model.

---

10) Compare models m1, m2, m3. Which model you prefer? Comment and Justify your choice.

Model M3, as previously discussed, falls short of being an effective and efficient choice due to the reasons outlined earlier. When comparing models M1 and M2, it's notable that their respective Akaike Information Criterion (AIC) values are -623.534 and -625.979.

Given that a lower AIC value is preferable as it indicates a better fit with fewer complexities, Model M2 emerges as the more suitable option. This preference is further justified by the fact that M2 is a refined version of the AutoRegressive model with nine lags (AR9), where insignificant lags have been prudently excluded to enhance its predictive accuracy and efficiency.
