Financial Econometrics - Homework 1 - José Barretto, Daniel Deutsch, Stéphane Roblet

September 19, 2021

1 Empirical Application 1 Financial Econometrics

By Daniel Deutsch, José Lucas Barretto, and Stéphane Roblet

[1]:
```
import itertools
import warnings

import matplotlib.pyplot as plt
import pandas as pd
import statsmodels.tsa.stattools as sm
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.regression.linear_model import OLS
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.tsa.seasonal import seasonal_decompose
```

[2]:
```
# Ignore warnings
warnings.filterwarnings('ignore')

# Matplotlib styles
plt.style.use('ggplot')
plt.rcParams.update({
    'figure.figsize': (15, 4),
    'axes.prop_cycle': plt.cycler(color=["#4C72B0", "#C44E52", "#55A868",␣
 ↪"#8172B2", "#CCB974", "#64B5CD"]),
    'axes.facecolor': "#EAEAF2"
})
```

2 Motivation

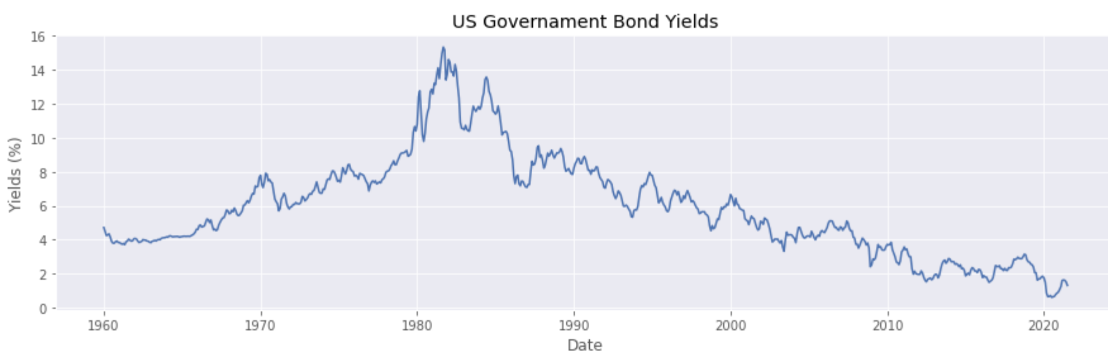
We have chosen these datasets for different reasons. First, we wanted to investigate the price evolution for financial assets when the issuers differ in nature. Thus, we chose to look at the evolution of the price of the New Stock Exchange and government bonds. In addition, we wanted to know if different trends could emerge depending on the nature of the financial asset of the same type of issuer (in this case a company), which is why we were also interested in the evolution of the valuation of AAA corporate bonds in addition to that of shares of companies listed on the NYSE.

3 US Governament Bond Yields

3.1 Load the Data

Notice that here we don't apply the log-transform in our data since it is already given in percentage.

[3]:
```
# Loads data and drops non available values
df_govbonds = pd.read_csv("datasets/US_govbonds_yields.csv", names=['date',␣
↪'value'], parse_dates=['date'], skiprows=[0], na_values='.')
df_govbonds.dropna(inplace=True)

# Plots the time series
plt.plot(df_govbonds['date'], df_govbonds['value'])
plt.title("US Governament Bond Yields")
plt.xlabel("Date")
plt.ylabel("Yields (%)")
plt.show()
```



3.2 Augmented Dickey-Fuller Test

The Augmented Dickey-Fuller test is a unit root test that checks for stationarity. It considers the following hypotesis:

$$\begin{aligned} H_0: &\quad \text{there is a unit root (the series contains a stochastic trend and is non-stationary)} \\ H_1: &\quad \text{there isn't a unit root (the series doesn't contain a stochastic trend and is stationary)} \end{aligned}$$

Once the Augmented Dickey-Fuller Test is performed and we have obtained our results, we should consider the following to take our conclusions:

- If the p-value is lower than 0.05, than we must reject the null hypotesys (and, consequently, accept the alternativel one).
- If the p-value is slightly above 0.05, then the critical values should be used to judge whether to reject the null hypotesis.

First, we are going to implement the augmented Dickey-Fuller Test to the most general regression:

$$\Delta X_t = b_0 + b_1 t + \rho X_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta_{t-j} + \varepsilon_t$$

In this case, the hypotesis can be written as:

$$\begin{aligned} H_0 : &\quad \rho = 0 \\ H_1 : &\quad \rho < 0 \end{aligned}$$

[4]:
```
# Obtains the ADF results for the general regression
adf_res = sm.adfuller(
    df_govbonds['value'],
    regression='ct',
    maxlag=12,
    autolag='AIC',
    regresults=True
)

# Show test results
print(f"T-statistic: {adf_res[0]}")
print(f"P-value: {adf_res[1]}")
print(f"Used lag: {adf_res[3].usedlag}")
```

```
T-statistic: -2.1596201496468836
P-value: 0.5125920716411859
Used lag: 12
```

From the results printed above we can conclude that **the null hypotesis shouldn't be rejected**, i.e., there isn't a unit root in the time series (it is non-stationary), since the p-value is way higher than 0.05. One could reach the same conclusion by observing the value of the T-statistic (since it's absolute value is lower than the absolute value of -3.45).

Once we didn't reject the null hypotesis, we should test wether the coefficient b_1 is significant or not.

[5]:
```
print(adf_res[3].resols.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.174
```

```
Model:                            OLS   Adj. R-squared:                  0.158
Method:                 Least Squares   F-statistic:                     10.69
Date:                Sun, 19 Sep 2021   Prob (F-statistic):           2.90e-22
Time:                        21:05:46   Log-Likelihood:                -27.423
No. Observations:                 726   AIC:                             84.85
Df Residuals:                     711   BIC:                             153.7
Df Model:                          14
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1            -0.0081      0.004     -2.160      0.031      -0.015      -0.001
x2             0.3916      0.037     10.504      0.000       0.318       0.465
x3            -0.2393      0.040     -5.993      0.000      -0.318      -0.161
x4             0.1005      0.041      2.455      0.014       0.020       0.181
x5            -0.0675      0.041     -1.643      0.101      -0.148       0.013
x6             0.0957      0.041      2.325      0.020       0.015       0.176
x7            -0.0728      0.041     -1.765      0.078      -0.154       0.008
x8            -0.0486      0.041     -1.178      0.239      -0.130       0.032
x9             0.0780      0.041      1.896      0.058      -0.003       0.159
x10            0.0042      0.041      0.101      0.919      -0.077       0.085
x11            0.0278      0.041      0.679      0.497      -0.053       0.108
x12            0.0904      0.040      2.264      0.024       0.012       0.169
x13           -0.0724      0.037     -1.938      0.053      -0.146       0.001
const          0.0860      0.037      2.348      0.019       0.014       0.158
x14           -0.0001   5.19e-05     -2.141      0.033      -0.000   -9.21e-06
==============================================================================
Omnibus:                       77.089   Durbin-Watson:                   2.000
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              510.423
Skew:                           0.146   Prob(JB):                    1.46e-111
Kurtosis:                       7.097   Cond. No.                     2.98e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.98e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

In the print above, the coefficient b_1 is represented by the variable x14 (which can only be seen once you click at "*show more (open the raw output data in a text editor) ...*"). We observe that its T-statistic is -2.141. Since the absolute value of its T-statistic is inferior than the T-statistic of 2.78 at the 5% confidence level, we conclude that **the trend coefficient b_1 is not statistically significant**.

Now, since we concluded that b_1 isn't statistically significant, we shall perform the ADF test again, but this time considering a restricted regression, as following:

$$\Delta X_t = b_0 + \rho X_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta_{t-j} + \varepsilon_t$$

Again, our hypotesis can be written as:

$$\begin{aligned} H_0 &: \quad \rho = 0 \\ H_1 &: \quad \rho < 0 \end{aligned}$$

[5]:
```
# Obtains the ADF results for the restricted regression
adf_res = sm.adfuller(
    df_govbonds['value'],
    regression='c',
    maxlag=12,
    autolag='AIC',
    regresults=True
)

# Show test results
print(f"T-statistic: {adf_res[0]}")
print(f"P-value: {adf_res[1]}")
print(f"Used lag: {adf_res[3].usedlag}")
```

```
T-statistic: -1.294841876108271
P-value: 0.6315288451340434
Used lag: 12
```

Once more, from the results printed above, we can conclude that **the null hypotesis shouldn't be rejected**, i.e., there isn't a unit root in the time series (it is non-stationary), since the p-value is way higher than 0.05. The same conclusion could be reached by observing the value of the T-statistic (since it's absolute value is lower than the absolute value of -2.89).

Once we didn't reject the null hypotesis, we should test wether the coefficient b_0 is significant or not.

[6]:
```
print(adf_res[3].resols.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.169
Model:                            OLS   Adj. R-squared:                  0.153
Method:                 Least Squares   F-statistic:                     11.10
Date:                Sun, 19 Sep 2021   Prob (F-statistic):           6.93e-22
Time:                        16:39:26   Log-Likelihood:                -29.756
No. Observations:                 726   AIC:                             87.51
Df Residuals:                     712   BIC:                             151.7
```

```
Df Model:                          13
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1            -0.0043      0.003     -1.295      0.196      -0.011       0.002
x2             0.3940      0.037     10.550      0.000       0.321       0.467
x3            -0.2394      0.040     -5.979      0.000      -0.318      -0.161
x4             0.1021      0.041      2.489      0.013       0.022       0.183
x5            -0.0666      0.041     -1.615      0.107      -0.147       0.014
x6             0.0973      0.041      2.359      0.019       0.016       0.178
x7            -0.0719      0.041     -1.738      0.083      -0.153       0.009
x8            -0.0473      0.041     -1.144      0.253      -0.128       0.034
x9             0.0795      0.041      1.929      0.054      -0.001       0.160
x10            0.0052      0.041      0.127      0.899      -0.076       0.086
x11            0.0291      0.041      0.707      0.480      -0.052       0.110
x12            0.0911      0.040      2.276      0.023       0.013       0.170
x13           -0.0721      0.037     -1.924      0.055      -0.146       0.001
const          0.0230      0.022      1.053      0.293      -0.020       0.066
==============================================================================
Omnibus:                       74.986   Durbin-Watson:                   1.999
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              508.396
Skew:                           0.064   Prob(JB):                    4.01e-111
Kurtosis:                       7.098   Cond. No.                         47.6
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

In the print above, the coefficient b_0 is represented by the variable const (which can only be seen once you click at "*show more (open the raw output data in a text editor) ...*"). We observe that its T-statistic is 1.053. Since the absolute value of its T-statistic is inferior than the T-statistic of 2.54 at the 5% confidence level, we conclude that **the drift coefficient b_0 is not statistically significant**.

Now, since we concluded that b_0 isn't statistically significant, we shall perform the ADF test again, but this time considering an even more restricted regression, as following:

$$\Delta X_t = \rho X_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta_{t-j} + \varepsilon_t$$

Again, our hypotesis can be written as:

$$\begin{aligned} H_0 : &\quad \rho = 0 \\ H_1 : &\quad \rho < 0 \end{aligned}$$

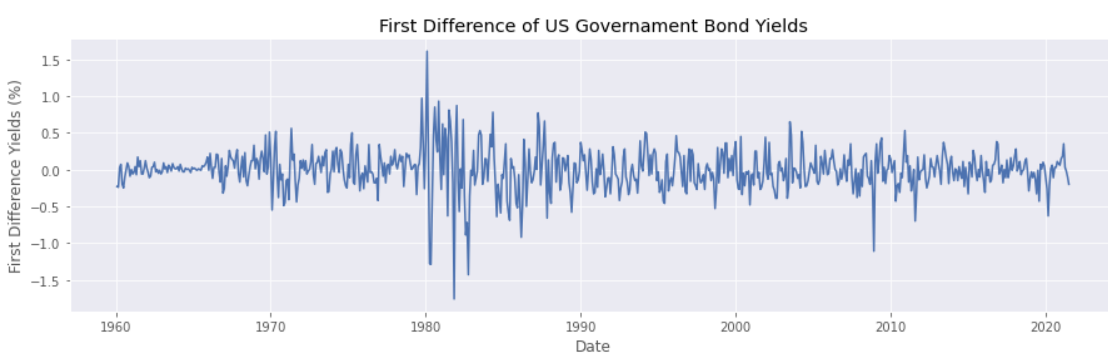
[7]:
```python
# Obtains the ADF results for the restricted regression
adf_res = sm.adfuller(
    df_govbonds['value'],
    regression='nc',
    maxlag=12,
    autolag='AIC',
    regresults=True
)

# Show test results
print(f"T-statistic: {adf_res[0]}")
print(f"P-value: {adf_res[1]}")
print(f"Used lag: {adf_res[3].usedlag}")
```

```
T-statistic: -0.7996081095853742
P-value: 0.3706833546535423
Used lag: 12
```

Once more, from the results printed above, we can conclude that **the null hypotesis shouldn't be rejected**, i.e., there isn't a unit root in the time series (it is non-stationary), since the p-value is way higher than 0.05. The same conclusion could be reached by observing the value of the T-statistic (since it's absolute value is lower than the absolute value of -1.95).

Once we didn't reject the null hypotesis, we finally conclude that **the series is non-stationary without constant**. We can verify this result by running the Augmented Dickey-Fuller test on the first difference of the series:

[8]:
```python
# Obtains the first difference of the time series
df_govbonds['diff'] = df_govbonds['value'].diff()

# Plots the first difference time series
plt.plot(df_govbonds['date'], df_govbonds['diff'])
plt.title("First Difference of US Governament Bond Yields")
plt.xlabel("Date")
plt.ylabel("First Difference Yields (%)")
plt.show()
```



[9]:
```python
# Obtains the ADF results for the restricted regression
adf_res = sm.adfuller(
    df_govbonds['diff'].dropna(),
    regression='ct',
    maxlag=12,
    autolag='AIC',
    regresults=True
)

# Show test results
print(f"T-statistic: {adf_res[0]}")
print(f"P-value: {adf_res[1]}")
print(f"Used lag: {adf_res[3].usedlag}")
```

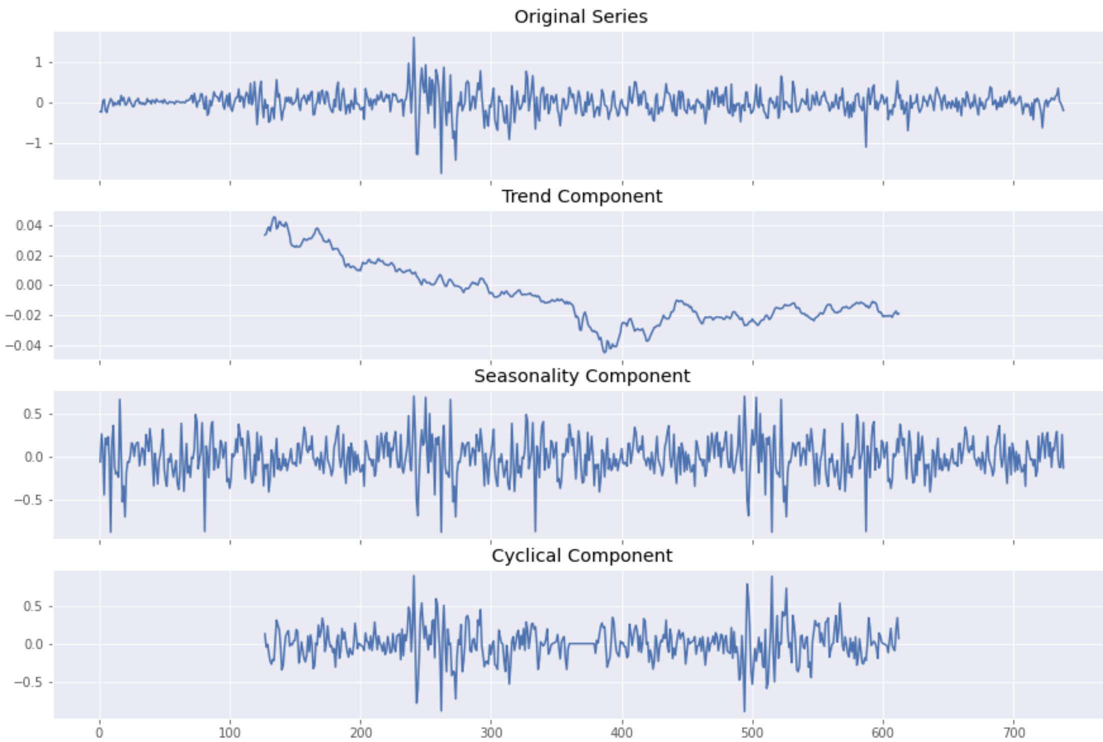
```
T-statistic: -7.363656510594449
P-value: 2.0960706199430693e-09
Used lag: 11
```

From the results printed above, we can conclude that **the null hypotesis should be rejected**, i.e., there is a unit root in the time series (it is stationary), since the p-value is way lower than 0.05. This means that US Governament Bond Yields are I(1).

3.3 Decomposition and Analysis of Drift, Trend, and Seasonality for the First Difference Series

Now, we shall decompose the first difference series into trend, seasonality, and residues (cyclical component). We will assume that the series follow an annual period, and thus, we consider a period of 253 days (average number of trading days per year).

[10]:
```python
# Decomposes the time series
decomposition = seasonal_decompose(df_govbonds['diff'].dropna(),␣
 ↪model='additive', period=253)

# Plot
fig, axs = plt.subplots(4, 1, sharex=True, figsize=(15, 10))
df_govbonds['diff'].plot(ax=axs[0])
decomposition.trend.plot(ax=axs[1])
decomposition.seasonal.plot(ax=axs[2])
decomposition.resid.plot(ax=axs[3])
axs[0].set_title("Original Series")
axs[1].set_title("Trend Component")
axs[2].set_title("Seasonality Component")
axs[3].set_title("Cyclical Component")
plt.show()
```



3.3.1 Deterministic Trend

[11]:
```
model = OLS(df_govbonds['diff'].dropna(), decomposition.trend, missing='drop')
print(model.fit().summary())
```

```
                                 OLS Regression Results
=======================================================================================
=======
Dep. Variable:                   diff   R-squared (uncentered):
0.004
Model:                            OLS   Adj. R-squared (uncentered):
0.001
Method:                 Least Squares   F-statistic:
1.728
Date:                Sun, 19 Sep 2021   Prob (F-statistic):
0.189
Time:                        16:39:27   Log-Likelihood:
-130.86
No. Observations:                 486   AIC:
263.7
Df Residuals:                     485   BIC:
267.9
```

```
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
trend          0.9128      0.694      1.314      0.189      -0.452       2.277
==============================================================================
Omnibus:                       72.329   Durbin-Watson:                   1.379
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              485.597
Skew:                          -0.406   Prob(JB):                    3.58e-106
Kurtosis:                       7.829   Cond. No.                         1.00
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not
contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

From the print above we can see that the trend component isn't statistically significant at the level 5%. This result indicates that the series doesn't have a deterministic trend when considering an annual period.

3.3.2 Seasonality

[12]:

```
model = OLS(df_govbonds['diff'].dropna(), decomposition.seasonal,␣
 ↪missing='drop')
print(model.fit().summary())
```

```
                                 OLS Regression Results
================================================================================
=======
Dep. Variable:                   diff   R-squared (uncentered):
0.317
Model:                            OLS   Adj. R-squared (uncentered):
0.316
Method:                 Least Squares   F-statistic:
341.6
Date:                Sun, 19 Sep 2021   Prob (F-statistic):
5.79e-63
Time:                        16:39:27   Log-Likelihood:
46.544
No. Observations:                 738   AIC:
-91.09
Df Residuals:                     737   BIC:
-86.48
Df Model:                           1
```

```
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
seasonal       0.6861      0.037     18.484      0.000       0.613       0.759
==============================================================================
Omnibus:                       59.605   Durbin-Watson:                   1.327
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              229.872
Skew:                          -0.262   Prob(JB):                     1.21e-50
Kurtosis:                       5.683   Cond. No.                         1.00
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not
contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

From the print above we can see that the seasonal component is statistically significant at the level 5%. This result indicates that the series has seasonality when considering an annual period.

We can also verify this by using dummy variables for each month and verifying their statistical significance.

[13]:
```python
# Creates dummy vars columns for week of the year
df_govbonds['period'] = df_govbonds['date'].dt.weekofyear

# Create a column for each dummy var
for period in df_govbonds['period'].unique():
    df_govbonds[f'is_{period}'] = 1*(df_govbonds['period'] == period)

# Include all dummy vars except one
X = df_govbonds[[f'is_{period}' for period in sorted(df_govbonds['period'].
 ↪unique())[:-1]]]

# Run regression
model = OLS(df_govbonds['diff'], X, missing='drop')
print(model.fit().summary())
```

```
                                 OLS Regression Results
================================================================================
=======
Dep. Variable:                   diff   R-squared (uncentered):
0.022
Model:                            OLS   Adj. R-squared (uncentered):
-0.009
Method:                 Least Squares   F-statistic:
0.7028
Date:                Sun, 19 Sep 2021   Prob (F-statistic):
```

```
0.846
Time:                        16:39:27   Log-Likelihood:
-85.748
No. Observations:                 738   AIC:
217.5
Df Residuals:                     715   BIC:
323.4
Df Model:                          23
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
is_1          -0.0017      0.047     -0.037      0.971      -0.093       0.090
is_5           0.0326      0.035      0.929      0.353      -0.036       0.101
is_9           0.0193      0.036      0.542      0.588      -0.051       0.089
is_10         -0.1550      0.195     -0.794      0.428      -0.538       0.228
is_13          0.0332      0.047      0.702      0.483      -0.060       0.126
is_14          0.0189      0.052      0.363      0.717      -0.084       0.121
is_17          0.0350      0.069      0.507      0.612      -0.101       0.171
is_18          0.0570      0.041      1.399      0.162      -0.023       0.137
is_22         -0.0467      0.039     -1.207      0.228      -0.123       0.029
is_23         -0.0545      0.083     -0.655      0.513      -0.218       0.109
is_26          0.0088      0.047      0.186      0.852      -0.084       0.102
is_27          0.0068      0.052      0.130      0.897      -0.096       0.109
is_30         -0.2133      0.113     -1.893      0.059      -0.435       0.008
is_31          0.0024      0.037      0.063      0.949      -0.071       0.075
is_35         -0.0652      0.043     -1.531      0.126      -0.149       0.018
is_36          0.0289      0.063      0.457      0.648      -0.095       0.153
is_39         -0.0350      0.056     -0.621      0.535      -0.146       0.076
is_40         -0.0376      0.045     -0.828      0.408      -0.127       0.052
is_44         -0.0412      0.036     -1.146      0.252      -0.112       0.029
is_45         -0.0150      0.195     -0.077      0.939      -0.398       0.368
is_48         -0.0033      0.043     -0.078      0.938      -0.087       0.080
is_49         -0.0653      0.063     -1.030      0.303      -0.190       0.059
is_52          0.0047      0.071      0.065      0.948      -0.135       0.145
==============================================================================
Omnibus:                      128.387   Durbin-Watson:                   1.404
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1245.605
Skew:                          -0.449   Prob(JB):                    3.31e-271
Kurtosis:                       9.301   Cond. No.                         5.57
==============================================================================
```

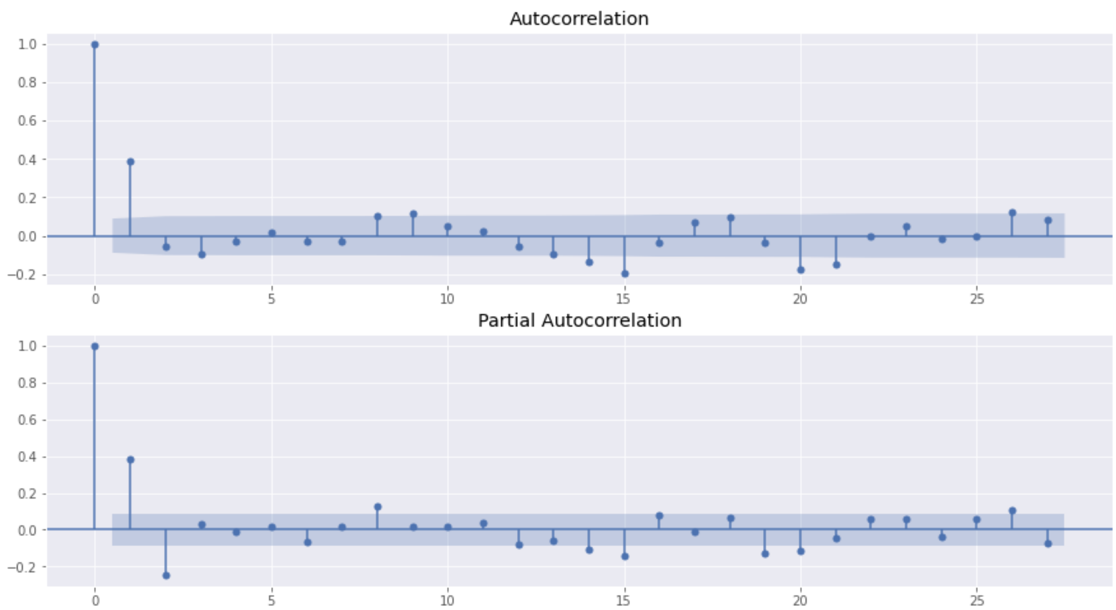
Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

The results indicate that some weeks of the year have statistically significant impacts on the value of the first difference series. This supports the previous finding that the series has a significant seasonal component.

3.4 Stationary ARMA Model

The goal here is to estimate the cyclical component through a stationary ARMA model. Firstly, we can plot the Autocorrelation and the Partial Autocorrelation of the time series to have an idea of the parameters of the model.

[14]:
```
# Plot
fig, axs = plt.subplots(2, 1, figsize=(15, 8))
plot_acf(decomposition.resid.dropna(), ax=axs[0])
plot_pacf(decomposition.resid.dropna(), ax=axs[1])
plt.show()
```



To obtain the best parameters for the ARMA model we run a cross-validation using the AIC score (select the model with the lowest AIC).

[15]:
```
best_model = None
for p, q in itertools.product(range(10), range(10)):

    print(f"\rCurrent {p=}, {q=}", end="")

    model = ARIMA(decomposition.resid.dropna(), order=(p, 0, q))
    model_fit = model.fit()

```

```python
    if p == 0 and q == 0:
        best_model = model_fit

    if model_fit.aic < best_model.aic:
        best_model = model_fit

print(best_model.summary())
```

```
Current p=9, q=9                                               SARIMAX Results
==============================================================================
Dep. Variable:                  resid   No. Observations:                  486
Model:                 ARIMA(4, 0, 9)   Log Likelihood                 125.189
Date:                Sun, 19 Sep 2021   AIC                           -220.378
Time:                        16:44:15   BIC                           -157.585
Sample:                             0   HQIC                          -195.708
                                - 486
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0072      0.012     -0.587      0.557      -0.031       0.017
ar.L1          0.0366      0.094      0.388      0.698      -0.148       0.221
ar.L2          0.3762      0.078      4.799      0.000       0.223       0.530
ar.L3         -0.2875      0.068     -4.214      0.000      -0.421      -0.154
ar.L4         -0.7768      0.088     -8.811      0.000      -0.950      -0.604
ma.L1          0.4193      0.101      4.137      0.000       0.221       0.618
ma.L2         -0.4006      0.074     -5.448      0.000      -0.545      -0.256
ma.L3          0.0834      0.082      1.016      0.310      -0.078       0.244
ma.L4          0.9305      0.075     12.487      0.000       0.784       1.077
ma.L5          0.3889      0.076      5.102      0.000       0.240       0.538
ma.L6          0.0357      0.050      0.708      0.479      -0.063       0.135
ma.L7         -0.0653      0.048     -1.375      0.169      -0.158       0.028
ma.L8         -0.0412      0.042     -0.982      0.326      -0.124       0.041
ma.L9          0.1229      0.042      2.931      0.003       0.041       0.205
sigma2         0.0333      0.002     17.078      0.000       0.030       0.037
================================================================================
===
Ljung-Box (L1) (Q):                   0.37   Jarque-Bera (JB):
116.61
Prob(Q):                              0.54   Prob(JB):
0.00
Heteroskedasticity (H):               1.07   Skew:
0.01
Prob(H) (two-sided):                  0.66   Kurtosis:
5.40
================================================================================
===
```

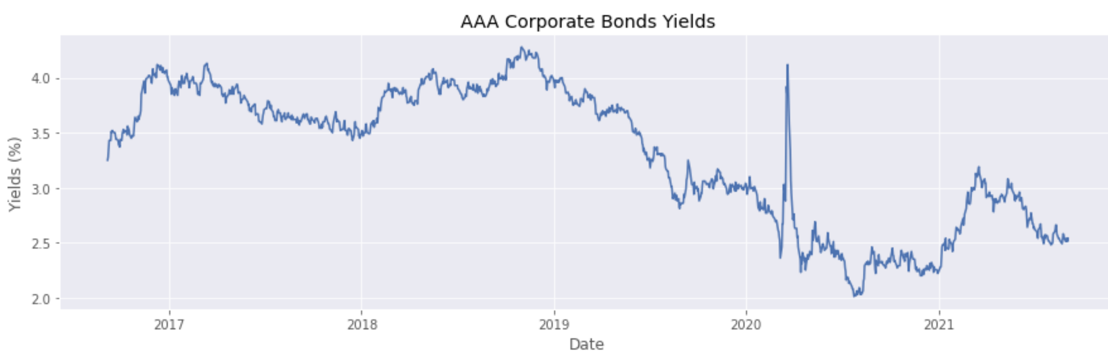
```
Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-
step).
```

4 Moody's AAA Corporate Bond Yields

4.1 Load the Data

Notice that here we don't apply the log-transform in our data since it is already given in percentage.

[16]:
```
# Loads data and drops non available values
df_aaa = pd.read_csv("datasets/AAA_corpbonds_yields.csv", names=['date',␣
 ↪'value'], parse_dates=['date'], skiprows=[0], na_values='.')
df_aaa.dropna(inplace=True)

# Plots the time series
plt.plot(df_aaa['date'], df_aaa['value'])
plt.title("AAA Corporate Bonds Yields")
plt.xlabel("Date")
plt.ylabel("Yields (%)")
plt.show()
```



4.2 Augmented Dickey-Fuller Test

First, we are going to implement the augmented Dickey-Fuller Test to the most general regression:

$$\Delta X_t = b_0 + b_1 t + \rho X_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta_{t-j} + \varepsilon_t$$

In this case, the hypotesis can be written as:

$$H_0: \quad \rho = 0$$
$$H_1: \quad \rho < 0$$

[17]:
```python
# Obtains the ADF results for the general regression
adf_res = sm.adfuller(
    df_aaa['value'],
    regression='ct',
    maxlag=12,
    autolag='AIC',
    regresults=True
)

# Show test results
print(f"T-statistic: {adf_res[0]}")
print(f"P-value: {adf_res[1]}")
print(f"Used lag: {adf_res[3].usedlag}\n")
print(adf_res[3].resols.summary())
```

```
T-statistic: -2.658179551223335
P-value: 0.25385353029949714
Used lag: 9

                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.046
Model:                            OLS   Adj. R-squared:                  0.037
Method:                 Least Squares   F-statistic:                     5.337
Date:                Sun, 19 Sep 2021   Prob (F-statistic):           2.63e-08
Time:                        16:44:16   Log-Likelihood:                 1955.7
No. Observations:                1241   AIC:                            -3887.
Df Residuals:                    1229   BIC:                            -3826.
Df Model:                          11
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1            -0.0105      0.004     -2.658      0.008      -0.018      -0.003
x2             0.0113      0.028      0.397      0.691      -0.044       0.067
x3             0.0911      0.028      3.208      0.001       0.035       0.147
x4             0.0729      0.028      2.567      0.010       0.017       0.129
x5            -0.0949      0.028     -3.329      0.001      -0.151      -0.039
x6             0.0284      0.029      0.994      0.320      -0.028       0.085
x7             0.0182      0.028      0.638      0.524      -0.038       0.074
x8            -0.0763      0.028     -2.688      0.007      -0.132      -0.021
x9            -0.0229      0.028     -0.807      0.420      -0.078       0.033
```

```
x10           -0.0777      0.028     -2.741      0.006      -0.133      -0.022
const          0.0446      0.017      2.670      0.008       0.012       0.077
x11        -1.706e-05   6.72e-06     -2.539      0.011   -3.03e-05   -3.88e-06
==============================================================================
Omnibus:                      537.820   Durbin-Watson:                   1.998
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            32528.550
Skew:                           1.177   Prob(JB):                         0.00
Kurtosis:                      27.971   Cond. No.                     1.63e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.63e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

We verify that p-value for the ADF is not low enough, which means that we do not reject $H_0 : \rho = 0$.

Then, we verify if the coefficient of the trend is statistically significant. Since the t-stat for x11 = -2.539, we conclude that the coef. of the trend is not statistically significant compared to the critical value of 2.78 at the 5% cofidence level.

The next step is to run the test again without the trend term.

[18]:
```python
# Obtains the ADF results for the restricted regression
adf_res = sm.adfuller(
    df_aaa['value'],
    regression='c',
    maxlag=12,
    autolag='AIC',
    regresults=True
)

# Show test results
print(f"T-statistic: {adf_res[0]}")
print(f"P-value: {adf_res[1]}")
print(f"Used lag: {adf_res[3].usedlag}\n")
print(adf_res[3].resols.summary())
```

```
T-statistic: -1.033952299163935
P-value: 0.7406804882311689
Used lag: 9

                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.041
Model:                            OLS   Adj. R-squared:                  0.033
Method:                 Least Squares   F-statistic:                     5.203
Date:                Sun, 19 Sep 2021   Prob (F-statistic):           1.66e-07
```

```
Time:                        16:44:16   Log-Likelihood:                 1952.5
No. Observations:                1241   AIC:                            -3883.
Df Residuals:                    1230   BIC:                            -3827.
Df Model:                          10
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1            -0.0024      0.002     -1.034      0.301      -0.007       0.002
x2             0.0086      0.028      0.303      0.762      -0.047       0.064
x3             0.0884      0.028      3.107      0.002       0.033       0.144
x4             0.0699      0.028      2.458      0.014       0.014       0.126
x5            -0.0982      0.029     -3.444      0.001      -0.154      -0.042
x6             0.0257      0.029      0.898      0.369      -0.030       0.082
x7             0.0154      0.029      0.540      0.590      -0.041       0.071
x8            -0.0791      0.028     -2.781      0.005      -0.135      -0.023
x9            -0.0251      0.028     -0.883      0.378      -0.081       0.031
x10           -0.0796      0.028     -2.804      0.005      -0.135      -0.024
const          0.0072      0.008      0.914      0.361      -0.008       0.023
==============================================================================
Omnibus:                      527.550   Durbin-Watson:                   1.998
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            32344.620
Skew:                           1.134   Prob(JB):                         0.00
Kurtosis:                      27.907   Cond. No.                         80.1
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

We do not reject the non-stationarity hypothesis, because the p-value is 0.713.

Then, we check that constant term is NOT statistically significant, because its t-value is not high enough.

The next step is to run a restricted regression (no constand and no trend.)

[19]:
```python
# Obtains the ADF results for the restricted regression
adf_res = sm.adfuller(
    df_aaa['value'],
    regression='nc',
    maxlag=12,
    autolag='AIC',
    regresults=True
)

# Show test results
print(f"T-statistic: {adf_res[0]}")
print(f"P-value: {adf_res[1]}")
```

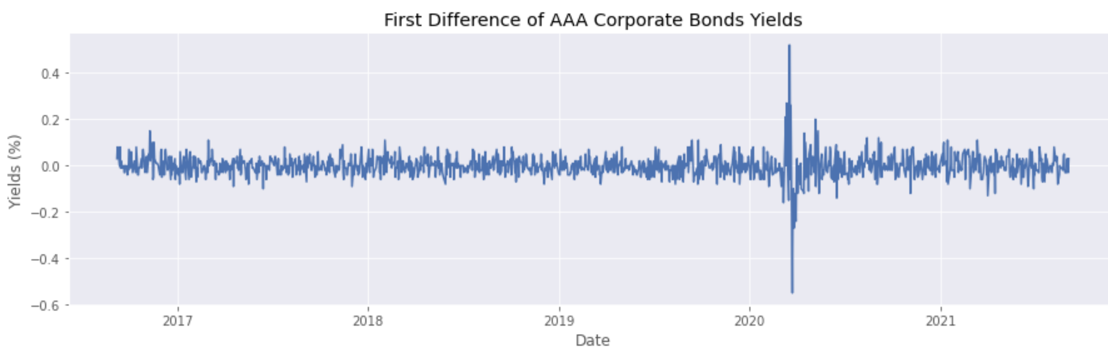
```
print(f"Used lag: {adf_res[3].usedlag}\n")
print(adf_res[3].resols.summary())
```

```
T-statistic: -0.7469882196814753
P-value: 0.39281257611444154
Used lag: 9

                                 OLS Regression Results
================================================================================
=======
Dep. Variable:                      y   R-squared (uncentered):
0.040
Model:                            OLS   Adj. R-squared (uncentered):
0.032
Method:                 Least Squares   F-statistic:
5.149
Date:                Sun, 19 Sep 2021   Prob (F-statistic):
2.06e-07
Time:                        16:44:16   Log-Likelihood:
1952.1
No. Observations:                1241   AIC:
-3884.
Df Residuals:                    1231   BIC:
-3833.
Df Model:                          10
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1            -0.0003      0.000     -0.747      0.455      -0.001       0.001
x2             0.0073      0.028      0.256      0.798      -0.048       0.063
x3             0.0871      0.028      3.065      0.002       0.031       0.143
x4             0.0687      0.028      2.417      0.016       0.013       0.124
x5            -0.0996      0.028     -3.496      0.000      -0.155      -0.044
x6             0.0244      0.029      0.854      0.393      -0.032       0.081
x7             0.0142      0.028      0.500      0.617      -0.042       0.070
x8            -0.0803      0.028     -2.828      0.005      -0.136      -0.025
x9            -0.0264      0.028     -0.931      0.352      -0.082       0.029
x10           -0.0810      0.028     -2.855      0.004      -0.137      -0.025
==============================================================================
Omnibus:                      535.547   Durbin-Watson:                   1.998
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            32614.863
Skew:                           1.167   Prob(JB):                         0.00
Kurtosis:                      28.006   Cond. No.                         76.8
==============================================================================

Notes:
```

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Since we cannot reject the non-stationarity hypothesis, the ADF test suggests that the series is non-stationary without constant. This means that the series has a unit root (stochastic trend), but no drift. We can check this result by running the ADF test on the first difference of the series.

[20]:
```
df_aaa['diff'] = df_aaa['value'].diff()

# Plot
plt.plot(df_aaa['date'], df_aaa['diff'])
plt.title("First Difference of AAA Corporate Bonds Yields")
plt.xlabel("Date")
plt.ylabel("Yields (%)")
plt.show()
```



[21]:
```
adf_res = sm.adfuller(
    df_aaa['diff'].dropna(),
    regression='ct',
    maxlag=12,
    autolag='AIC',
    regresults=True
)

# Show test results
print(f"T-statistic: {adf_res[0]}")
print(f"P-value: {adf_res[1]}")
print(f"Used lag: {adf_res[3].usedlag}\n")
print(adf_res[3].resols.summary())
```

```
T-statistic: -13.713251329274184
P-value: 2.0647792264106046e-21
Used lag: 8
```

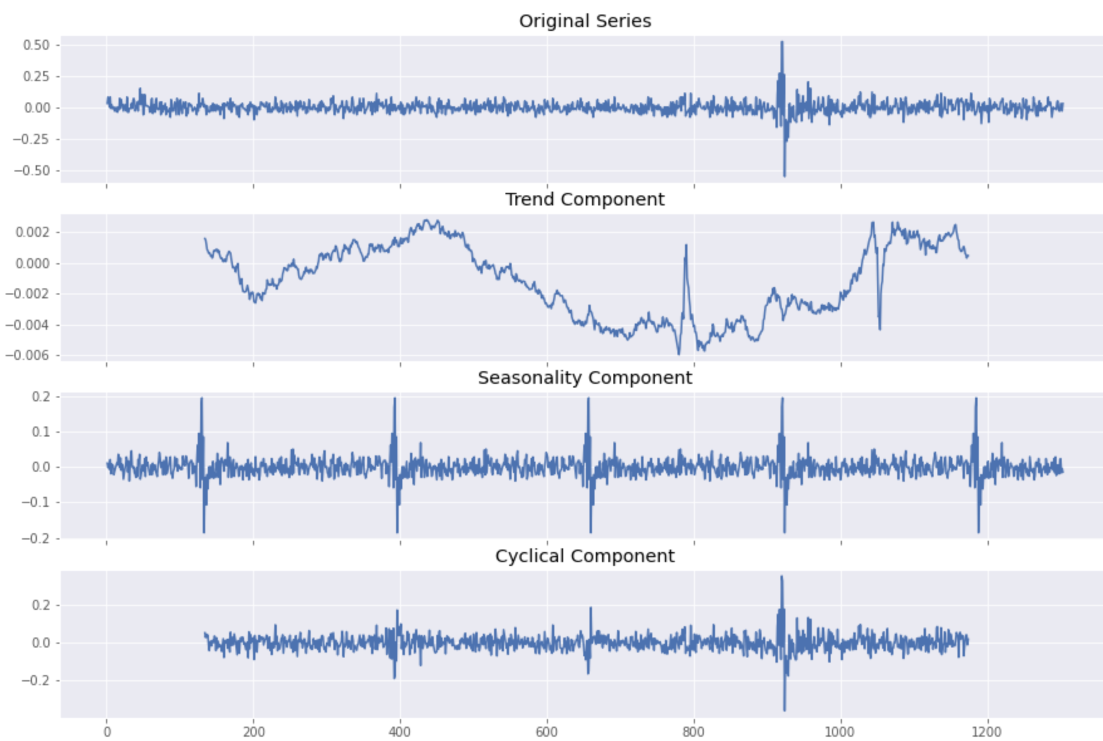
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.516
Model:                            OLS   Adj. R-squared:                  0.512
Method:                 Least Squares   F-statistic:                     131.1
Date:                Sun, 19 Sep 2021   Prob (F-statistic):         6.81e-186
Time:                        16:44:16   Log-Likelihood:                 1952.2
No. Observations:                1241   AIC:                            -3882.
Df Residuals:                    1230   BIC:                            -3826.
Df Model:                          10
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1            -1.0896      0.079    -13.713      0.000      -1.245      -0.934
x2             0.0964      0.074      1.301      0.194      -0.049       0.242
x3             0.1830      0.069      2.665      0.008       0.048       0.318
x4             0.2513      0.064      3.933      0.000       0.126       0.377
x5             0.1514      0.059      2.575      0.010       0.036       0.267
x6             0.1754      0.053      3.335      0.001       0.072       0.279
x7             0.1893      0.047      4.013      0.000       0.097       0.282
x8             0.1086      0.040      2.719      0.007       0.030       0.187
x9             0.0816      0.028      2.875      0.004       0.026       0.137
const          0.0009      0.003      0.297      0.766      -0.005       0.006
x10        -2.687e-06      4e-06     -0.672      0.502   -1.05e-05    5.16e-06
==============================================================================
Omnibus:                      542.753   Durbin-Watson:                   1.998
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            32827.037
Skew:                           1.197   Prob(JB):                         0.00
Kurtosis:                      28.082   Cond. No.                     7.93e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 7.93e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

By running the ADF test on the first difference series, we obtain the result that the non-stationarity hypothesis is rejected, which means that AAA Corporate Bond Yields are I(1). We can also see that both the constant and trend coefficients are not significant, suggesting that the first difference series does not have drift or trend.

4.3 Decomposition and Analysis of Drift, Trend, and Seasonality for the First Difference Series

Now, we shall decompose the first difference series into trend, seasonality, and residues (cyclical component). We will assume that the series follow an annual period, and thus, we consider a period of 253 days (average number of trading days per year).

[22]:
```
# Decomposes the time series
decomposition = seasonal_decompose(df_aaa['diff'].dropna(), model='additive',␣
↪period=253)

# Plot
fig, axs = plt.subplots(4, 1, sharex=True, figsize=(15, 10))
df_aaa['diff'].plot(ax=axs[0])
decomposition.trend.plot(ax=axs[1])
decomposition.seasonal.plot(ax=axs[2])
decomposition.resid.plot(ax=axs[3])
axs[0].set_title("Original Series")
axs[1].set_title("Trend Component")
axs[2].set_title("Seasonality Component")
axs[3].set_title("Cyclical Component")
plt.show()
```



4.3.1 Deterministic Trend

[23]:
```
model = OLS(df_aaa['diff'].dropna(), decomposition.trend, missing='drop')
print(model.fit().summary())
```

```
                            OLS Regression Results
==============================================================================
=======
Dep. Variable:                   diff   R-squared (uncentered):
0.003
Model:                            OLS   Adj. R-squared (uncentered):
0.002
Method:                 Least Squares   F-statistic:
2.915
Date:                Sun, 19 Sep 2021   Prob (F-statistic):
0.0881
Time:                        16:44:17   Log-Likelihood:
1511.8
No. Observations:                 998   AIC:
-3022.
Df Residuals:                     997   BIC:
-3017.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
trend          1.1026      0.646      1.707      0.088      -0.165       2.370
==============================================================================
Omnibus:                      423.075   Durbin-Watson:                   1.977
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            40074.391
Skew:                           0.977   Prob(JB):                         0.00
Kurtosis:                      33.982   Cond. No.                         1.00
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not
contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

We can see that the trend component is not statistical significant at the 5% level, which confirms our previous findings. Therefore, these results indicate that the series don't have a deterministic trend when considering an annual period.

4.3.2 Seasonality

[24]:
```python
model = OLS(df_aaa['diff'].dropna(), decomposition.seasonal, missing='drop')
print(model.fit().summary())
```

```
                                 OLS Regression Results
================================================================================
=======
Dep. Variable:                   diff   R-squared (uncentered):
0.173
Model:                            OLS   Adj. R-squared (uncentered):
0.172
Method:                 Least Squares   F-statistic:
260.8
Date:                Sun, 19 Sep 2021   Prob (F-statistic):
2.00e-53
Time:                        16:44:17   Log-Likelihood:
2060.0
No. Observations:                1250   AIC:
-4118.
Df Residuals:                    1249   BIC:
-4113.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
seasonal       0.6959      0.043     16.149      0.000       0.611       0.780
==============================================================================
Omnibus:                      316.210   Durbin-Watson:                   1.999
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            10588.675
Skew:                           0.457   Prob(JB):                         0.00
Kurtosis:                      17.229   Cond. No.                         1.00
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not
contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

We can see that the seasonal component is statistically significant, when considering a period equivalent to the trading year, indicating that there's seasonality in the series. We can also verify this by using dummy variables for each month and verifying their statistical significance.

[25]:
```python
# create dummy vars columns for week of the year
df_aaa['period'] = df_aaa['date'].dt.weekofyear
```

```
# create a column for each dummy var
for period in df_aaa['period'].unique():
    df_aaa[f'is_{period}'] = 1*(df_aaa['period'] == period)

# include all dummy vars except one
X = df_aaa[[f'is_{period}' for period in sorted(df_aaa['period'].unique())[:
↪-1]]]

# run regression
model = OLS(df_aaa['diff'], X, missing='drop')
res = model.fit()
print(res.summary())
```

```
                                 OLS Regression Results
================================================================================
=======
Dep. Variable:                   diff   R-squared (uncentered):
0.065
Model:                            OLS   Adj. R-squared (uncentered):
0.024
Method:                 Least Squares   F-statistic:
1.599
Date:                Sun, 19 Sep 2021   Prob (F-statistic):
0.00493
Time:                        16:44:17   Log-Likelihood:
1983.4
No. Observations:                1250   AIC:
-3863.
Df Residuals:                    1198   BIC:
-3596.
Df Model:                          52
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
is_1           0.0071      0.011      0.647      0.518      -0.015       0.029
is_2        4.337e-19      0.010   4.29e-17      1.000      -0.020       0.020
is_3           0.0100      0.011      0.928      0.354      -0.011       0.031
is_4          -0.0135      0.011     -1.278      0.201      -0.034       0.007
is_5           0.0124      0.010      1.226      0.220      -0.007       0.032
is_6          -0.0020      0.010     -0.198      0.843      -0.022       0.018
is_7           0.0062      0.010      0.605      0.545      -0.014       0.027
is_8           0.0010      0.011      0.086      0.931      -0.021       0.023
is_9           0.0108      0.010      1.068      0.286      -0.009       0.031
is_10         -0.0040      0.010     -0.395      0.693      -0.024       0.016
is_11          0.0236      0.010      2.333      0.020       0.004       0.043
is_12          0.0304      0.010      3.006      0.003       0.011       0.050
```

is_13	-0.0643	0.011	-6.103	0.000	-0.085	-0.044
is_14	-0.0104	0.010	-1.028	0.304	-0.030	0.009
is_15	-0.0126	0.011	-1.196	0.232	-0.033	0.008
is_16	0.0029	0.010	0.283	0.778	-0.017	0.023
is_17	0.0004	0.010	0.040	0.968	-0.019	0.020
is_18	0.0100	0.010	0.989	0.323	-0.010	0.030
is_19	0.0052	0.010	0.514	0.607	-0.015	0.025
is_20	-0.0040	0.010	-0.395	0.693	-0.024	0.016
is_21	-0.0124	0.010	-1.226	0.220	-0.032	0.007
is_22	-0.0125	0.011	-1.105	0.269	-0.035	0.010
is_23	0.0032	0.010	0.316	0.752	-0.017	0.023
is_24	-0.0156	0.010	-1.542	0.123	-0.035	0.004
is_25	0.0008	0.010	0.079	0.937	-0.019	0.021
is_26	-0.0116	0.010	-1.147	0.252	-0.031	0.008
is_27	0.0030	0.011	0.265	0.791	-0.019	0.025
is_28	-0.0128	0.010	-1.266	0.206	-0.033	0.007
is_29	-0.0032	0.010	-0.316	0.752	-0.023	0.017
is_30	-0.0044	0.010	-0.435	0.664	-0.024	0.015
is_31	-0.0036	0.010	-0.356	0.722	-0.023	0.016
is_32	-0.0052	0.010	-0.514	0.607	-0.025	0.015
is_33	0.0016	0.010	0.158	0.874	-0.018	0.021
is_34	-0.0024	0.010	-0.237	0.812	-0.022	0.017
is_35	0.0068	0.010	0.672	0.501	-0.013	0.027
is_36	0.0110	0.011	0.973	0.331	-0.011	0.033
is_37	0.0158	0.010	1.534	0.125	-0.004	0.036
is_38	-0.0108	0.010	-1.068	0.286	-0.031	0.009
is_39	-0.0008	0.010	-0.079	0.937	-0.021	0.019
is_40	0.0076	0.010	0.751	0.453	-0.012	0.027
is_41	0.0029	0.010	0.283	0.778	-0.017	0.023
is_42	-0.0030	0.011	-0.289	0.773	-0.024	0.018
is_43	0.0140	0.010	1.384	0.167	-0.006	0.034
is_44	-0.0028	0.010	-0.277	0.782	-0.023	0.017
is_45	0.0122	0.011	1.155	0.249	-0.009	0.033
is_46	0.0068	0.011	0.632	0.527	-0.014	0.028
is_47	-0.0118	0.011	-1.096	0.273	-0.033	0.009
is_48	-0.0035	0.011	-0.330	0.742	-0.024	0.017
is_49	0.0037	0.010	0.363	0.716	-0.017	0.024
is_50	-0.0084	0.010	-0.831	0.406	-0.028	0.011
is_51	0.0036	0.010	0.356	0.722	-0.016	0.023
is_52	-0.0100	0.011	-0.884	0.377	-0.032	0.012

Omnibus:	469.106	Durbin-Watson:	2.068
Prob(Omnibus):	0.000	Jarque-Bera (JB):	28206.685
Skew:	0.901	Prob(JB):	0.00
Kurtosis:	26.202	Cond. No.	1.12

Notes:

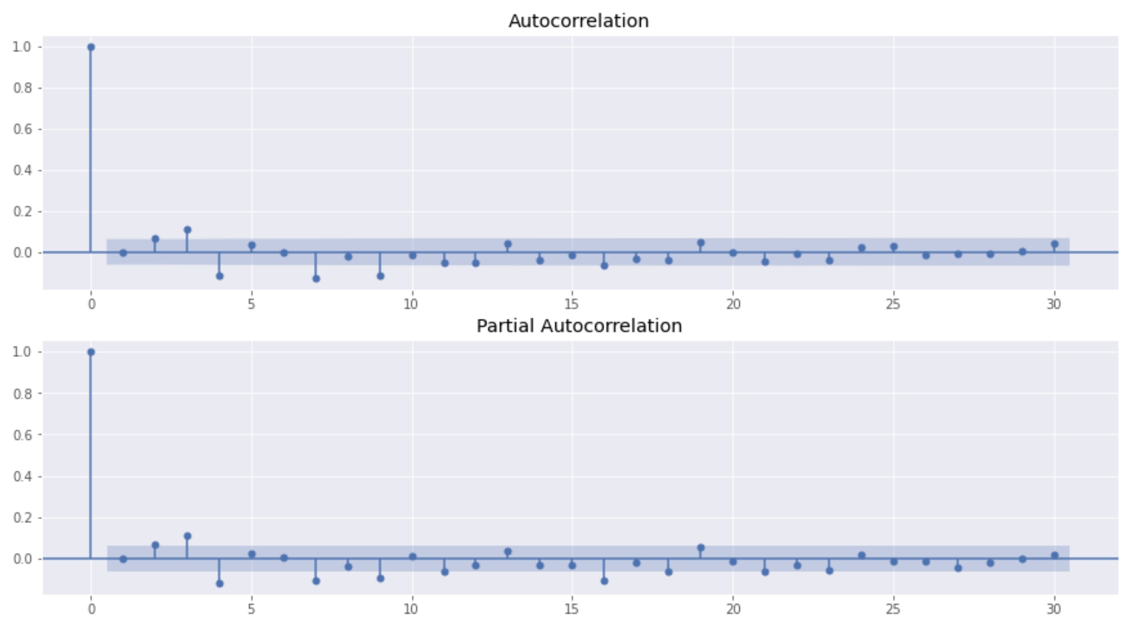
```
[1] R² is computed without centering (uncentered) since the model does not
contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

The results indicate that some weeks of the year have statistically significant impacts on the value of the first difference series. This supports the previous finding that the series has a significant seasonal component.

4.4 Stationary ARMA Model

The goal here is to estimate the cyclical component through a stationary ARMA model. First, let's estimate the parameters q and p through the autocorrelation and partial autocorrelation of the series.

[26]:
```python
# Plot
fig, axs = plt.subplots(2, 1, figsize=(15, 8))
plot_acf(decomposition.resid.dropna(), ax=axs[0])
plot_pacf(decomposition.resid.dropna(), ax=axs[1])
plt.show()
```



From the autocorrelation plot we can estimate that values are close enough to zero when $q > 9$, and from the partial autocorrelation plot, we can estimate the same when $p > 7$. Now, we can fit the model and verify the results.

[27]:
```python
arma_model = ARIMA(decomposition.resid.dropna(), order=(9, 0, 7))
print(arma_model.fit().summary())
```

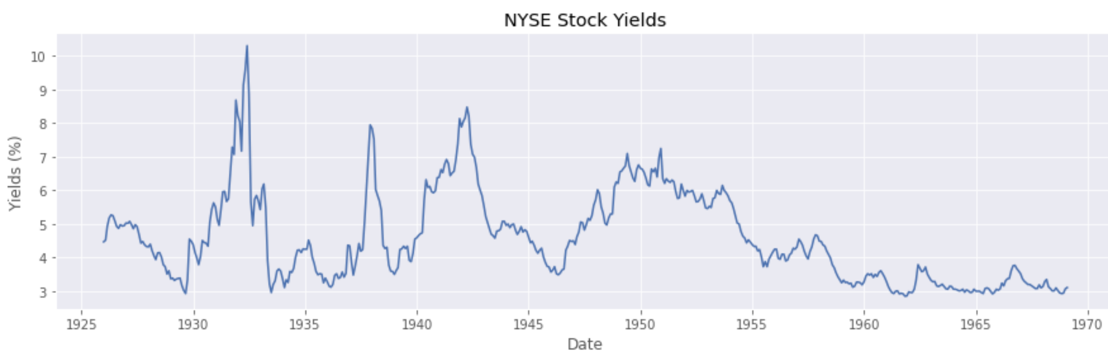
```
                              SARIMAX Results
==============================================================================
Dep. Variable:                  resid   No. Observations:                  998
Model:                 ARIMA(9, 0, 7)   Log Likelihood                1716.102
Date:                Sun, 19 Sep 2021   AIC                          -3396.204
Time:                        16:44:28   BIC                          -3307.901
Sample:                             0   HQIC                         -3362.640
                                - 998
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0003      0.001      0.286      0.775      -0.001       0.002
ar.L1          0.0299      0.365      0.082      0.935      -0.686       0.745
ar.L2          0.1119      0.262      0.427      0.669      -0.401       0.625
ar.L3          0.0959      0.263      0.365      0.715      -0.419       0.610
ar.L4          0.0137      0.218      0.063      0.950      -0.414       0.441
ar.L5          0.0764      0.224      0.341      0.733      -0.363       0.515
ar.L6         -0.0467      0.213     -0.219      0.827      -0.465       0.372
ar.L7          0.2691      0.178      1.513      0.130      -0.079       0.618
ar.L8         -0.0568      0.074     -0.769      0.442      -0.201       0.088
ar.L9         -0.1341      0.070     -1.904      0.057      -0.272       0.004
ma.L1         -0.0192      0.366     -0.052      0.958      -0.736       0.698
ma.L2         -0.0493      0.263     -0.187      0.851      -0.565       0.467
ma.L3         -0.0213      0.249     -0.085      0.932      -0.509       0.467
ma.L4         -0.1085      0.211     -0.515      0.607      -0.521       0.305
ma.L5         -0.0794      0.217     -0.366      0.715      -0.505       0.346
ma.L6          0.0583      0.213      0.274      0.784      -0.359       0.475
ma.L7         -0.4056      0.173     -2.340      0.019      -0.745      -0.066
sigma2         0.0019   5.38e-05     35.315      0.000       0.002       0.002
================================================================================
===
Ljung-Box (L1) (Q):                   0.36   Jarque-Bera (JB):
4006.25
Prob(Q):                              0.55   Prob(JB):
0.00
Heteroskedasticity (H):               2.17   Skew:
0.36
Prob(H) (two-sided):                  0.00   Kurtosis:
12.79
================================================================================
===

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

5 NYSE Stock Yields

5.1 Load the Data

Notice that here we don't apply the log-transform in our data since it is already given in percentage.

[28]:
```
# Loads data and drops non available values
df_nyse = pd.read_csv("./datasets/NYSE_stock_yields.csv", names=['date',␣
↪'value'], parse_dates=['date'], skiprows=[0], na_values='.')
df_nyse.dropna(inplace=True)

# Plots the time series
plt.plot(df_nyse['date'], df_nyse['value'])
plt.title("NYSE Stock Yields")
plt.xlabel("Date")
plt.ylabel("Yields (%)")
plt.show()
```



First, we are going to implement the augmented Dickey-Fuller Test to the most general regression:

$$\Delta X_t = b_0 + b_1 t + \rho X_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta_{t-j} + \varepsilon_t$$

In this case, the hypotesis can be written as:

$$\begin{aligned} H_0: &\quad \rho = 0 \\ H_1: &\quad \rho < 0 \end{aligned}$$

[29]:
```
# Obtains the ADF results for the general regression
adf_res = sm.adfuller(
    df_nyse['value'],
```

```python
    regression='ct',
    maxlag=12,
    autolag='AIC',
    regresults=True
)

# Show test results
print(f"T-statistic: {adf_res[0]}")
print(f"P-value: {adf_res[1]}")
print(f"Used lag: {adf_res[3].usedlag}\n")
print(adf_res[3].resols.summary())
```

```
T-statistic: -3.4096114441815564
P-value: 0.050116958198374056
Used lag: 4

                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.150
Model:                            OLS   Adj. R-squared:                  0.140
Method:                 Least Squares   F-statistic:                     14.88
Date:                Sun, 19 Sep 2021   Prob (F-statistic):           1.06e-15
Time:                        16:44:29   Log-Likelihood:                -110.74
No. Observations:                 513   AIC:                             235.5
Df Residuals:                     506   BIC:                             265.2
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1            -0.0378      0.011     -3.410      0.001      -0.060      -0.016
x2             0.3334      0.044      7.581      0.000       0.247       0.420
x3            -0.0507      0.045     -1.116      0.265      -0.140       0.039
x4            -0.1785      0.045     -3.947      0.000      -0.267      -0.090
x5             0.1344      0.044      3.054      0.002       0.048       0.221
const          0.2057      0.066      3.126      0.002       0.076       0.335
x6            -0.0001   9.73e-05     -1.450      0.148      -0.000    5.01e-05
==============================================================================
Omnibus:                      155.383   Durbin-Watson:                   1.998
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             9645.278
Skew:                          -0.331   Prob(JB):                         0.00
Kurtosis:                      24.232   Cond. No.                     1.50e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

```
[2] The condition number is large, 1.5e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

We can see that the t-statistic is not statistically significant at a 1% risk level so we can not reject the null hypothesis. b1 (here x6) is not statistically significant.

Thus, we perform the restricted regression without the trend term.

[30]:
```python
# Obtains the ADF results
adf_res = sm.adfuller(
    df_nyse['value'],
    regression='c',
    maxlag=12,
    autolag='AIC',
    regresults=True
)

# Show test results
print(f"T-statistic: {adf_res[0]}")
print(f"P-value: {adf_res[1]}")
print(f"Used lag: {adf_res[3].usedlag}\n")
print(adf_res[3].resols.summary())
```

```
T-statistic: -3.0879577428483445
P-value: 0.02745730031042599
Used lag: 4

                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.146
Model:                            OLS   Adj. R-squared:                  0.138
Method:                 Least Squares   F-statistic:                     17.40
Date:                Sun, 19 Sep 2021   Prob (F-statistic):           6.53e-16
Time:                        16:44:29   Log-Likelihood:                -111.81
No. Observations:                 513   AIC:                             235.6
Df Residuals:                     507   BIC:                             261.1
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1            -0.0317      0.010     -3.088      0.002      -0.052      -0.012
x2             0.3311      0.044      7.526      0.000       0.245       0.418
x3            -0.0535      0.045     -1.178      0.239      -0.143       0.036
x4            -0.1807      0.045     -3.992      0.000      -0.270      -0.092
x5             0.1321      0.044      3.002      0.003       0.046       0.219
const          0.1418      0.049      2.900      0.004       0.046       0.238
==============================================================================
Omnibus:                      151.753   Durbin-Watson:                   1.998
```

```
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             9615.288
Skew:                          -0.257   Prob(JB):                         0.00
Kurtosis:                      24.203   Cond. No.                         20.0
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

We can see that the t-statistic is not statistically significant at a 1% risk level so we can not reject the null hypothesis.

b0 (here constant) is not statistically significant at a 1% risk level

Thus, we perform the restricted regression without the trend term and without the constant

[31]:
```python
# Obtains the ADF results
adf_res = sm.adfuller(
    df_nyse['value'],
    regression='nc',
    maxlag=12,
    autolag='AIC',
    regresults=True
)

# Show test results
print(f"T-statistic: {adf_res[0]}")
print(f"P-value: {adf_res[1]}")
print(f"Used lag: {adf_res[3].usedlag}\n")
print(adf_res[3].resols.summary())
```

```
T-statistic: -1.0848432040960392
P-value: 0.2513731607012235
Used lag: 4

                                 OLS Regression Results
================================================================================
=======
Dep. Variable:                      y   R-squared (uncentered):
0.132
Model:                            OLS   Adj. R-squared (uncentered):
0.124
Method:                 Least Squares   F-statistic:
15.51
Date:                Sun, 19 Sep 2021   Prob (F-statistic):
3.27e-14
Time:                        16:44:30   Log-Likelihood:
-116.03
No. Observations:                 513   AIC:
```

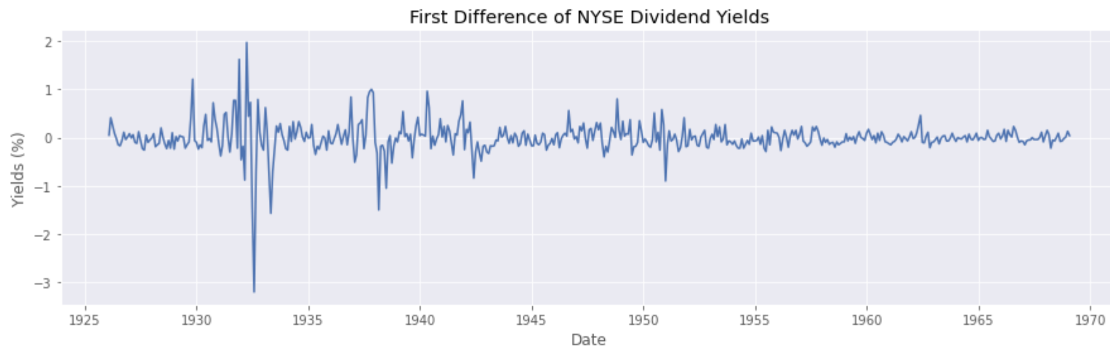
```
242.1
Df Residuals:                     508   BIC:
263.3
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1            -0.0031      0.003     -1.085      0.279      -0.009       0.002
x2             0.3167      0.044      7.192      0.000       0.230       0.403
x3            -0.0697      0.045     -1.536      0.125      -0.159       0.019
x4            -0.1944      0.045     -4.287      0.000      -0.283      -0.105
x5             0.1159      0.044      2.635      0.009       0.029       0.202
==============================================================================
Omnibus:                      181.982   Durbin-Watson:                   1.994
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            10308.014
Skew:                          -0.680   Prob(JB):                         0.00
Kurtosis:                      24.918   Cond. No.                         19.6
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not
contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

The t-value is not statistically significant

We conclude that the series is not stationnary without a drift.

We will now investigate the first difference time series in order to see if the series is I(1).

[32]:
```python
df_nyse['diff'] = df_nyse['value'].diff()

# Plot
plt.plot(df_nyse['date'], df_nyse['diff'])
plt.title("First Difference of NYSE Dividend Yields")
plt.xlabel("Date")
plt.ylabel("Yields (%)")
plt.show()
```



[33]:
```python
adf_res = sm.adfuller(
    df_nyse['diff'].dropna(),
    regression='ct',
    maxlag=12,
    autolag='AIC',
    regresults=True
)

# Show test results
print(f"T-statistic: {adf_res[0]}")
print(f"P-value: {adf_res[1]}")
print(f"Used lag: {adf_res[3].usedlag}\n")
print(adf_res[3].resols.summary())
```

```
T-statistic: -6.9019835918490875
P-value: 2.3517361165049426e-08
Used lag: 12

                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.419
Model:                            OLS   Adj. R-squared:                  0.402
Method:                 Least Squares   F-statistic:                     25.15
Date:                Sun, 19 Sep 2021   Prob (F-statistic):           4.69e-49
Time:                        16:44:30   Log-Likelihood:                -108.70
No. Observations:                 504   AIC:                             247.4
Df Residuals:                     489   BIC:                             310.7
Df Model:                          14
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1            -0.9346      0.135     -6.902      0.000      -1.201      -0.669
x2             0.2694      0.131      2.059      0.040       0.012       0.526
```

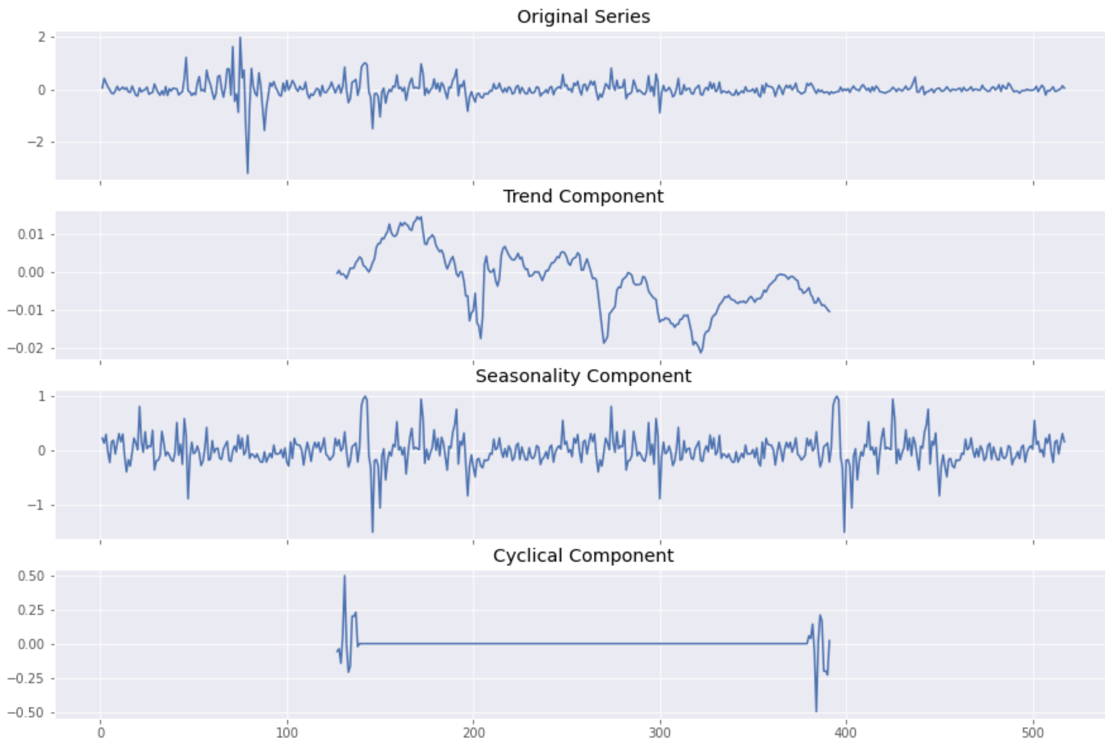
```
x3             0.1881      0.125      1.504      0.133      -0.058       0.434
x4            -0.0168      0.119     -0.141      0.888      -0.251       0.217
x5             0.1295      0.114      1.138      0.256      -0.094       0.353
x6             0.0689      0.107      0.646      0.519      -0.141       0.278
x7             0.1229      0.100      1.231      0.219      -0.073       0.319
x8             0.1021      0.093      1.098      0.273      -0.081       0.285
x9             0.0471      0.084      0.558      0.577      -0.119       0.213
x10            0.1573      0.076      2.065      0.039       0.008       0.307
x11            0.0724      0.063      1.144      0.253      -0.052       0.197
x12            0.0423      0.054      0.787      0.432      -0.063       0.148
x13            0.1573      0.045      3.528      0.000       0.070       0.245
const          0.0027      0.027      0.099      0.921      -0.051       0.056
x14        -2.448e-05   9.34e-05     -0.262      0.793      -0.000       0.000
==============================================================================
Omnibus:                      178.880   Durbin-Watson:                   1.991
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             9810.814
Skew:                          -0.688   Prob(JB):                         0.00
Kurtosis:                      24.571   Cond. No.                     6.93e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 6.93e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

We can see that the t-value is statistically significant.

The b0 (constant) and b1 (trend term) are not statiscally significant, whichconfirms our previous findings.

We conclude that the time series is I(1) with no drift.

[34]:
```python
# Decomposes the time series
decomposition = seasonal_decompose(df_nyse['diff'].dropna(), model='additive',␣
 ↪period=253)

# Plot
fig, axs = plt.subplots(4, 1, sharex=True, figsize=(15, 10))
df_nyse['diff'].plot(ax=axs[0])
decomposition.trend.plot(ax=axs[1])
decomposition.seasonal.plot(ax=axs[2])
decomposition.resid.plot(ax=axs[3])
axs[0].set_title("Original Series")
axs[1].set_title("Trend Component")
axs[2].set_title("Seasonality Component")
axs[3].set_title("Cyclical Component")
plt.show()
```



5.1.1 Deterministic Trend

[35]:
```
model = OLS(df_nyse['diff'].dropna(), decomposition.trend, missing='drop')
print(model.fit().summary())
```

```
                            OLS Regression Results
================================================================================
=======
Dep. Variable:                   diff   R-squared (uncentered):
0.019
Model:                            OLS   Adj. R-squared (uncentered):
0.015
Method:                 Least Squares   F-statistic:
5.098
Date:                Sun, 19 Sep 2021   Prob (F-statistic):
0.0248
Time:                        16:44:32   Log-Likelihood:
-36.898
No. Observations:                 265   AIC:
75.80
Df Residuals:                     264   BIC:
79.38
```

```
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
trend          4.8204      2.135      2.258      0.025       0.617       9.024
==============================================================================
Omnibus:                       41.037   Durbin-Watson:                   1.354
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              359.956
Skew:                          -0.069   Prob(JB):                     6.86e-79
Kurtosis:                       8.708   Cond. No.                         1.00
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not
contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

We find that b1 is not statiscally significant

5.1.2 Seasonality

[36]:
```
model = OLS(df_nyse['diff'].dropna(), decomposition.seasonal, missing='drop')
print(model.fit().summary())
```

```
                                 OLS Regression Results
================================================================================
=======
Dep. Variable:                   diff   R-squared (uncentered):
0.174
Model:                            OLS   Adj. R-squared (uncentered):
0.172
Method:                 Least Squares   F-statistic:
108.6
Date:                Sun, 19 Sep 2021   Prob (F-statistic):
3.26e-23
Time:                        16:44:32   Log-Likelihood:
-103.40
No. Observations:                 517   AIC:
208.8
Df Residuals:                     516   BIC:
213.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
```

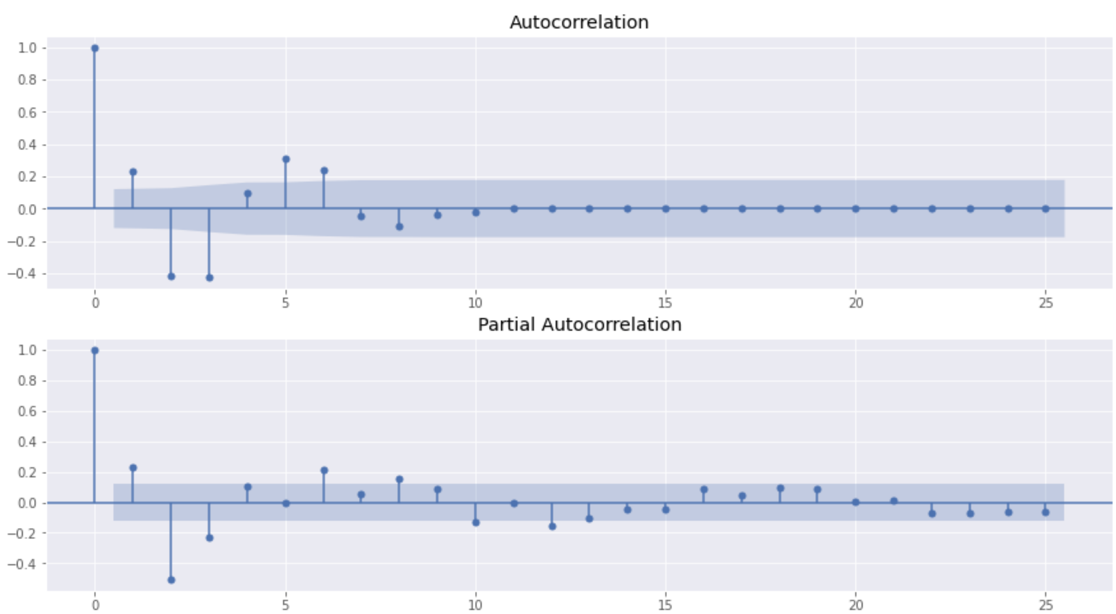
```
------------------------------------------------------------------------------
seasonal        0.4923      0.047     10.423      0.000       0.400       0.585
==============================================================================
Omnibus:                      358.426   Durbin-Watson:                   1.464
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            31957.458
Skew:                          -2.194   Prob(JB):                         0.00
Kurtosis:                      41.266   Cond. No.                         1.00
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not
contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

The seasonal coefficient is statiscally significant. The time series has a seasonal component.

5.2 Stationary ARMA Model

The goal here is to estimate the cyclical component through a stationary ARMA model. First, let's estimate the parameters q and p through the autocorrelation and partial autocorrelation of the series.

[37]:
```
# Plot
fig, axs = plt.subplots(2, 1, figsize=(15, 8))
plot_acf(decomposition.resid.dropna(), ax=axs[0])
plot_pacf(decomposition.resid.dropna(), ax=axs[1])
plt.show()
```



Graphically we can estimate $q > 6$.

[39]:
```
arma_model = ARIMA(decomposition.resid.dropna(), order=(6, 0, 12))
print(arma_model.fit().summary())
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                  resid   No. Observations:                  265
Model:                ARIMA(6, 0, 12)   Log Likelihood                 455.402
Date:                Sun, 19 Sep 2021   AIC                           -870.804
Time:                        16:45:48   BIC                           -799.210
Sample:                             0   HQIC                          -842.039
                                - 265
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0034      0.004      0.781      0.435      -0.005       0.012
ar.L1          0.0639      2.432      0.026      0.979      -4.702       4.830
ar.L2         -0.2614      2.058     -0.127      0.899      -4.295       3.773
ar.L3         -0.1048      2.008     -0.052      0.958      -4.041       3.831
ar.L4          0.1677      1.781      0.094      0.925      -3.323       3.658
ar.L5          0.0546      1.114      0.049      0.961      -2.130       2.239
ar.L6          0.2892      0.712      0.406      0.684      -1.105       1.684
ma.L1          0.1085      2.431      0.045      0.964      -4.657       4.874
ma.L2         -0.2300      2.289     -0.100      0.920      -4.717       4.256
ma.L3         -0.1931      1.224     -0.158      0.875      -2.592       2.206
ma.L4          0.1384      0.777      0.178      0.859      -1.385       1.662
ma.L5          0.1245      0.810      0.154      0.878      -1.463       1.712
ma.L6          0.1774      0.997      0.178      0.859      -1.777       2.132
ma.L7          0.1370      1.214      0.113      0.910      -2.243       2.517
ma.L8         -0.1600      1.282     -0.125      0.901      -2.673       2.353
ma.L9         -0.0005      0.713     -0.001      0.999      -1.398       1.397
ma.L10        -0.3131      0.507     -0.617      0.537      -1.308       0.681
ma.L11        -0.0047      0.512     -0.009      0.993      -1.009       0.999
ma.L12         0.0528      0.432      0.122      0.903      -0.794       0.899
sigma2         0.0019   5.78e-05     32.117      0.000       0.002       0.002
================================================================================
===
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):
28310.12
Prob(Q):                              0.96   Prob(JB):
0.00
Heteroskedasticity (H):               1.00   Skew:
-1.21
Prob(H) (two-sided):                  0.99   Kurtosis:
```

```
53.58
==============================================================================
===

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