Econometrics Assignment 4

Shalem Sumanthiran - sps9893

Question 1

In [ ]:
```
import pandas as pd
from fredapi import Fred
import matplotlib.pyplot as plt
import numpy as np
from statsmodels.tsa.stattools import acf
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.tsa.ar_model import AutoReg
import statsmodels.tsa.ardl as ardl

fred = Fred(api_key='b1f5ecf8eb1278d11cc738e92436e960')
housing = fred.get_series('HOUST', observation_start = '1959-01-01')
housing = pd.DataFrame(housing)
housing.columns = ['HOUST']
housing = housing.dropna()
print(housing.head(3))
print(housing.tail(3))
```

```
             HOUST
1959-01-01  1657.0
1959-02-01  1667.0
1959-03-01  1620.0
             HOUST
2023-08-01  1305.0
2023-09-01  1346.0
2023-10-01  1372.0
```

Data: New Privately-Owned Housing Units Started: Total Units

Source: U.S. Census Bureau Source: U.S. Department of Housing and Urban Development
Release: New Residential Construction
Units: Thousands of Units, Seasonally Adjusted Annual Rate

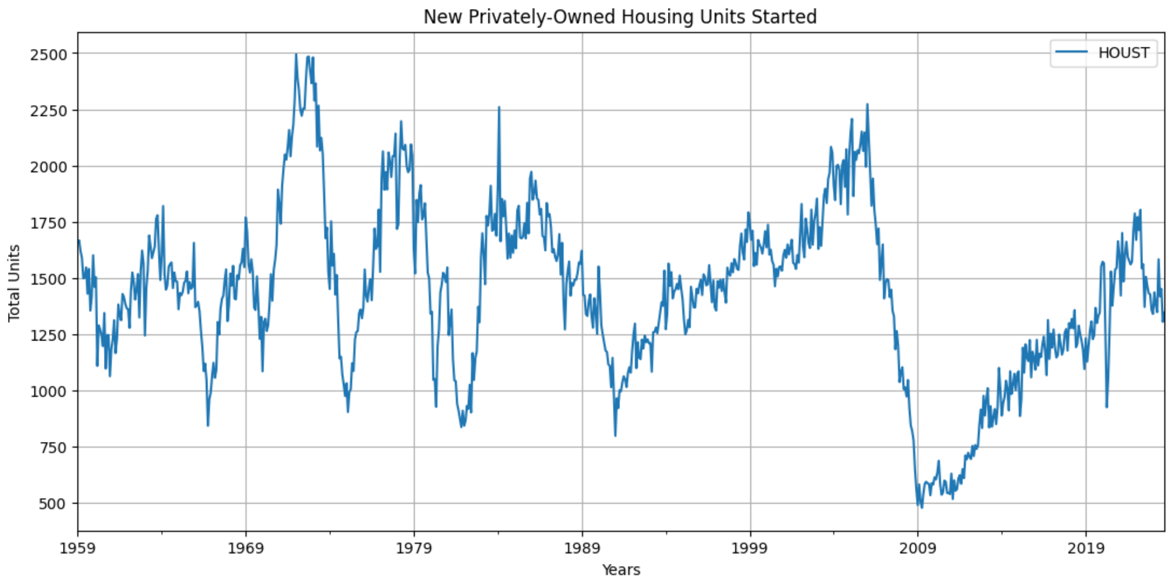
Frequency: Monthly

As provided by the Census, start occurs when excavation begins for the footings or foundation of a building. All housing units in a multifamily building are defined as being started when this excavation begins. Beginning with data for September 1992, estimates of housing starts include units in structures being totally rebuilt on an existing foundation.

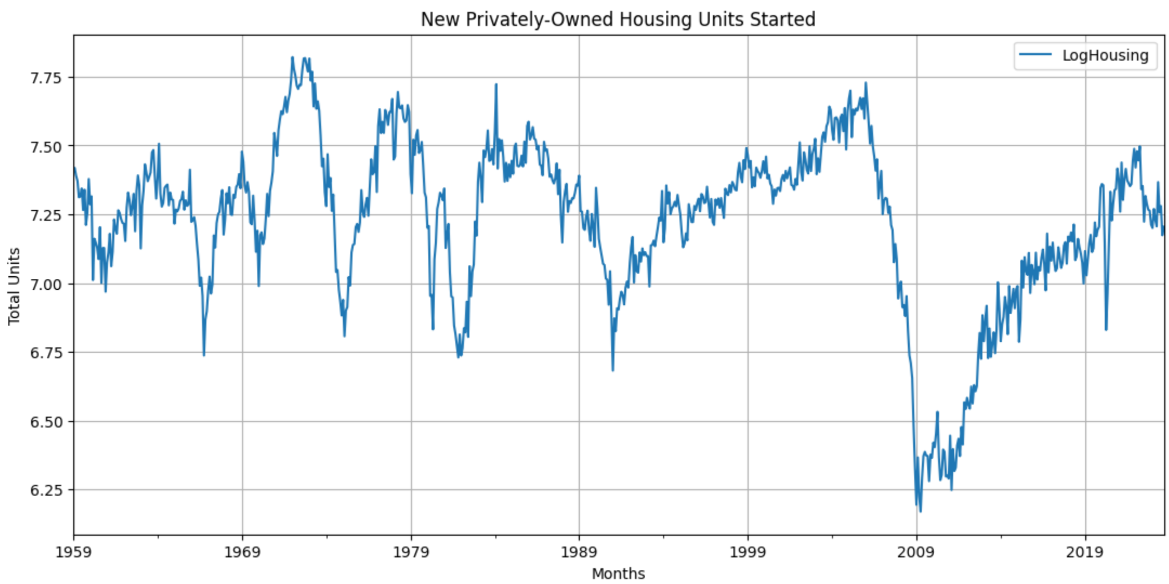
Suggested Citation: U.S. Census Bureau and U.S. Department of Housing and Urban Development, New Privately-Owned Housing Units Started: Total Units [HOUST], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/HOUST, December 11, 2023.

Question 2

In [ ]:
```
plt.figure(figsize=(13, 6))
dates = housing.index
housing['HOUST'][dates < '2023-10-1'].plot()
plt.title('New Privately-Owned Housing Units Started')
plt.xlabel('Years')
plt.ylabel('Total Units')
plt.legend()
plt.grid(True)
plt.show()
```

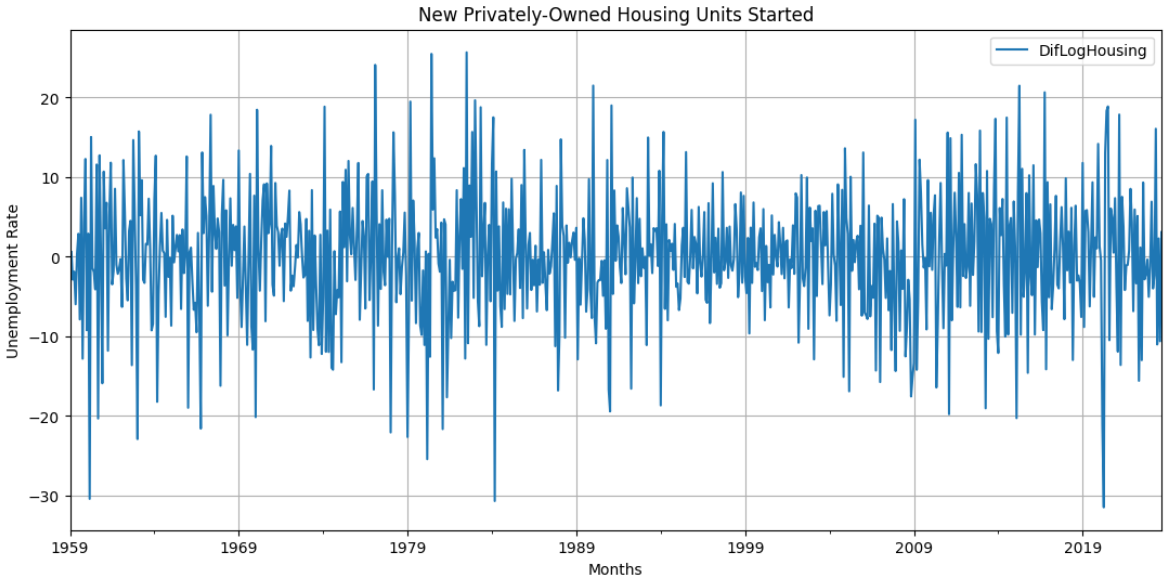


In [ ]:
```
plt.figure(figsize=(13, 6))
housing['LogHousing'] = np.log(housing['HOUST'])
housing['LogHousing'][dates < '2023-10-1'].plot()
plt.title('New Privately-Owned Housing Units Started')
plt.xlabel('Months')
plt.ylabel('Total Units')
plt.legend()
plt.grid(True)
plt.show()
```



In [ ]:
```
plt.figure(figsize=(13, 6))
dates = housing.index
housing['DifLogHousing'] = 100 * (np.log(housing['HOUST']) - np.log(housing['HOU
```

```
housing['DifLogHousing'][dates < '2023-10-1'].plot()
plt.title('New Privately-Owned Housing Units Started')
plt.xlabel('Months')
plt.ylabel('Unemployment Rate')
plt.legend()
plt.grid(True)
plt.show()
```



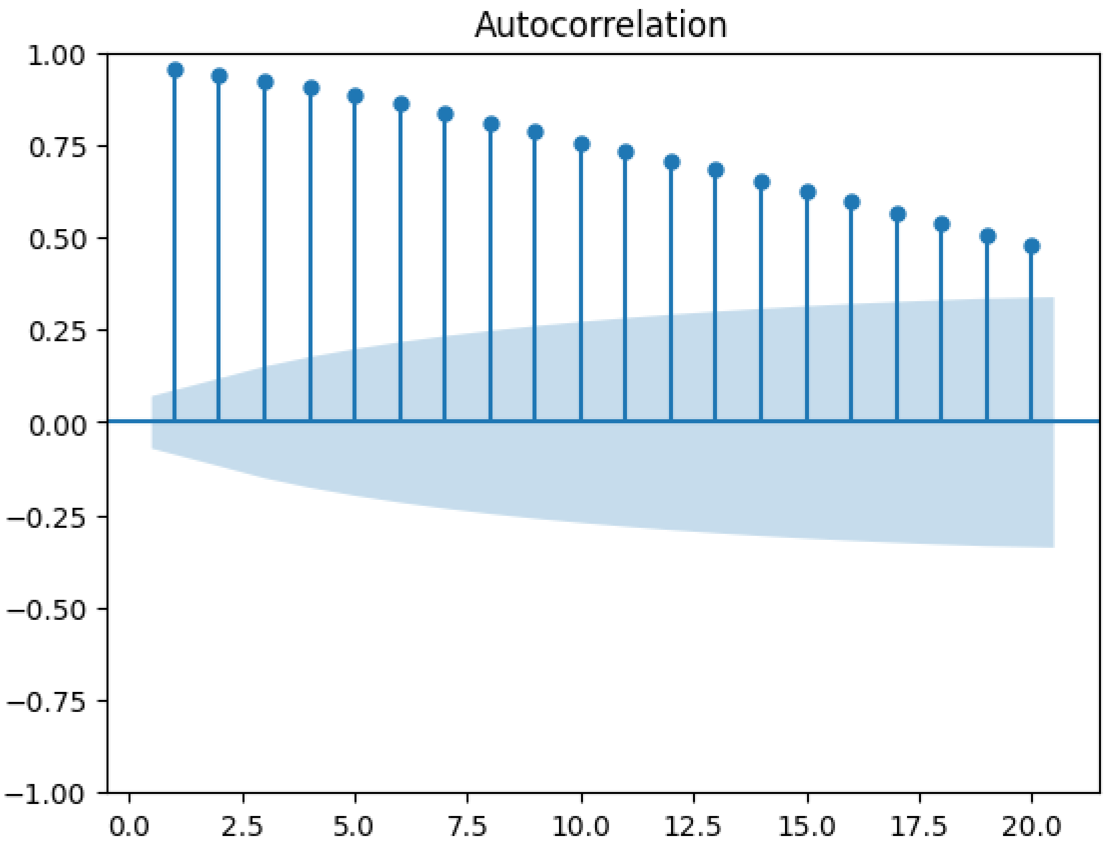
The Housing Start data series which is already seasonally adjusted does not exhibit any particular behavior of a trend, although notably the values decrease significantly during the 2008 housing crisis, and there is a slight dip during the COVID-19 pandemic as well. The original series shows a range of variation of around 500 to 2500 total units per month. The graph returns to a level of around 1500 total units per month. The peaks and troughs in the data may be explained by expansions and recessions in the US economy during those periods.

Log transforming the series does not seem necessary as it does not show a trend of exponential growth, and the first two graphs above are quite similar.

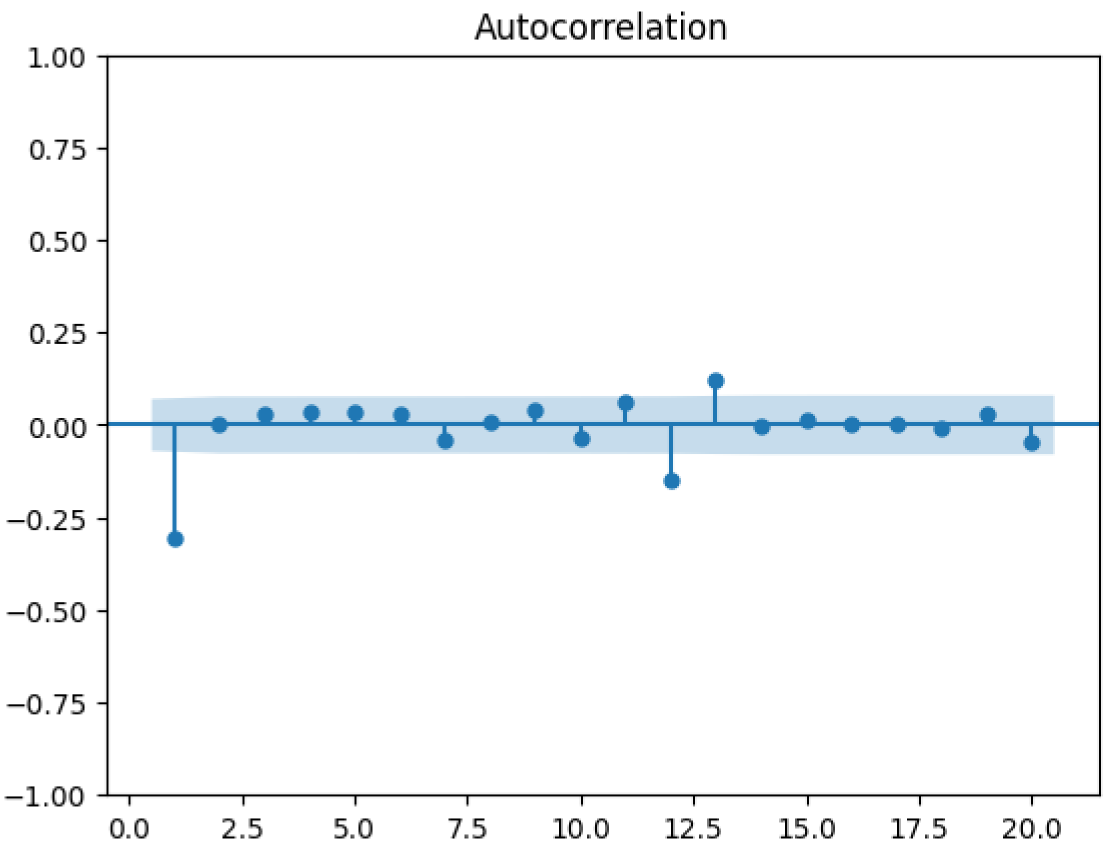
Differencing the logs, however, creates a significant difference in that the graph appears non-stationary and mean-reverting. This can be considered later in the analysis (Question 4)

Question 3

In [ ]:
```
plot_acf(housing['HOUST'], lags=20, missing = 'drop',
          zero = False)
plt.show()
```



In [ ]:
```
plot_acf(housing['DifLogHousing'], lags=20, missing = 'drop',
           zero = False)
plt.show()
```



The ACF of the full sample and the ACF of the sample excluding the COVID-19 pandemic are not significantly different, therefore we opt to use the full sample.

The ACF shows positive autocorrelations, therefore observations are positively correlated with previous observations. The ACF of the original series also exhibits a decay in autocorrelation as the lags increase, which suggests the presence of a significant autocorrelation structure in the time series. The consistently elevated spikes suggest a persistent autocorrelation structure in the data. This could be indicative of a trend or repeating patterns at specific intervals.

The ACF of the log-differenced series exhibits a plot where all autocorrelation values are within the boundary except the first one, which is indicative of a well-behaved and possibly stationary time series. This again is useful in deciding the answer to the next question.

Question 4

In [ ]:
```
from statsmodels.tsa.stattools import adfuller
adf_test = adfuller(housing['HOUST'][dates < '2023-10-1'], maxlag = 4, autolag='
print(adf_test)

result = adf_test

# Extract and print the test statistic and p-value
adf_statistic = result[0]
p_value = result[1]

print(f'ADF Statistic: {adf_statistic}')
print(f'p-value: {p_value}')

# Interpret the results
if p_value <= 0.05:
    print('Reject the null hypothesis. The time series is likely stationary.')
else:
    print('Fail to reject the null hypothesis. The time series may be non-statio
```

```
(-2.753900837771641, 0.06516638465994573, 2, 774, {'1%': -3.4388268991356936,
'5%': -2.8652813916285518, '10%': -2.5687622857867782}, 9451.182849190369)
ADF Statistic: -2.753900837771641
p-value: 0.06516638465994573
Fail to reject the null hypothesis. The time series may be non-stationary.
```

The data is already seasonally adjusted, and the plots do not exhibit a trend or drift in its behavior. Therefore we do not consider drift or trend when performing the test. The ADF test indicates that the original time series is non-stationary, which is consistent with the earlier interpretations.

In [ ]:
```
housing = housing.dropna()
np.roll(housing['DifLogHousing'], 1)
adf_test = adfuller(housing['DifLogHousing'], maxlag = 4, autolag='BIC')
print(adf_test)

result = adf_test

# Extract and print the test statistic and p-value
adf_statistic = result[0]
p_value = result[1]

print(f'ADF Statistic: {adf_statistic}')
```

```
print(f'p-value: {p_value}')

# Interpret the results
if p_value <= 0.05:
    print('Reject the null hypothesis. The time series is likely stationary.')
else:
    print('Fail to reject the null hypothesis. The time series may be non-statio
```

```
(-24.778345325326132, 0.0, 1, 775, {'1%': -3.4388159246521433, '5%': -2.865276554
696385, '10%': -2.5687597090530696}, 5389.060238121698)
ADF Statistic: -24.778345325326132
p-value: 0.0
Reject the null hypothesis. The time series is likely stationary.
```

The log-differenced series also does not show any indication of drift or trend, therefore we do the ADF test above.

Again as shown in the earlier questions, the log-differenced series clearly exhibits stationary behavior, which is confirmed by the ADF test above. Therefore, we should consider the log-difference when modelling the series.

Question 5

In [ ]:
```
housing.index = pd.date_range(start='1959-01-01', periods=len(housing), freq='MS

ar1_model = AutoReg(housing['DifLogHousing'], lags=1, missing = "drop").fit()
print(ar1_model.summary())

ar2_model = AutoReg(housing['DifLogHousing'], lags=2, missing = "drop").fit()
print(ar2_model.summary())

ar3_model = AutoReg(housing['DifLogHousing'], lags=3, missing = "drop").fit()
print(ar3_model.summary())

ar4_model = AutoReg(housing['DifLogHousing'], lags=4, missing = "drop").fit()
print(ar4_model.summary())

bic_ar1 = ar1_model.bic
print('The BIC of the AR(1) model is: ' + str(bic_ar1))

bic_ar2 = ar2_model.bic
print('The BIC of the AR(2) model is: ' + str(bic_ar2))

bic_ar3 = ar3_model.bic
print('The BIC of the AR(3) model is: ' + str(bic_ar3))

bic_ar4 = ar4_model.bic
print('The BIC of the AR(4) model is: ' + str(bic_ar4))
```

```
                            AutoReg Model Results
==============================================================================
Dep. Variable:          DifLogHousing   No. Observations:                  777
Model:                     AutoReg(1)   Log Likelihood               -2700.776
Method:               Conditional MLE   S.D. of innovations              7.857
Date:                Tue, 12 Dec 2023   AIC                           5407.551
Time:                        10:58:33   BIC                           5421.514
Sample:                    02-01-1959   HQIC                          5412.923
                         - 09-01-2023
=================================================================================
===
                       coef    std err          z      P>|z|      [0.025      0.9
75]
---------------------------------------------------------------------------------
---
const               -0.0333      0.282     -0.118      0.906      -0.586       0.
520
DifLogHousing.L1    -0.3056      0.034     -8.939      0.000      -0.373      -0.
239
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -3.2727           +0.0000j            3.2727            0.5000
-----------------------------------------------------------------------------
                            AutoReg Model Results
==============================================================================
Dep. Variable:          DifLogHousing   No. Observations:                  777
Model:                     AutoReg(2)   Log Likelihood               -2693.958
Method:               Conditional MLE   S.D. of innovations              7.824
Date:                Tue, 12 Dec 2023   AIC                           5395.916
Time:                        10:58:33   BIC                           5414.527
Sample:                    03-01-1959   HQIC                          5403.076
                         - 09-01-2023
=================================================================================
===
                       coef    std err          z      P>|z|      [0.025      0.9
75]
---------------------------------------------------------------------------------
---
const               -0.0337      0.281     -0.120      0.904      -0.585       0.
517
DifLogHousing.L1    -0.3356      0.036     -9.390      0.000      -0.406      -0.
266
DifLogHousing.L2    -0.0985      0.036     -2.757      0.006      -0.169      -0.
028
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -1.7030           -2.6921j            3.1855           -0.3398
AR.2           -1.7030           +2.6921j            3.1855            0.3398
-----------------------------------------------------------------------------
                            AutoReg Model Results
==============================================================================
Dep. Variable:          DifLogHousing   No. Observations:                  777
Model:                     AutoReg(3)   Log Likelihood               -2690.920
Method:               Conditional MLE   S.D. of innovations              7.828
Date:                Tue, 12 Dec 2023   AIC                           5391.840
Time:                        10:58:33   BIC                           5415.098
```

```
Sample:                    04-01-1959   HQIC                          5400.789
                         - 09-01-2023
==================================================================================
===
                       coef    std err          z      P>|z|      [0.025      0.9
75]
----------------------------------------------------------------------------------
---
const               -0.0302      0.281     -0.107      0.914      -0.582       0.
521
DifLogHousing.L1    -0.3358      0.036     -9.343      0.000      -0.406      -0.
265
DifLogHousing.L2    -0.0987      0.038     -2.614      0.009      -0.173      -0.
025
DifLogHousing.L3    -0.0003      0.036     -0.008      0.993      -0.071       0.
070
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -1.7041          -2.7089j           3.2003           -0.3394
AR.2           -1.7041          +2.7089j           3.2003            0.3394
AR.3         -326.0398          -0.0000j         326.0398           -0.5000
-----------------------------------------------------------------------------
                            AutoReg Model Results
==============================================================================
Dep. Variable:          DifLogHousing   No. Observations:                  777
Model:                     AutoReg(4)   Log Likelihood               -2686.621
Method:               Conditional MLE   S.D. of innovations              7.820
Date:                Tue, 12 Dec 2023   AIC                           5385.241
Time:                        10:58:33   BIC                           5413.143
Sample:                    05-01-1959   HQIC                          5395.977
                         - 09-01-2023
==================================================================================
===
                       coef    std err          z      P>|z|      [0.025      0.9
75]
----------------------------------------------------------------------------------
---
const               -0.0200      0.281     -0.071      0.943      -0.571       0.
531
DifLogHousing.L1    -0.3362      0.036     -9.363      0.000      -0.407      -0.
266
DifLogHousing.L2    -0.0943      0.038     -2.491      0.013      -0.169      -0.
020
DifLogHousing.L3     0.0162      0.038      0.426      0.670      -0.058       0.
090
DifLogHousing.L4     0.0493      0.036      1.372      0.170      -0.021       0.
120
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -2.0327          -0.0000j           2.0327           -0.5000
AR.2           -0.4430          -1.9113j           1.9620           -0.2863
AR.3           -0.4430          +1.9113j           1.9620            0.2863
AR.4            2.5911          -0.0000j           2.5911           -0.0000
-----------------------------------------------------------------------------
The BIC of the AR(1) model is: 5421.513875092892
The BIC of the AR(2) model is: 5414.527020796055
```

```
The BIC of the AR(3) model is: 5415.098225697888
The BIC of the AR(4) model is: 5413.143061241818
```

We consider AR(1) to AR(4) and compare the BIC values. Notably, the BIC values decrease as the lags increase. The data is monthly, therefore we should consider other lags. The code below considers the BIC values of AR(1) to AR(35) models and selects the model with the lowest BIC.

In [ ]:
```
endog_data = housing['DifLogHousing']

endog_data.index = pd.date_range(start='1959-01-01', periods=len(endog_data), fr

# Initialize variables to track the minimum BIC and corresponding order
min_bic = float('inf')
best_order = None

# Iterate through AR orders 1 to 30
for order in range(1, 35):
    # Fit AutoReg model
    ar_model = AutoReg(endog_data, lags=order, missing="drop").fit()

    # Get BIC value
    bic_value = ar_model.bic

    # Update minimum BIC and corresponding order if a new minimum is found
    if bic_value < min_bic:
        min_bic = bic_value
        best_order = order

# Fit the best model
best_model = AutoReg(endog_data, lags=best_order, missing="drop").fit()

# Print summary of the best model
print(best_model.summary())

# Print the BIC value of the best model
print(f'The BIC of the best AR({best_order}) model is: {min_bic}')
```

```
                            AutoReg Model Results
==============================================================================
Dep. Variable:          DifLogHousing   No. Observations:                  777
Model:                   AutoReg(29)   Log Likelihood               -2565.816
Method:               Conditional MLE   S.D. of innovations              7.473
Date:                Tue, 12 Dec 2023   AIC                           5193.633
Time:                        10:58:33   BIC                           5336.772
Sample:                    06-01-1961   HQIC                          5248.795
                         - 09-01-2023
==============================================================================
====
                        coef    std err          z      P>|z|      [0.025      0.
975]
------------------------------------------------------------------------------
----
const                -0.0040      0.273     -0.015      0.988      -0.540
0.532
DifLogHousing.L1     -0.3163      0.036     -8.674      0.000      -0.388      -
0.245
DifLogHousing.L2     -0.0777      0.038     -2.033      0.042      -0.153      -
0.003
DifLogHousing.L3      0.0388      0.038      1.012      0.311      -0.036
0.114
DifLogHousing.L4      0.0783      0.038      2.042      0.041       0.003
0.153
DifLogHousing.L5      0.0823      0.038      2.141      0.032       0.007
0.158
DifLogHousing.L6      0.0701      0.038      1.837      0.066      -0.005
0.145
DifLogHousing.L7     -0.0072      0.038     -0.189      0.850      -0.082
0.068
DifLogHousing.L8     -0.0045      0.038     -0.117      0.907      -0.079
0.070
DifLogHousing.L9      0.0175      0.038      0.458      0.647      -0.057
0.092
DifLogHousing.L10    -0.0301      0.038     -0.791      0.429      -0.105
0.045
DifLogHousing.L11     0.0067      0.038      0.175      0.861      -0.068
0.081
DifLogHousing.L12    -0.1324      0.038     -3.475      0.001      -0.207      -
0.058
DifLogHousing.L13     0.0432      0.038      1.125      0.261      -0.032
0.118
DifLogHousing.L14     0.0487      0.038      1.269      0.205      -0.027
0.124
DifLogHousing.L15     0.0729      0.038      1.898      0.058      -0.002
0.148
DifLogHousing.L16     0.0506      0.038      1.325      0.185      -0.024
0.125
DifLogHousing.L17     0.0108      0.038      0.283      0.777      -0.064
0.086
DifLogHousing.L18    -0.0050      0.038     -0.131      0.896      -0.079
0.069
DifLogHousing.L19     0.0015      0.038      0.039      0.969      -0.073
0.076
DifLogHousing.L20    -0.0391      0.038     -1.032      0.302      -0.113
0.035
DifLogHousing.L21    -0.0132      0.038     -0.349      0.727      -0.087
0.061
DifLogHousing.L22     0.0298      0.038      0.787      0.431      -0.044
```

```
0.104
DifLogHousing.L23     0.0126      0.038      0.332      0.740      -0.062
0.087
DifLogHousing.L24    -0.1519      0.038     -4.011      0.000      -0.226      -
0.078
DifLogHousing.L25    -0.0216      0.038     -0.567      0.570      -0.096
0.053
DifLogHousing.L26     0.0545      0.038      1.432      0.152      -0.020
0.129
DifLogHousing.L27     0.0288      0.038      0.758      0.448      -0.046
0.103
DifLogHousing.L28     0.0239      0.038      0.630      0.529      -0.050
0.098
DifLogHousing.L29    -0.0272      0.036     -0.753      0.451      -0.098
0.043
                                    Roots
==============================================================================
                  Real          Imaginary           Modulus         Frequency
------------------------------------------------------------------------------
AR.1           -0.9996           -0.4187j            1.0837           -0.4369
AR.2           -0.9996           +0.4187j            1.0837            0.4369
AR.3           -1.0544           -0.2068j            1.0745           -0.4692
AR.4           -1.0544           +0.2068j            1.0745            0.4692
AR.5           -1.2157           -0.0000j            1.2157           -0.5000
AR.6           -0.8327           -0.6727j            1.0705           -0.3919
AR.7           -0.8327           +0.6727j            1.0705            0.3919
AR.8           -0.6749           -0.8260j            1.0667           -0.3590
AR.9           -0.6749           +0.8260j            1.0667            0.3590
AR.10          -0.4137           -0.9864j            1.0697           -0.3132
AR.11          -0.4137           +0.9864j            1.0697            0.3132
AR.12          -0.1417           -1.0564j            1.0659           -0.2712
AR.13          -0.1417           +1.0564j            1.0659            0.2712
AR.14          -0.5025           -1.3380j            1.4292           -0.3072
AR.15          -0.5025           +1.3380j            1.4292            0.3072
AR.16           0.1526           -1.0359j            1.0471           -0.2267
AR.17           0.1526           +1.0359j            1.0471            0.2267
AR.18           0.3830           -0.9878j            1.0594           -0.1911
AR.19           0.3830           +0.9878j            1.0594            0.1911
AR.20           0.6482           -0.8535j            1.0717           -0.1466
AR.21           0.6482           +0.8535j            1.0717            0.1466
AR.22           0.8438           -0.7292j            1.1153           -0.1134
AR.23           0.8438           +0.7292j            1.1153            0.1134
AR.24           1.0017           -0.4362j            1.0926           -0.0654
AR.25           1.0017           +0.4362j            1.0926            0.0654
AR.26           1.1335           -0.1358j            1.1416           -0.0190
AR.27           1.1335           +0.1358j            1.1416            0.0190
AR.28           1.5041           -0.3111j            1.5360           -0.0325
AR.29           1.5041           +0.3111j            1.5360            0.0325
------------------------------------------------------------------------------
The BIC of the best AR(29) model is: 5336.772441130901

The BIC of the best AR(29) model is: 5336.772441130901
```

The model with the lowest BIC is the AR(29) model, therefore we use this as the model with the best goodness-of-fit.

However, not all of the coefficients are statistically significant (only are 7 are statistically significant), and the coefficient is not significant either. The coefficient values do not have any causal interpretation.

Question 6

In [ ]:
```
ur = fred.get_series('UNRATE', observation_start = '1959-01-01', observation_end
ur = pd.DataFrame(ur)
ur.columns = ['UNRATE']
ur = ur.dropna()
print(ur.head(3))
print(ur.tail(3))
```

```
            UNRATE
1959-01-01     6.0
1959-02-01     5.9
1959-03-01     5.6
            UNRATE
2023-08-01     3.8
2023-09-01     3.8
2023-10-01     3.9
```

Data: Unemployment Rate

Source: U.S. Bureau of Labor Statistics Release: Employment Situation
Units: Percent, Seasonally Adjusted

Frequency: Monthly

The unemployment rate represents the number of unemployed as a percentage of the labor force. Labor force data are restricted to people 16 years of age and older, who currently reside in 1 of the 50 states or the District of Columbia, who do not reside in institutions (e.g., penal and mental facilities, homes for the aged), and who are not on active duty in the Armed Forces.

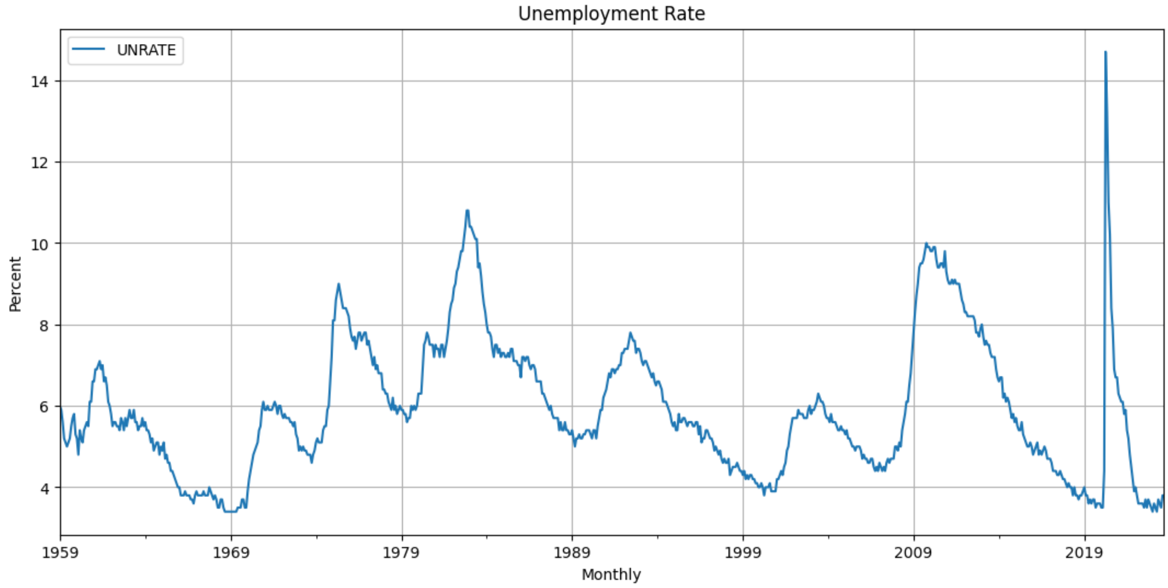
This rate is also defined as the U-3 measure of labor underutilization.

The series comes from the 'Current Population Survey (Household Survey)'

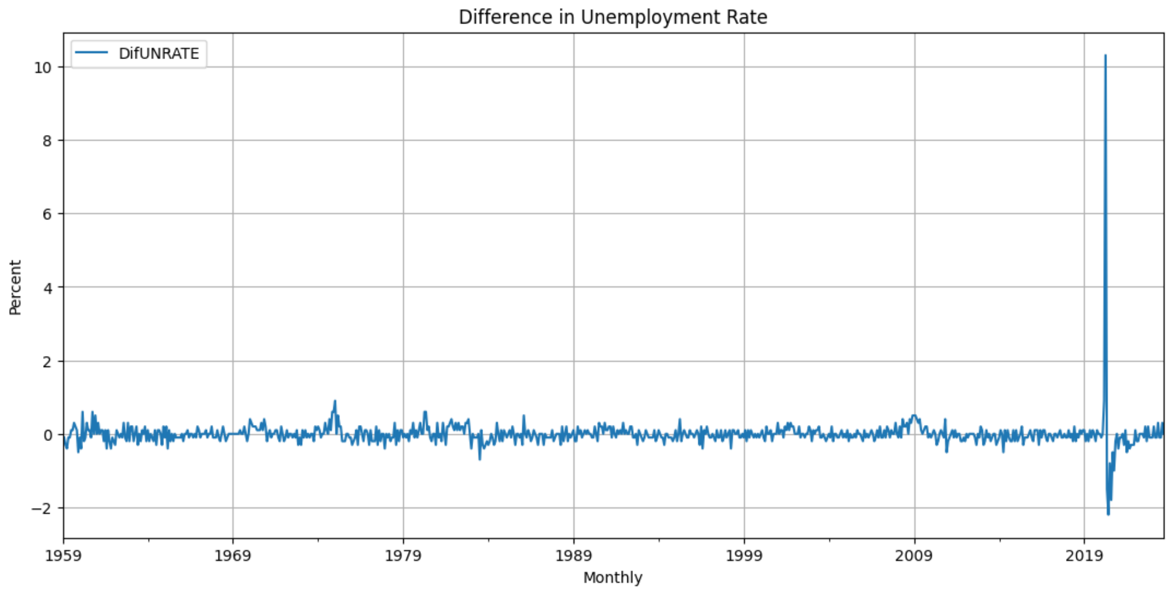
The source code is: LNS14000000

Suggested Citation: U.S. Bureau of Labor Statistics, Unemployment Rate [UNRATE], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/UNRATE, December 11, 2023.

In [ ]:
```
plt.figure(figsize=(13, 6))
dates = ur.index
ur['UNRATE'][dates < '2023-10-1'].plot()
plt.title('Unemployment Rate')
plt.xlabel('Monthly')
plt.ylabel('Percent')
plt.legend()
plt.grid(True)
plt.show()
```



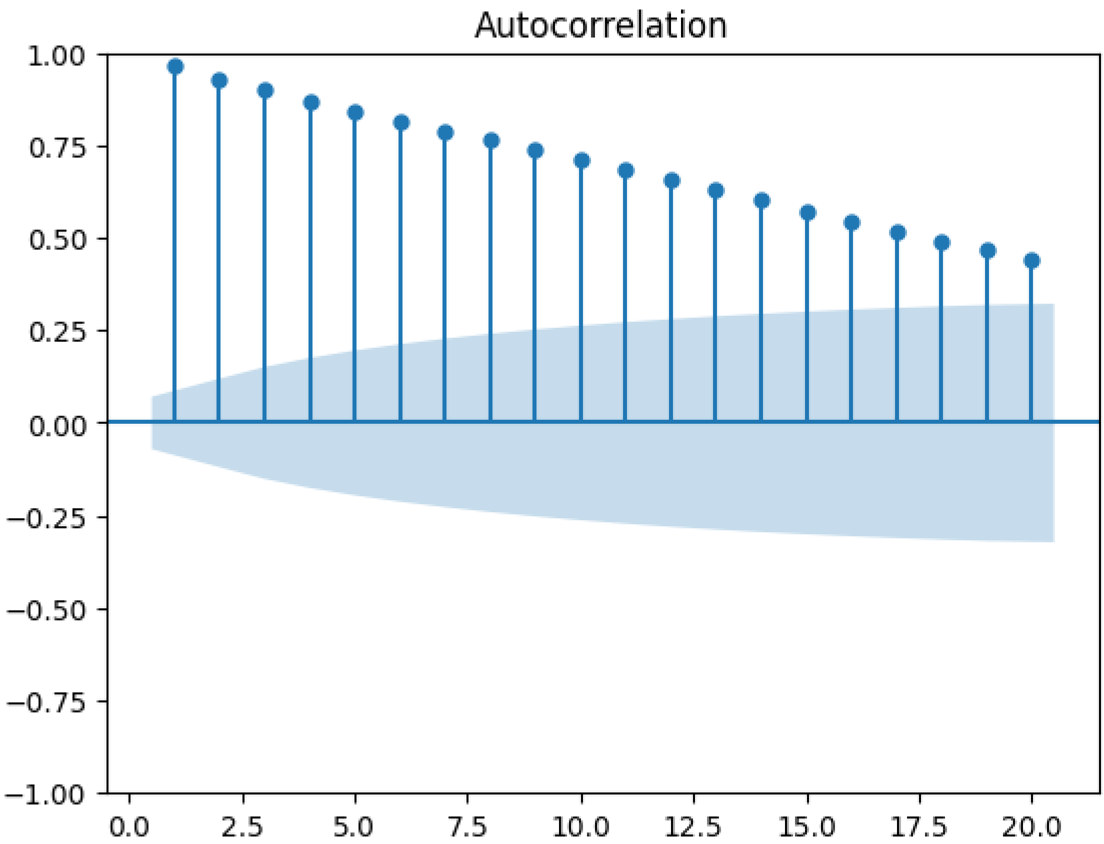
In [ ]:
```
ur['DifUNRATE'] = ur['UNRATE'] - ur['UNRATE'].shift(1)
plt.figure(figsize=(13, 6))
ur['DifUNRATE'][dates < '2023-10-1'].plot()
plt.title('Difference in Unemployment Rate')
plt.xlabel('Monthly')
plt.ylabel('Percent')
plt.legend()
plt.grid(True)
plt.show()
```



The data is already in percentage values and is seasonally adjusted. Therefore taking the log values does not make a large difference, and we will use the level data. Notably although the COVID-19 pandemic appears to have had a significant impact on the unemployment rate, removing the data from the sample does not make a large difference in any of the tests, therefore we use the full sample.

In [ ]:
```
plot_acf(ur['UNRATE'], lags=20, missing = 'drop',
          zero = False)
plt.show()

adf_test = adfuller(ur['UNRATE'], maxlag = 4, autolag='BIC')
```

```
print(adf_test)

result = adf_test

# Extract and print the test statistic and p-value
adf_statistic = result[0]
p_value = result[1]

print(f'ADF Statistic: {adf_statistic}')
print(f'p-value: {p_value}')

# Interpret the results
if p_value <= 0.05:
    print('Reject the null hypothesis. The time series is likely stationary.')
else:
    print('Fail to reject the null hypothesis. The time series may be non-statio
```

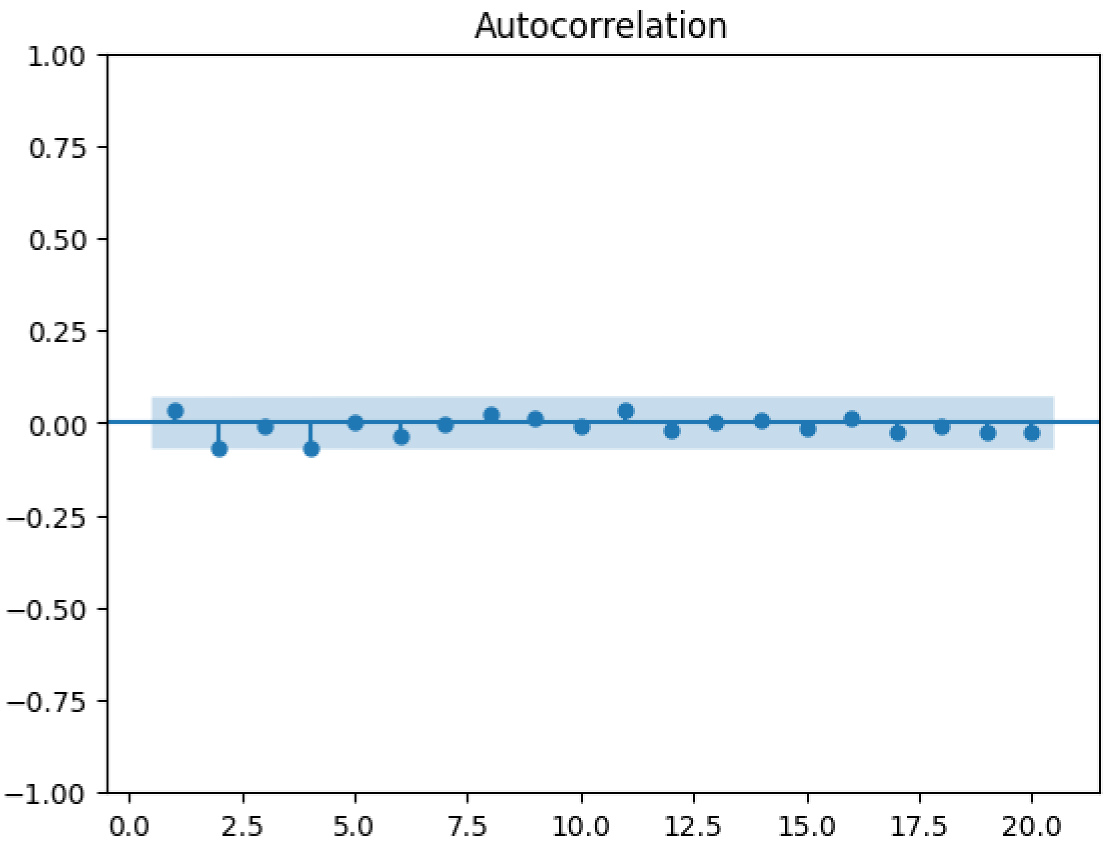



```
(-3.496109211181302, 0.008087581816052507, 0, 777, {'1%': -3.4387940607132887,
'5%': -2.8652669182555943, '10%': -2.5687545755297494}, 894.1671056421721)
ADF Statistic: -3.496109211181302
p-value: 0.008087581816052507
Reject the null hypothesis. The time series is likely stationary.
```

In [ ]:

```
plot_acf(ur['DifUNRATE'], lags=20, missing = 'drop',
          zero = False)
plt.show()

ur = ur.dropna()
np.roll(ur['DifUNRATE'], 1)
adf_test = adfuller(ur['DifUNRATE'], maxlag = 4, autolag='BIC')
print(adf_test)

result = adf_test

# Extract and print the test statistic and p-value
```

```
adf_statistic = result[0]
p_value = result[1]

print(f'ADF Statistic: {adf_statistic}')
print(f'p-value: {p_value}')

# Interpret the results
if p_value <= 0.05:
    print('Reject the null hypothesis. The time series is likely stationary.')
else:
    print('Fail to reject the null hypothesis. The time series may be non-statio
```




```
(-26.929710669493147, 0.0, 0, 776, {'1%': -3.438804978547988, '5%': -2.8652717302
548396, '10%': -2.5687571389759802}, 905.3931353056731)
ADF Statistic: -26.929710669493147
p-value: 0.0
Reject the null hypothesis. The time series is likely stationary.
```

There does not appear to be drift or trend in this series either. Both ACFs indicate that the series is stationary, however we will use the differenced data as the graph of the differenced data appears to exhibit more stationary behavior, and the ACF of the original series indicates some non-stationarity in the original series as well.

In [ ]:
```
print(housing.columns)
print(ur.columns)
```

```
Index(['HOUST', 'LogHousing', 'DifLogHousing'], dtype='object')
Index(['UNRATE', 'DifUNRATE'], dtype='object')
```

In [ ]:
```
dataset = pd.concat([housing, ur], axis = 1).dropna()
dataset.index = pd.date_range(start='1959-01-01', periods=len(dataset), freq='MS

ardl_model1 = ardl.ARDL(endog = dataset['DifLogHousing'], exog = dataset[['DifUN
print(ardl_model1.summary())
```

```
ardl_model = ardl.ARDL(endog = dataset['DifLogHousing'], exog = dataset[['DifUNR
print(ardl_model.summary())
```

```
                              ARDL Model Results
==============================================================================
Dep. Variable:          DifLogHousing   No. Observations:                  776
Model:                    ARDL(29, 1)   Log Likelihood               -2560.752
Method:               Conditional MLE   S.D. of innovations              7.457
Date:                Tue, 12 Dec 2023   AIC                           5185.505
Time:                        10:58:34   BIC                           5333.219
Sample:                    06-01-1961   HQIC                          5242.433
                         - 08-01-2023
=================================================================================
====
                        coef    std err          z      P>|z|      [0.025      0.
975]
---------------------------------------------------------------------------------
----
const                 0.0014      0.279      0.005      0.996      -0.546
0.549
DifLogHousing.L1     -0.3121      0.037     -8.365      0.000      -0.385      -
0.239
DifLogHousing.L2     -0.0577      0.040     -1.435      0.152      -0.137
0.021
DifLogHousing.L3      0.0561      0.040      1.401      0.162      -0.023
0.135
DifLogHousing.L4      0.0850      0.039      2.169      0.030       0.008
0.162
DifLogHousing.L5      0.0808      0.039      2.062      0.040       0.004
0.158
DifLogHousing.L6      0.0652      0.039      1.673      0.095      -0.011
0.142
DifLogHousing.L7     -0.0071      0.039     -0.181      0.857      -0.084
0.069
DifLogHousing.L8     -0.0034      0.039     -0.086      0.931      -0.080
0.073
DifLogHousing.L9      0.0208      0.039      0.535      0.593      -0.056
0.097
DifLogHousing.L10    -0.0279      0.039     -0.717      0.474      -0.104
0.049
DifLogHousing.L11     0.0108      0.039      0.277      0.782      -0.066
0.087
DifLogHousing.L12    -0.1251      0.039     -3.208      0.001      -0.202      -
0.049
DifLogHousing.L13     0.0459      0.039      1.171      0.242      -0.031
0.123
DifLogHousing.L14     0.0536      0.039      1.366      0.172      -0.023
0.131
DifLogHousing.L15     0.0745      0.039      1.902      0.058      -0.002
0.151
DifLogHousing.L16     0.0527      0.039      1.342      0.180      -0.024
0.130
DifLogHousing.L17     0.0104      0.039      0.267      0.789      -0.066
0.087
DifLogHousing.L18    -0.0008      0.039     -0.021      0.983      -0.077
0.075
DifLogHousing.L19     0.0049      0.039      0.126      0.900      -0.071
0.081
DifLogHousing.L20    -0.0365      0.039     -0.945      0.345      -0.112
0.039
DifLogHousing.L21    -0.0107      0.039     -0.276      0.783      -0.087
0.065
DifLogHousing.L22     0.0297      0.039      0.769      0.442      -0.046
```

```
0.106
DifLogHousing.L23     0.0169      0.039      0.436      0.663      -0.059
0.093
DifLogHousing.L24    -0.1483      0.039     -3.839      0.000      -0.224      -
0.072
DifLogHousing.L25    -0.0223      0.039     -0.574      0.566      -0.099
0.054
DifLogHousing.L26     0.0576      0.039      1.485      0.138      -0.019
0.134
DifLogHousing.L27     0.0284      0.039      0.731      0.465      -0.048
0.105
DifLogHousing.L28     0.0214      0.039      0.552      0.581      -0.055
0.097
DifLogHousing.L29    -0.0316      0.037     -0.859      0.391      -0.104
0.041
DifUNRATE.L1          1.3728      0.679      2.023      0.043       0.040
2.705
=================================================================================
====
                              ARDL Model Results
==============================================================================
Dep. Variable:          DifLogHousing   No. Observations:                  776
Model:                   ARDL(29, 4)   Log Likelihood               -2550.175
Method:              Conditional MLE   S.D. of innovations              7.352
Date:               Tue, 12 Dec 2023   AIC                           5170.350
Time:                       10:58:34   BIC                           5331.912
Sample:                   06-01-1961   HQIC                          5232.615
                        - 08-01-2023
=================================================================================
====
                        coef    std err          z      P>|z|      [0.025      0.
975]
---------------------------------------------------------------------------------
----
const                 0.0180      0.275      0.065      0.948      -0.523
0.559
DifLogHousing.L1     -0.3338      0.037     -8.929      0.000      -0.407      -
0.260
DifLogHousing.L2     -0.0685      0.040     -1.695      0.091      -0.148
0.011
DifLogHousing.L3      0.0899      0.041      2.195      0.028       0.010
0.170
DifLogHousing.L4      0.1369      0.041      3.335      0.001       0.056
0.217
DifLogHousing.L5      0.1261      0.041      3.052      0.002       0.045
0.207
DifLogHousing.L6      0.0859      0.040      2.156      0.031       0.008
0.164
DifLogHousing.L7     -0.0090      0.039     -0.231      0.818      -0.085
0.067
DifLogHousing.L8     -0.0062      0.039     -0.161      0.872      -0.082
0.069
DifLogHousing.L9      0.0196      0.039      0.507      0.612      -0.056
0.095
DifLogHousing.L10    -0.0193      0.039     -0.502      0.616      -0.095
0.056
DifLogHousing.L11     0.0192      0.039      0.500      0.618      -0.056
0.095
DifLogHousing.L12    -0.1117      0.039     -2.889      0.004      -0.188      -
0.036
```

```
DifLogHousing.L13      0.0629      0.039      1.617      0.106      -0.013      
0.139
DifLogHousing.L14      0.0670      0.039      1.718      0.086      -0.010      
0.143
DifLogHousing.L15      0.0928      0.039      2.381      0.018       0.016      
0.169
DifLogHousing.L16      0.0634      0.039      1.628      0.104      -0.013      
0.140
DifLogHousing.L17      0.0222      0.039      0.575      0.565      -0.054      
0.098
DifLogHousing.L18      0.0044      0.038      0.116      0.908      -0.071      
0.080
DifLogHousing.L19      0.0163      0.038      0.426      0.670      -0.059      
0.092
DifLogHousing.L20     -0.0259      0.038     -0.678      0.498      -0.101      
0.049
DifLogHousing.L21     -0.0003      0.038     -0.008      0.994      -0.075      
0.075
DifLogHousing.L22      0.0389      0.038      1.019      0.309      -0.036      
0.114
DifLogHousing.L23      0.0214      0.038      0.558      0.577      -0.054      
0.097
DifLogHousing.L24     -0.1371      0.038     -3.584      0.000      -0.212      -
0.062
DifLogHousing.L25     -0.0162      0.038     -0.422      0.673      -0.092      
0.059
DifLogHousing.L26      0.0590      0.038      1.537      0.125      -0.016      
0.134
DifLogHousing.L27      0.0365      0.038      0.951      0.342      -0.039      
0.112
DifLogHousing.L28      0.0239      0.038      0.624      0.533      -0.051      
0.099
DifLogHousing.L29     -0.0297      0.036     -0.817      0.414      -0.101      
0.042
DifUNRATE.L1           1.4902      0.673      2.213      0.027       0.168      
2.812
DifUNRATE.L2           2.7786      0.675      4.120      0.000       1.454      
4.103
DifUNRATE.L3           1.0498      0.682      1.540      0.124      -0.289      
2.388
DifUNRATE.L4           1.1564      0.682      1.695      0.091      -0.183      
2.496
==================================================================================
====
```

The ADL(29, 4) model has a lower BIC as compared to the ADL(29, 1) model, therefore we pick the former as the model we will use.

When comparing the goodness-of-fit of the ADL model with the AR model, we can also compare BIC. The ADL(29, 4) model has a lower BIC than the AR(29) model, indicating that it is more suitable to model the series.

Question 7

In [ ]:
```
ardl_model_params = ardl_model.params
last_29_values_df1 = housing.iloc[-29:]['DifLogHousing']
last_4_values_df2 = ur.iloc[-4:]['DifUNRATE']
const_coefficient = ardl_model_params['const']
```

```
# Coefficients for DifLogHousing series
lag_coefficients_df1 = [
    ardl_model_params[f'DifLogHousing.L{i}'] for i in range(1, 30)
]

# Coefficients for DifUNRATE series
lag_coefficients_df2 = [
    ardl_model_params[f'DifUNRATE.L{i}'] for i in range(1, 5)
]

# Multiply each value by its respective coefficient
last_29_values_df1 *= lag_coefficients_df1
last_4_values_df2 *= lag_coefficients_df2

sum_df1 = last_29_values_df1.sum()
sum_df2 = last_4_values_df2.sum()

result = (
    const_coefficient +
    sum_df1 +
    sum_df2
)

# Display the result
print(result)
```

```
2.8005918097093456
```

This value is the first difference in the natural log of the housing start data series. Therefore we do the following to obtain the actual value:

In [ ]:

```
final_point = housing['HOUST'].iloc[-1]
remove_log = np.exp(result)
original_value = remove_log + final_point
print(f'The original value is: {original_value}')
```

```
The original value is: 1388.4543817530673
```

Therefore the forecast indicates that there will be an increase of approximately 16,000 in the total units of housing starts in the next period (1st November 2023). This forecast seems quite reasonable as it is consistent with the increases displayed in the previous months.