TIME SERIES MODELING & ANALYSIS

Instructor Name: Reza Jafari

HW#: 7

Submitted by: Dinesh Kumar Padmanabhan

Date: 06-Nov-2020

ANSWERS TO ASKED QUESTIONS

File - unknown

```
C:\ProgramData\Anaconda3\python.exe "C:\Program Files\
JetBrains\PyCharm 2019.3.1\plugins\python\helpers\pydev\
pydevconsole.py" --mode=client --port=62749

import sys; print('Python %s on %s' % (sys.version, sys.
platform))
sys.path.extend(['C:\\Users\\nsree_000\\Desktop\\Python-
Quiz', 'C:/Users/nsree_000/Desktop/Python-Quiz'])

Python 3.7.4 (default, Aug  9 2019, 18:34:13) [MSC v.1915
64 bit (AMD64)]
Type 'copyright', 'credits' or 'license' for more
information
IPython 7.8.0 -- An enhanced Interactive Python. Type '?'
for help.
PyDev console: using IPython 7.8.0

Python 3.7.4 (default, Aug  9 2019, 18:34:13) [MSC v.1915
64 bit (AMD64)] on win32
In[2]: runfile('C:/Users/nsree_000/Desktop/Python-Quiz/TIME
 SERIES/Homework7.py', wdir='C:/Users/nsree_000/Desktop/
Python-Quiz/TIME SERIES')
Coefficients for AR(1) are [-0.98913539]
MSE of the residual error for AR(1) model: 19081.
163430888555
Q value of residual errors for AR(1) model: 410.
80298794886966
Mean of residual errors for AR(1) model: 10.024060387306319
Variance of residual errors for AR(1) model: 18980.
68164424019
C:\ProgramData\Anaconda3\lib\site-packages\pandas\plotting\
_matplotlib\converter.py:103: FutureWarning: Using an
implicitly registered datetime converter for a matplotlib
plotting method. The converter was registered by pandas on
import. Future versions of pandas will require you to
explicitly register matplotlib converters.

To register the converters:
    >>> from pandas.plotting import
register_matplotlib_converters
    >>> register_matplotlib_converters()
  warnings.warn(msg, FutureWarning)
Coefficients for AR(2) are [-0.85284528 -0.13884815]
MSE of the residual error for AR(2) model: 18760.
390797398282
Q value of residual errors for AR(2) model: 435.
```

File - unknown

```
64526187287566
Mean of residual errors for AR(2) model: 8.65205744304954
Variance of residual errors for AR(2) model: 18685.
532699400454
Coefficients for AR(3) are [-0.75401983  0.49895401 -0.
74432075]
MSE of the residual error for AR(3) model: 8524.
498759432317
Q value of residual errors for AR(3) model: 257.
3538122295325
Mean of residual errors for AR(3) model: 2.285944381872274
Variance of residual errors for AR(3) model: 8519.
273217715303
Coefficients for AR(4) are [-0.18272727  0.10965715 -0.
14155409 -0.78693594]
MSE of the residual error for AR(4) model: 3267.
2350153036955
Q value of residual errors for AR(4) model: 18.
726306655079398
Mean of residual errors for AR(4) model: 0.4748552045113321
Variance of residual errors for AR(4) model: 3267.
009527838444
Coefficients for AR(5) are [-0.22398531  0.09670691 -0.
13425086 -0.7969546   0.05746722]
MSE of the residual error for AR(5) model: 3279.09644713048
Q value of residual errors for AR(5) model: 18.
73168031345683
Mean of residual errors for AR(5) model: 0.5018120401201426
Variance of residual errors for AR(5) model: 3278.
8446318068727
                  AR(1)         AR(2)        AR(3)        AR
(4)        AR(5)
MSE       19081.163431  18760.390797  8524.498759  3267.
235015  3279.096447
Q-value     410.802988    435.645262   257.353812    18.
726307    18.731680
Mean          10.024060      8.652057     2.285944     0.
474855     0.501812
Variance  18980.681644  18685.532699  8519.273218  3267.
009528  3278.844632
Best order number of AR model is: 4 and their coefficients
are: [-0.18272727  0.10965715 -0.14155409 -0.78693594]

```

1- Display the estimated coefficients on the console.

   Coefficients for AR(1) are [-0.98913539]

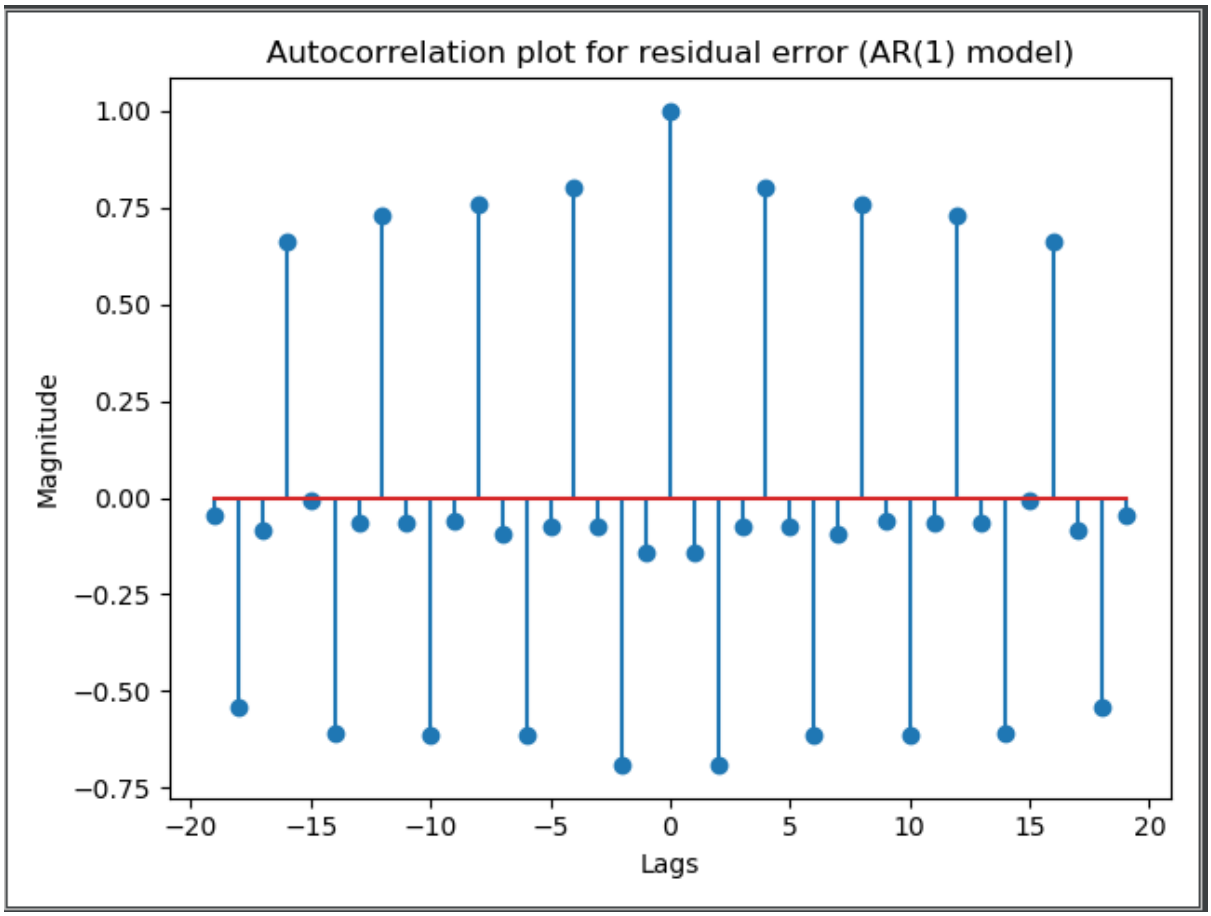
2- Calculate the Mean Square Error (MSE) and display it with an appropriate message on the console.

   MSE of the residual error for AR(1) model: 19081.163430888555

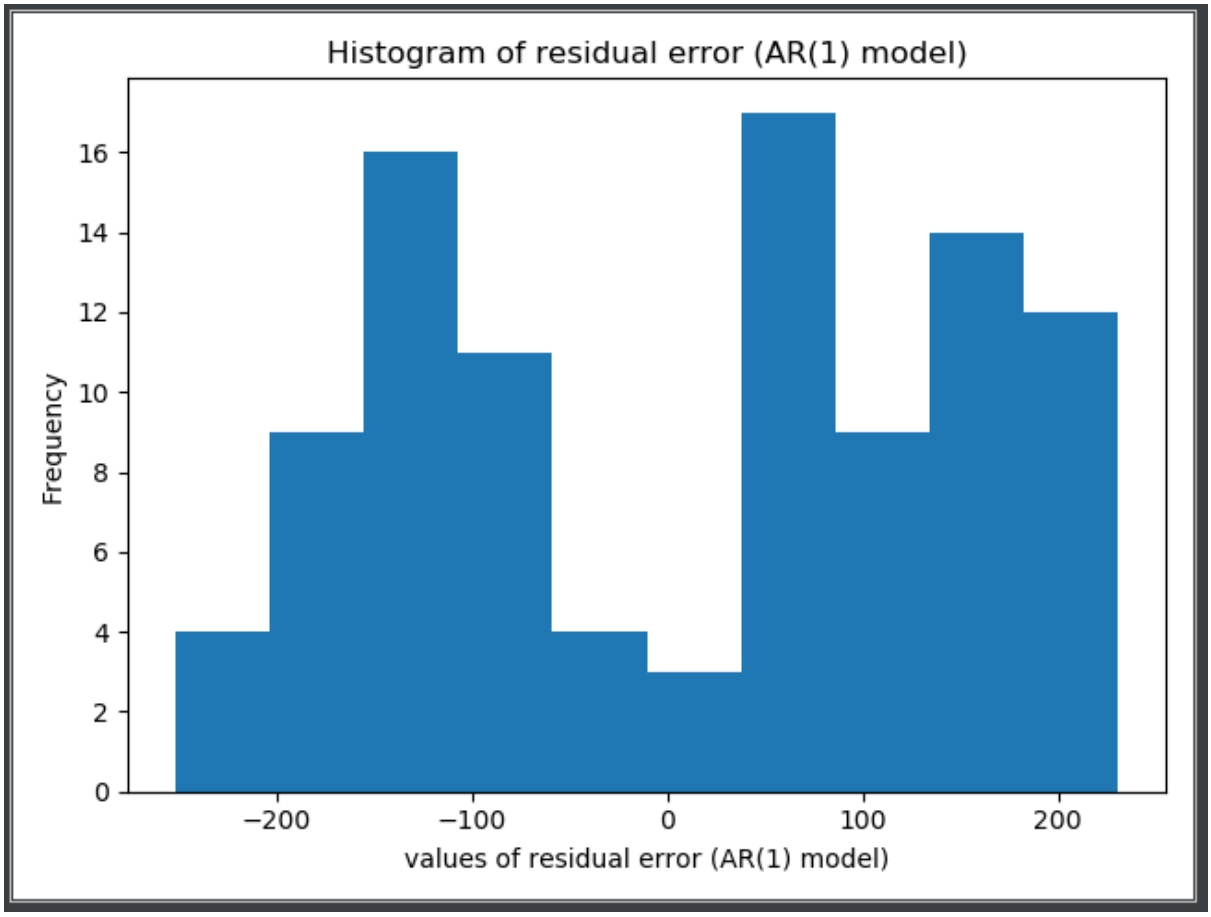
3- Calculate Q and display it with an appropriate message on the console. (20 lags)

   Q value of residual errors for AR(1) model: 410.80298794886966

4- Plot the AFC of residuals. (20 lags)



5- Plot the histogram plot of the residuals.

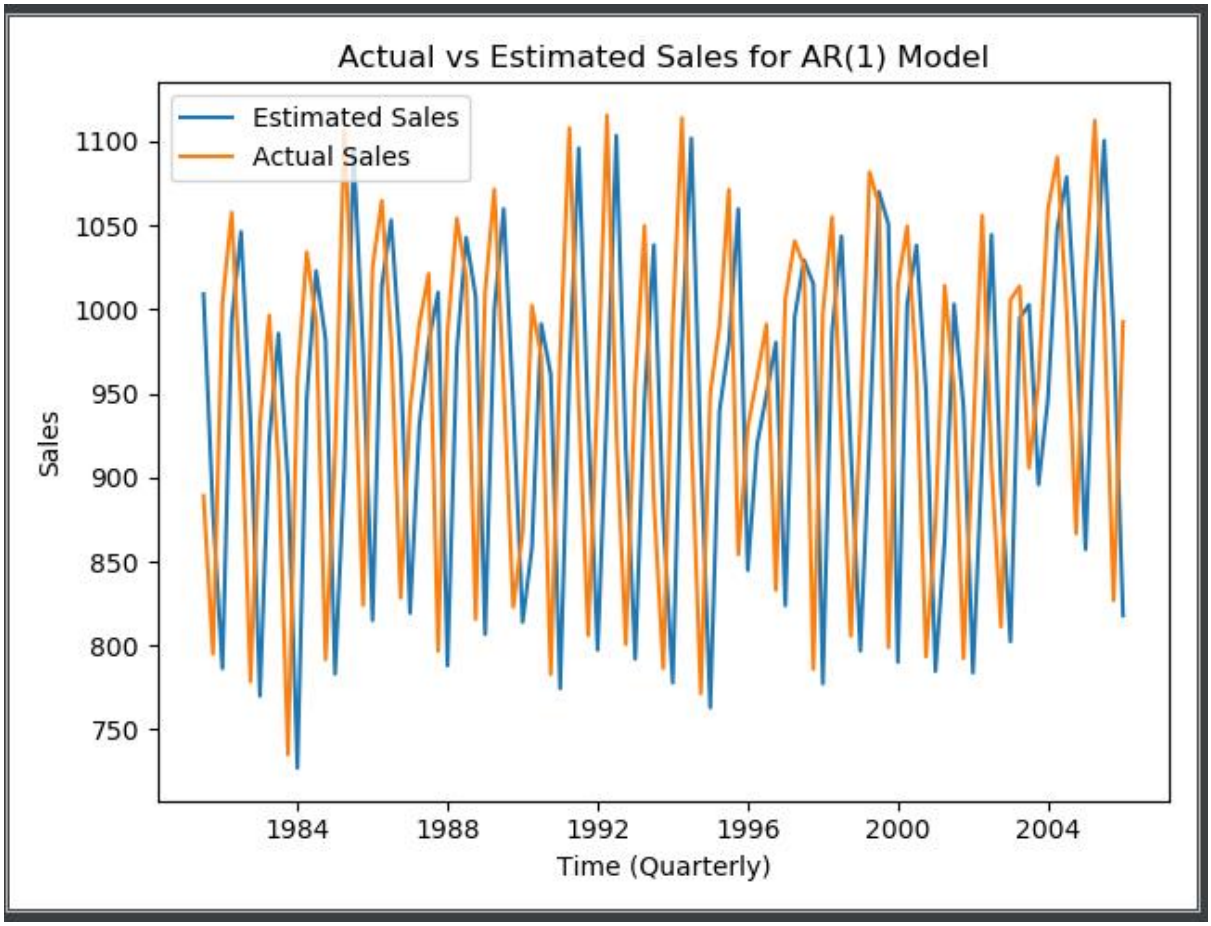


6- Calculate the mean & variance of residuals and display it with an appropriate message on the console.

Mean of residual errors for AR(1) model: 10.024060387306319

Variance of residual errors for AR(1) model: 18980.68164424019

7- Plot the estimated values versus the true values with respect to time.



8- Change the order to 2, 3, 4 and 5 and repeat the step 2 through 8.

AR(2)

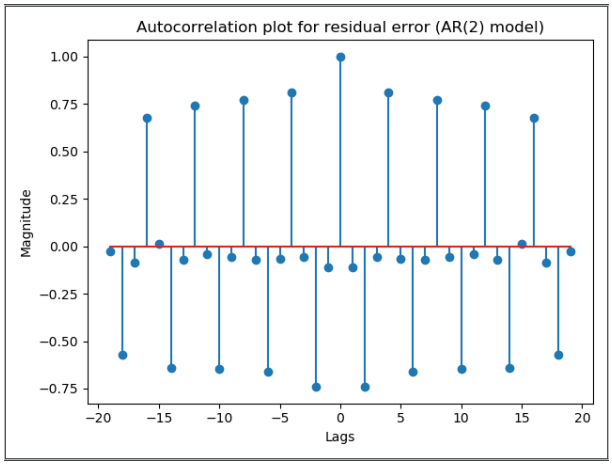
Coefficients for AR(2) are [-0.85284528 -0.13884815]

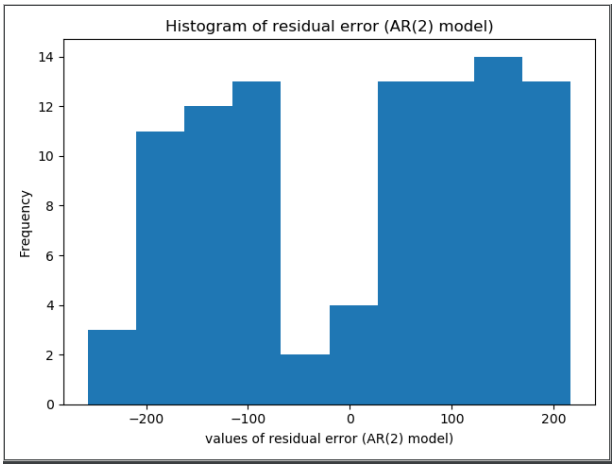
MSE of the residual error for AR(2) model: 18760.390797398282

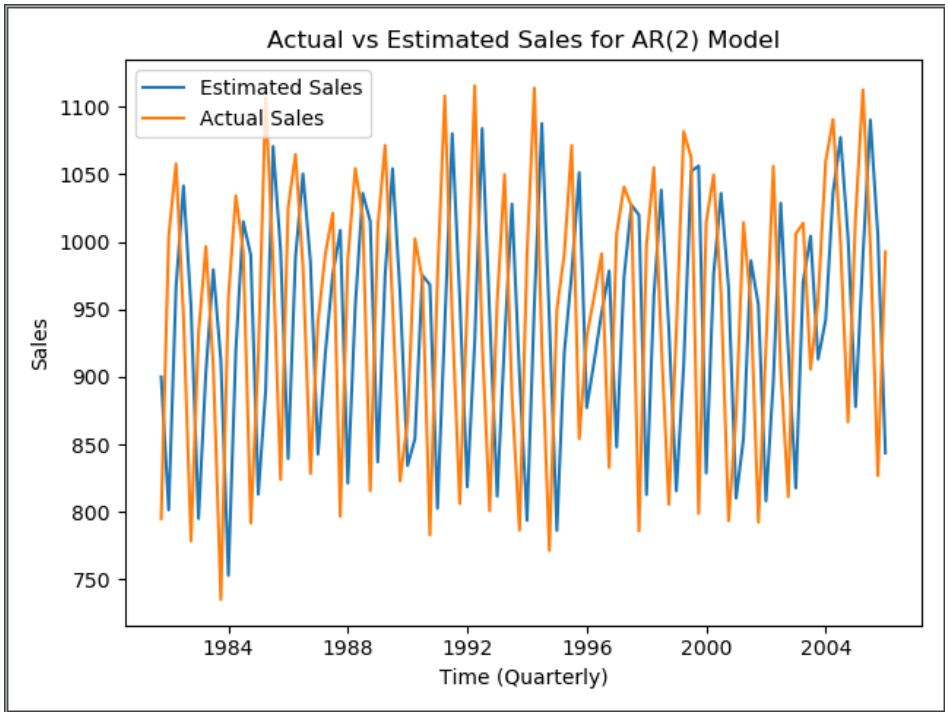
Q value of residual errors for AR(2) model: 435.64526187287566

Mean of residual errors for AR(2) model: 8.65205744304954

Variance of residual errors for AR(2) model: 18685.532699400454







AR(3)

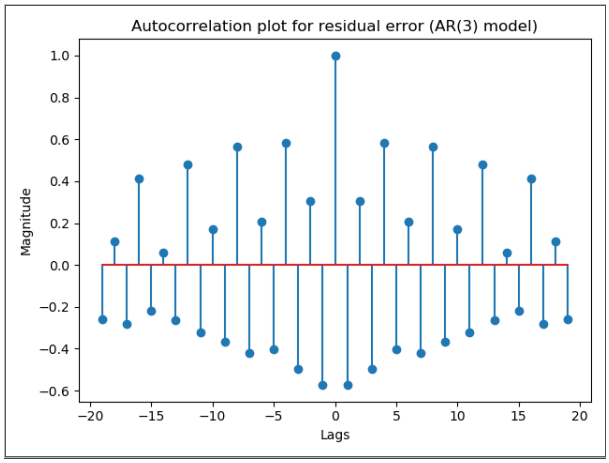
Coefficients for AR(3) are [-0.75401983 0.49895401 -0.74432075]

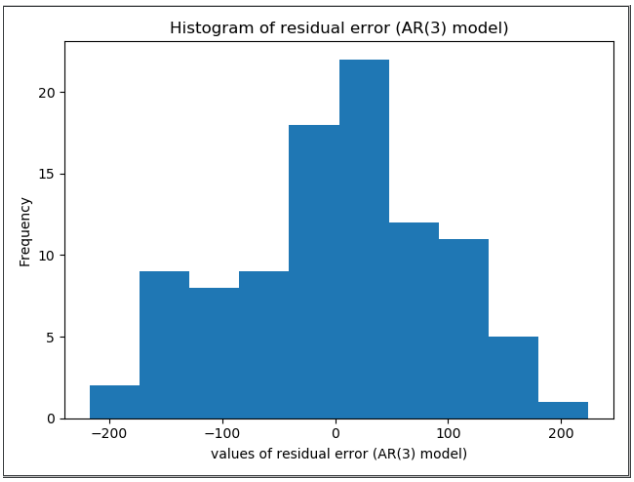
MSE of the residual error for AR(3) model: 8524.498759432317

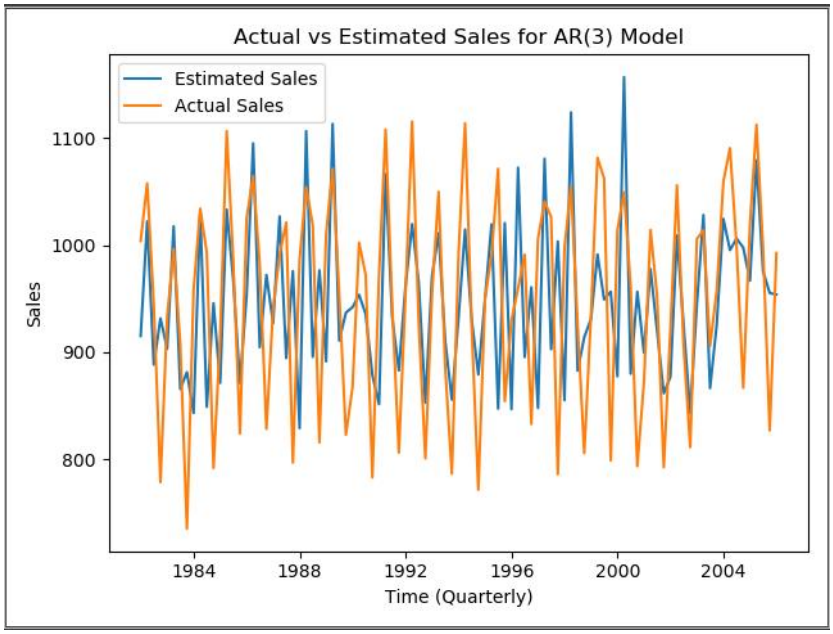
Q value of residual errors for AR(3) model: 257.3538122295325

Mean of residual errors for AR(3) model: 2.285944381872274

Variance of residual errors for AR(3) model: 8519.273217715303







AR(4)

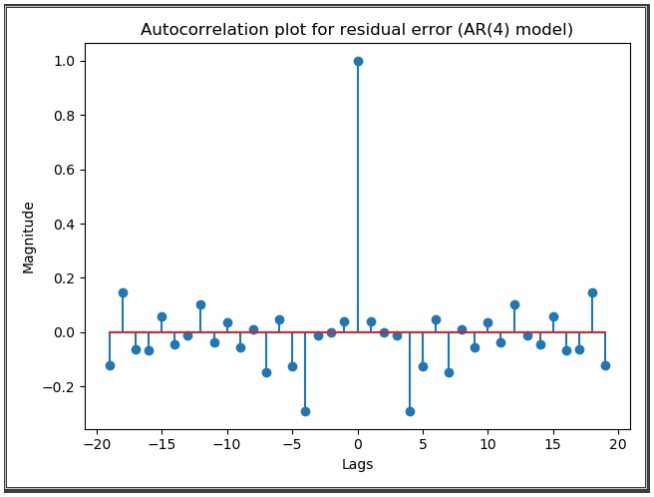
Coefficients for AR(4) are [-0.18272727 0.10965715 -0.14155409 -0.78693594]

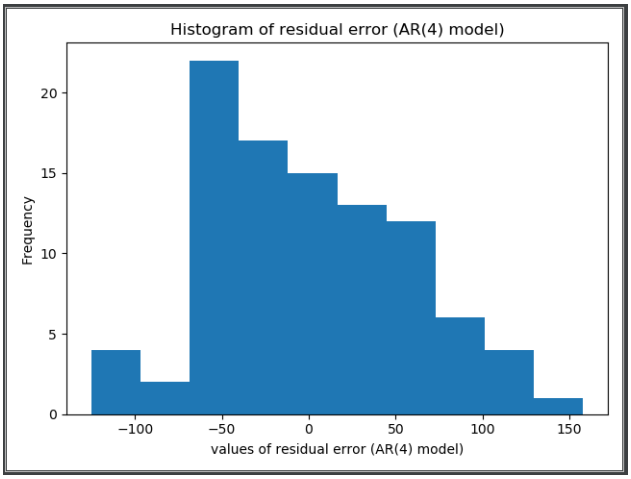
MSE of the residual error for AR(4) model: 3267.2350153036955

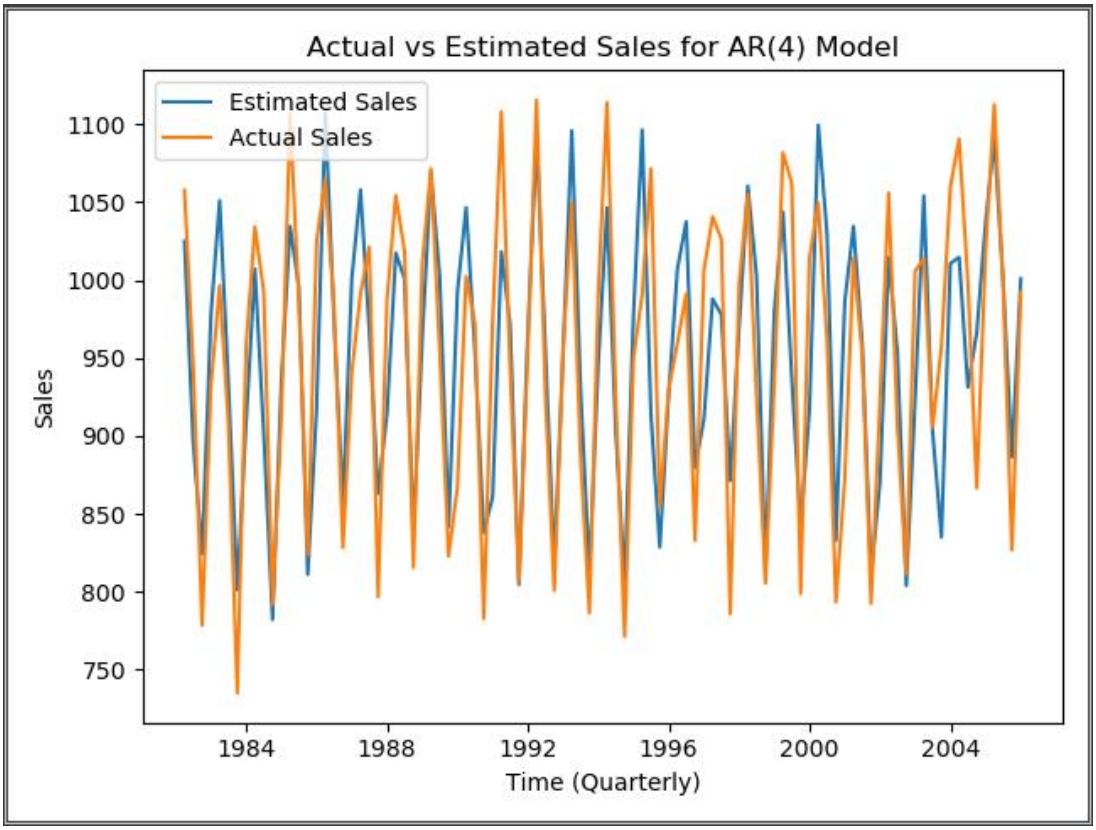
Q value of residual errors for AR(4) model: 18.726306655079398

Mean of residual errors for AR(4) model: 0.4748552045113321

Variance of residual errors for AR(4) model: 3267.009527838444







AR(5)

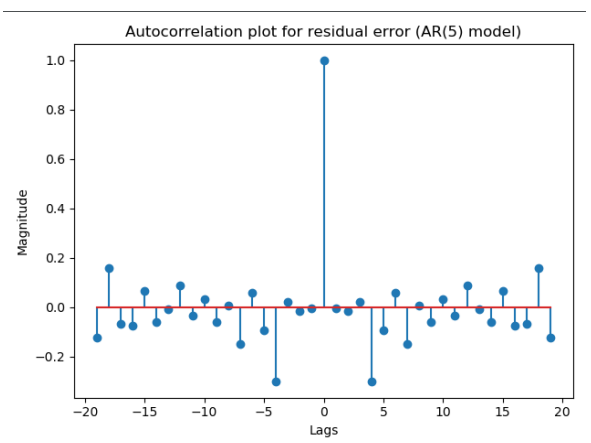
Coefficients for AR(5) are [-0.22398531 0.09670691 -0.13425086 -0.7969546   0.05746722]

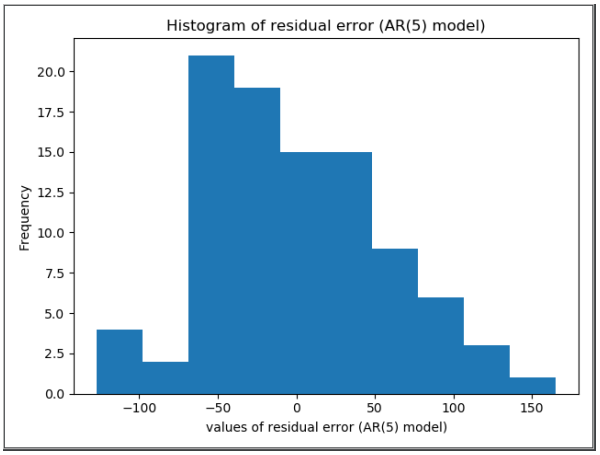
MSE of the residual error for AR(5) model: 3279.09644713048

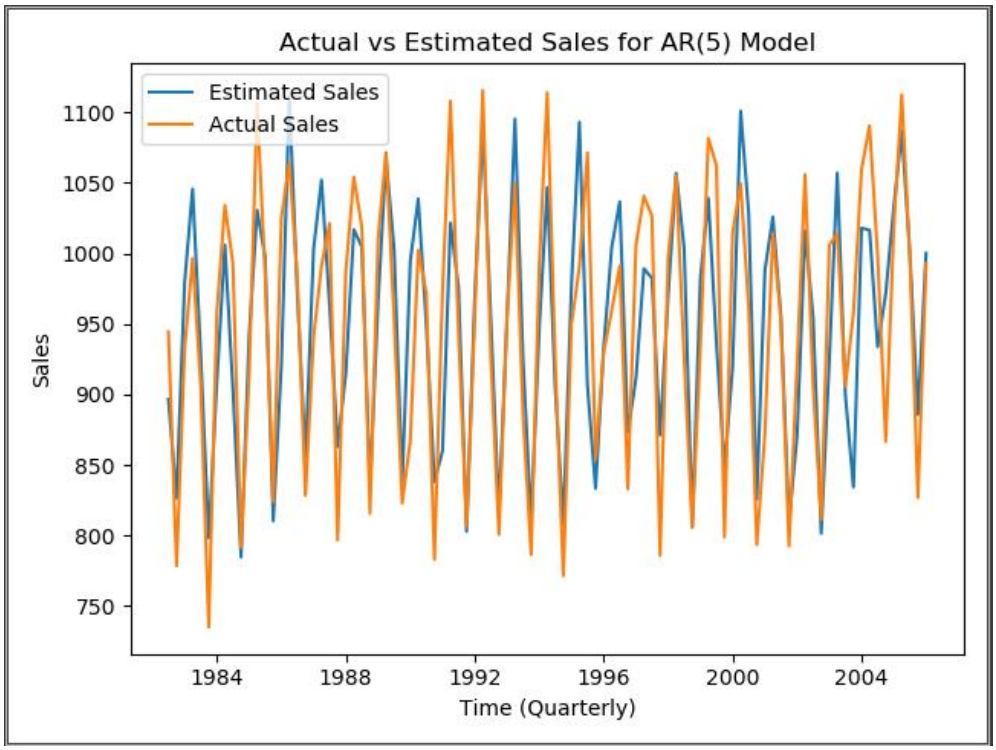
Q value of residual errors for AR(5) model: 18.73168031345683

Mean of residual errors for AR(5) model: 0.5018120401201426

Variance of residual errors for AR(5) model: 3278.8446318068727







9- Create a table and put all this information for different order number inside a table.

	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)
MSE	19081.163431	18760.390797	8524.498759	3267.235015	3279.096447
Q-value	410.802988	435.645262	257.353812	18.726307	18.731680
Mean	10.024060	8.652057	2.285944	0.474855	0.501812
Variance	18980.681644	18685.532699	8519.273218	3267.009528	3278.844632

10- Pick the best AR model order which represents the “Sales” dataset the best.You need to justify why the picked order number makes sense.

AR(4) is the best model since residual error are closer to zero meaning the mean is least (0.474855) among all other models.

11- Write the final AR model with the best order number and the corresponding coefficients as the model that best represents this dataset.

Best order number of AR model is: 4 and their coefficients are: [-0.18272727 0.10965715 -0.14155409 -0.78693594]

APPENDIX

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from Autocorrelation import cal_auto_corr

df = pd.read_csv('tute1.csv', header=0)
df['Date'] = pd.date_range(start='1981-3-1', end='2006-3-1', freq='Q')
y = df['Sales']
y.index = df.Date
N = 100


def calc_coeff(y, order):
    T_new = len(y) - order - 1
    Y = np.array(y[order:len(y)])
    X = np.zeros((T_new + 1, order))
    y_l = list(y)
    k = 1
    for j in range(order):
        for i in range(T_new + 1):
            X[i][j] = y_l[order + i - k]
        k += 1
    X = -1 * pd.DataFrame(X)
    X = np.array(X)
    x_transpose = X.T
    coeff = np.matmul(np.matmul(np.linalg.inv(np.matmul(x_transpose, X)),
x_transpose), Y)
    print('Coefficients for AR({}) are {}'.format(order, coeff))
    return coeff, X, Y


def MSE(error):
    return np.mean(error ** 2)


def cal_Q_value(residual_error, residual_error_acf, lags):
    k = len(residual_error)
    return k * np.sum(np.array(residual_error_acf[lags:]) ** 2)


def plot_acf(residual_error_acf, lags, title):
    plt.figure()
    plt.stem(range(-(lags - 1), lags), residual_error_acf,
use_line_collection=True)
    plt.xlabel('Lags')
    plt.ylabel('Magnitude')
    plt.title('Autocorrelation plot for {}'.format(title))
    plt.show()


def histogram_plot(residual_error, title):
    plt.figure()
    plt.hist(residual_error)
    plt.title('Histogram of {}'.format(title))
```

```
    plt.xlabel('values of {}'.format(title))
    plt.ylabel('Frequency')
    plt.show()


def plot_truevsestimated(df, order, Y, y_hat, title):
    plt.figure()
    plt.plot(df['Date'][order:], y_hat, label='Estimated Sales')
    plt.plot(df['Date'][order:], Y, label='Actual Sales')
    plt.title('Actual vs Estimated Sales for {}'.format(title))
    plt.xlabel('Time (Quarterly)')
    plt.ylabel('Sales')
    plt.legend(loc='upper left')
    plt.show()


def AR_model(df, y, order):
    result = []
    lags=20
    coeff, X, Y = calc_coeff(y, order)
    y_hat = np.matmul(X, coeff)
    residual_error = np.subtract(Y, y_hat)
    result.append(MSE(residual_error))
    print("MSE of the residual error for AR({}) model: {}".format(order,
MSE(residual_error)))
    residual_error_acf = cal_auto_corr(residual_error, lags)
    Q_value = cal_Q_value(residual_error, residual_error_acf, lags)
    result.append(Q_value)
    print('Q value of residual errors for AR({}) model: {}'.format(order,
Q_value))
    title = 'residual error (AR({}) model)'.format(order)
    plot_acf(residual_error_acf, lags, title=title)
    histogram_plot(residual_error, title)
    result.append(np.mean(residual_error))
    result.append(np.var(residual_error))
    print('Mean of residual errors for AR({}) model: {}'.format(order,
np.mean(residual_error)))
    print('Variance of residual errors for AR({}) model: {}'.format(order,
np.var(residual_error)))
    plot_truevsestimated(df, order, Y, y_hat, title='AR({}) Model'.format(order))
    return result, coeff

results = []
coefficients = {}
col_names = []
for i in range(1,6):
    result, coeff = AR_model(df, y, order=i)
    results.append(result)
    coefficients[i] = coeff
    col_names.append('AR({})'.format(i))

results_df = pd.DataFrame(data=results)
results_df = results_df.transpose()
results_df.columns = col_names
results_df.index = ['MSE', 'Q-value', 'Mean', 'Variance']
print(results_df.head())

# AR(4) is the best model
```

```
print('Best order number of AR model is: {} and their coefficients are: {}'.
format(4, coefficients[4]))

#Autocorrelation Function

import numpy as np

def auto_corr(y, k):
    T = len(y)
    y_mean = np.mean(y)
    res_num = 0
    res_den = 0
    for t in range(k, T):
        res_num += (y[t] - y_mean) * (y[t-k] - y_mean)

    for t in range(0, T):
        res_den += (y[t] - y_mean)**2

    result = res_num/res_den
    return result

def cal_auto_corr(y, k):
    res = []
    res1 = []
    for t in range(0, k):
        result = auto_corr(y, t)
        res.append(result)
    for t in range(k-1, 0, -1):
        res1.append(res[t])
    res1.extend(res)
    return res1
```