Time Series and Forecasting

Assignment Report

Supervised by: Dr. Jitendra Kumar

Implementation report of ARIMA model

This report consists of some time series analysis for time vs diet chart which will deal with the following problems:

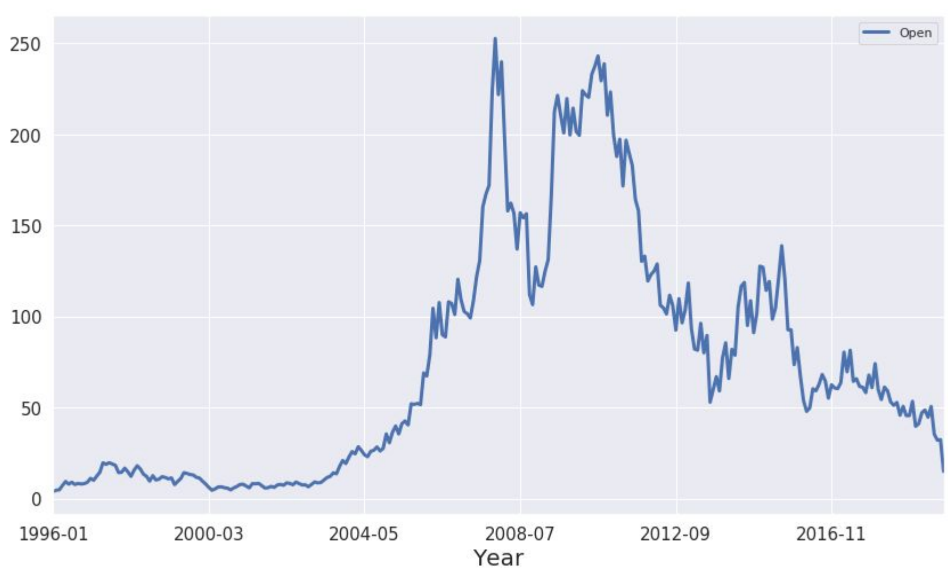
- Checking for the stationarity of data.
- Converting into stationary data.
- Finding Autocorrelation, Partial autocorrelation, and Yule-Walker equations.
- Finding the estimates upto ARMA(5,5) models.
- Forecasting.
- Simulating Data

Dataset Description :

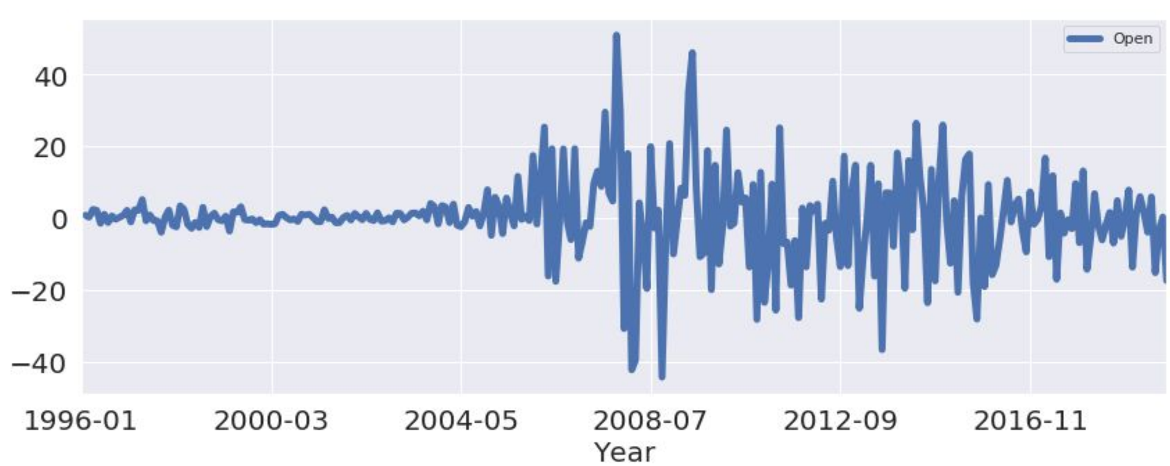
This dataset contains information about opening stock market data of Bharat Heavy Electronics Limited. This dataset has the shape (287,2).

	Open
Date	
1996-01	3.804667
1996-02	4.592889
1996-03	4.839000
1996-04	7.231664
1996-05	9.382107

Dataset plot :



Removing Trend using differencing the series :



Checking for Stationarity:

- ## Using the Augmented Dickey-Fuller test :

The null hypothesis of the test is that the time series can be represented by a unit root, that it is not stationary (has some time-dependent structure). The alternate hypothesis (rejecting the null hypothesis) is that the time series is stationary.

- **Null Hypothesis ($H_{\circ}$)**: If failed to be rejected, it suggests the time series has a unit root, meaning it is non-stationary. It has some time-dependent structure.
- **Alternate Hypothesis (H1)**: The null hypothesis is rejected; it suggests the time series does not have a unit root, meaning it is stationary. It does not have a time-dependent structure.

We have tested our time series in two-phase :

- Without Differencing

  ```
  Dickey-Fuller = -1.4102, Lag order = 6, p-value = 0.8247
  alternative hypothesis: stationary
  ```

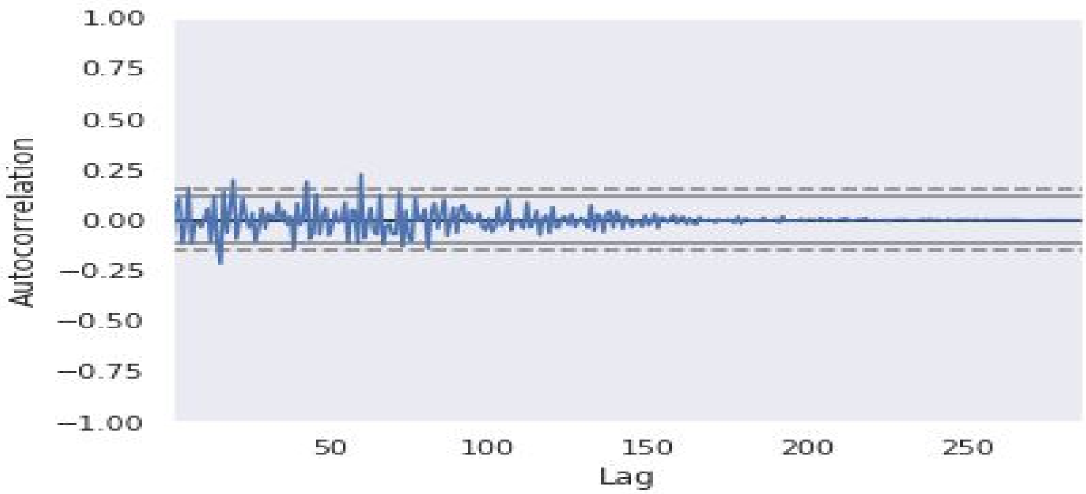
  Here p-value is greater than 0.05 (significance level), So we cannot reject the null hypothesis. It concludes that data is non-stationary.
- With Differencing

  ```
  Dickey-Fuller = -6.0444, Lag order = 6, p-value = 0.01
  alternative hypothesis: stationary
  ```

  Here p-value is smaller than 0.05 (significance level), So we can reject the null hypothesis. It concludes that data is stationary after differencing or it does not have a time-dependent structure.

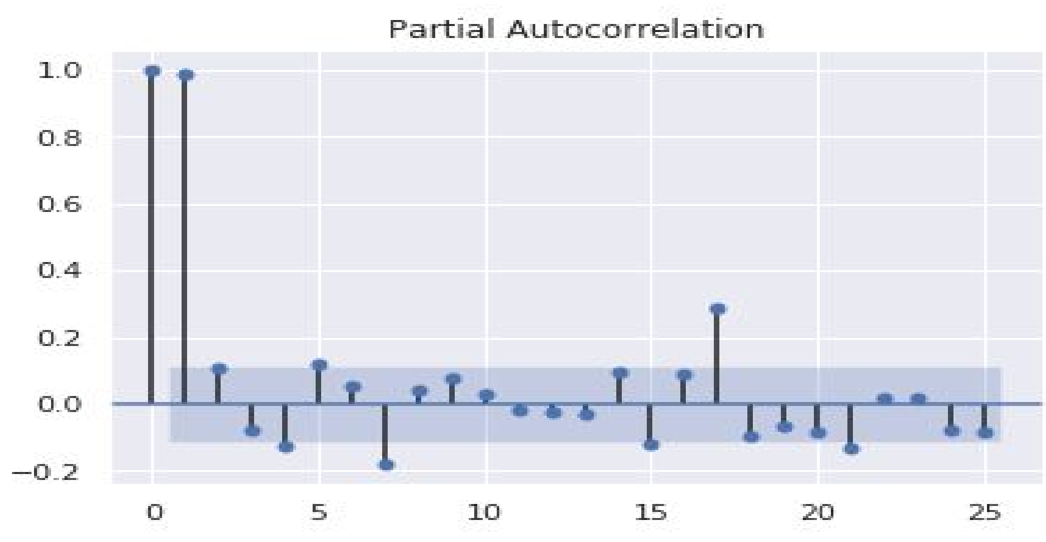
- ## Autocorrelation :

  array([1. , 0.98032364, 0.96446886, 0.94611695, 0.92471574, 0.90794461, 0.8917042 , 0.87047321, 0.85291754, 0.83588488, 0.81919919, 0.8043837 , 0.78888947,0.77177848, 0.75838201,0.74106668, 0.72790943, 0.7221426 , 0.71187908, 0.70398395, 0.69432622, 0.67782947, 0.66449741, 0.65205547, 0.63572796,0.61940331, 0.60376779, 0.58735036, 0.57460403, 0.56315418,0.54965439, 0.53827604, 0.52515539, 0.51224739, 0.49845362, 0.48162077, 0.46502702, 0.44711624, 0.42971697, 0.4121937 , 0.39943554])



- **Partial Autocorrelation :**

array([1. , 0.98375134, 0.10766393, -0.07504894, -0.12539581,0.11823655, 0.05537802, -0.18208189, 0.03880767, 0.07584407,0.03253517, -0.02035495, -0.02689218, -0.03056387, 0.09296362,-0.11771712, 0.09083488, 0.28663107, -0.09729856, -0.06690255, -0.08312814, -0.13163163, 0.01782478, 0.018729 , -0.07576997,-0.08510793, 0.06910983, 0.07032087, 0.01764756, 0.01213496, -0.04715283, 0.0039577 , 0.02996412, 0.0177848 , -0.15391363, -0.10356822, -0.07440875, -0.14288517, 0.09581656, -0.0111916 , 0.16973703])



Fitting the data using the AR(p) model:

- Standardize data
- Fitting data into the AR(p) model using the statsmodels library provides an autoregression model that automatically selects an appropriate lag value using statistical tests and trains a linear regression model.

Result:

Lag: 15

Coefficient :

array([-0.05896711, 0.11461257, -0.02781628, -0.15929061, -0.00413298, 0.02718052, -0.10922659, -0.0573905 , 0.01577114, -0.16018188,0.02129072, -0.02882241, -0.017635 , -0.08372528, -0.00370547, -0.01427569])

Mean squared error: 1.2695

Yule-Walker equation :

```
from statsmodels.regression.linear_model import yule_walker
```

```
param=yule_walker(x,order=5,method='mle')
```

```
param
```

```
(array([-0.08687052,  0.09320513,  0.11151227, -0.10997591, -0.06492007]),
 11.745914417817504)
```

It is used to estimate parameters and standard deviation of residuals of the AR(p) model where p is lagged value given by the user.

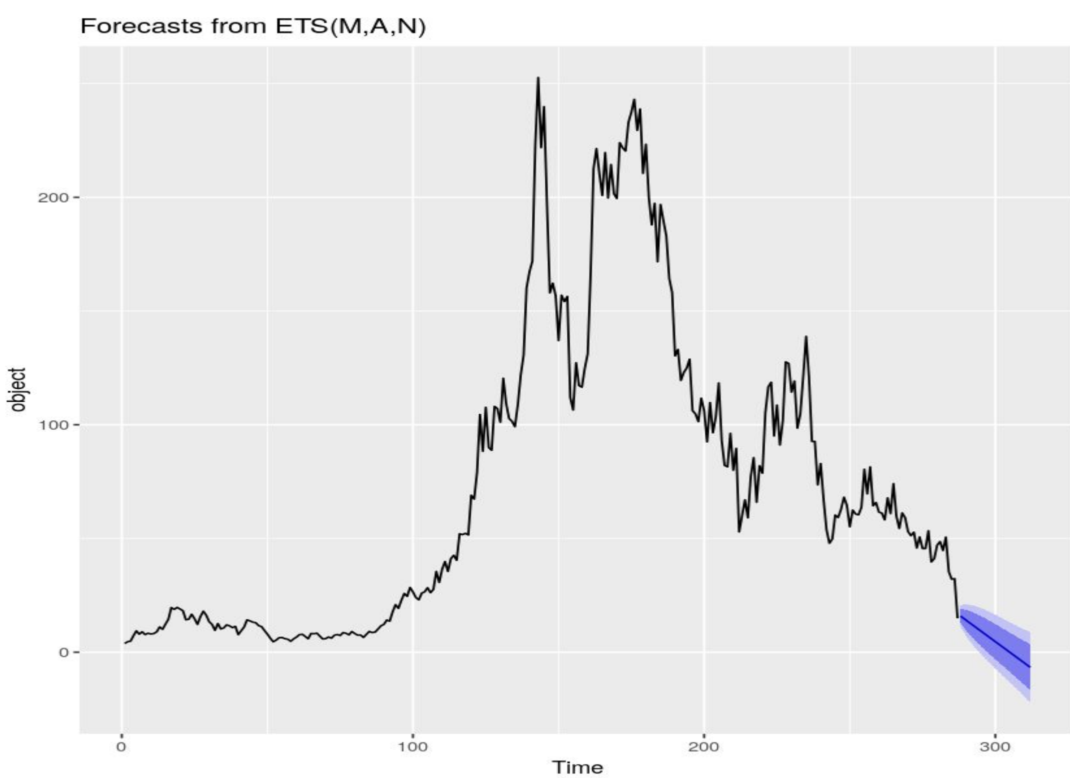
Here I have shown an example of the AR(5) model.

ARIMA(p,0,q) model :

```
ARIMA(1, 0, 1) MSE=150.735
ARIMA(1, 0, 2) MSE=150.620
ARIMA(1, 0, 3) MSE=158.959
ARIMA(1, 0, 4) MSE=154.252
ARIMA(1, 0, 5) MSE=155.440
ARIMA(2, 0, 1) MSE=149.820
ARIMA(2, 0, 2) MSE=153.343
ARIMA(2, 0, 3) MSE=159.080
ARIMA(2, 0, 4) MSE=154.692
ARIMA(2, 0, 5) MSE=155.972
ARIMA(3, 0, 1) MSE=152.003
ARIMA(3, 0, 2) MSE=155.040
ARIMA(3, 0, 3) MSE=156.235
ARIMA(3, 0, 4) MSE=159.563
ARIMA(3, 0, 5) MSE=155.211
ARIMA(4, 0, 1) MSE=152.219
ARIMA(4, 0, 2) MSE=153.455
ARIMA(4, 0, 3) MSE=158.436
ARIMA(4, 0, 4) MSE=152.670
ARIMA(4, 0, 5) MSE=159.091
ARIMA(5, 0, 1) MSE=160.091
ARIMA(5, 0, 2) MSE=161.103
ARIMA(5, 0, 3) MSE=163.864
ARIMA(5, 0, 4) MSE=162.343
ARIMA(5, 0, 5) MSE=160.094
```

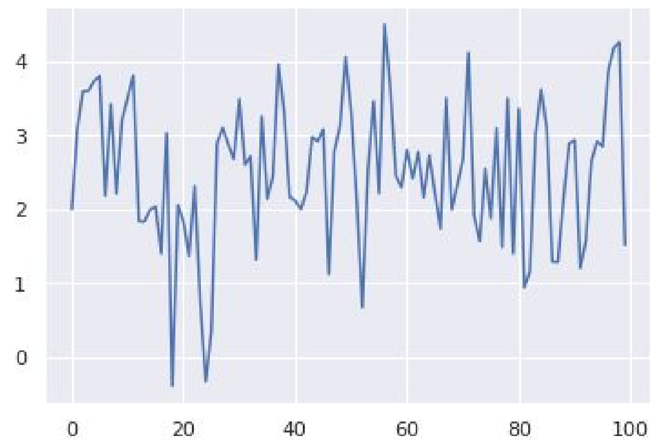
Here, I used the mean square error as metrics and got 149.820 MSE using ARMA(2,1) model.

Forecast Using Auto ARIMA(p,0,q) :



Simulating Data

- **Data generating process for AR(1) :**
  We have generated the data using the AR process with C=3, Φ^1=0.44, and ε Ɛ[0,1]. The dataset can be visualized as :



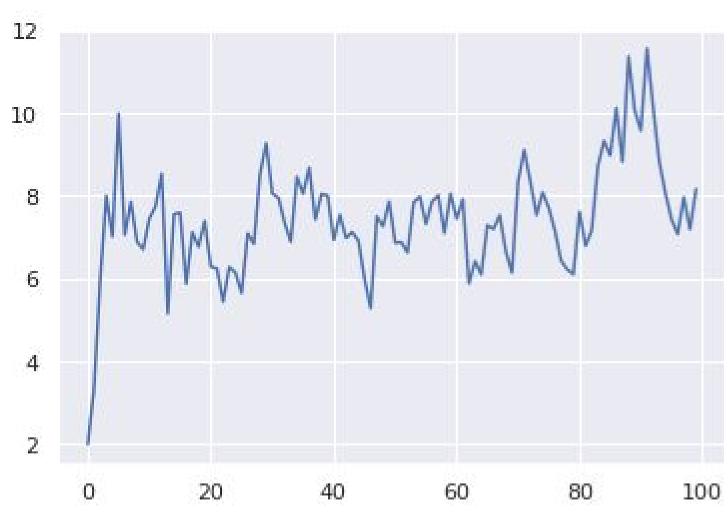
- **Test for Stationarity:**

  Dickey-Fuller = -8.218, p-value =0.00103
  Alternative hypothesis: stationary

  p-value<=0.05. Reject the null hypothesis (H,0), the data does not have a unit root and is stationary.

- **Parameter Estimation:**
  Estimated Parameter are C=2.50058493, Φ^1=0.24080762

- **Data generating process for AR(2) :**
  We have generated the data using the AR process with C=3, Φ^1=0.20, Φ^2=0.42 and ε Ɛ[0,1]. The dataset can be visualized as :



- **Test for Stationarity:**

  Dickey-Fuller = -4.8189, p-value =0.000582
  Alternative hypothesis: stationary

  p-value<=0.05. Reject the null hypothesis (H,0), the data does not have a unit root and is stationary.

- **Parameter Estimation:**
  Estimated Parameter are C = 7.26387486, Φ^1 = 0.55484425, Φ^2 = 0.21907865

Conclusion:

In this report, we have successfully estimated the parameters of the ARIMA(p,0,q) model which fits the stock data of Bharat Heavy Electricals Limited (BHEL). The presented model can be used to forecast the approximate stock market open price for the future. The model has been selected on the basis of the minimization of the AIC score among 36 models where p ε [1,5] and q ε [1,5]. So, according to this report, we can approximate the stock price data using ARIMA(p,0,q).