ECON M 524- FINANCIAL ECONOMETRICS

FINAL PROJECT REPORT

-KAUSHIK PARVATHANENI(kaparvat@iu.edu)

Forecasting SENSEX Stock using ARIMA model

By KAUSHIK PARVATHANENI

Data Science Graduate Student

Indiana University Bloomington

Index:

Table of Figures:

1)Research Question:

Using the ARIMA model to learn forecasting of economic data, with SENSEX data as the example dataset.

2) Introduction:

- The term Sensex(Stock Exchange Sensitive Index) refers to the benchmark index of the BSE in India.
- The Sensex is comprised of 30 of the largest and most actively traded stocks on the BSE and provides a gauge of India's economy.
- It is float-adjusted and market capitalization-weighted.
- The companies that make up the Sensex are drawn from the Bombay Stock Exchange, which is the largest in India and one of the largest stock exchanges in the world.
- Many investors throughout the world use the Sensex as a barometer of the overall state of the Indian economy, which has grown substantially in recent decades.

3)Data Source:

- https://www.bseindia.com/Indices/IndexArchiveData.html

Indices : S&P BSE SENSEX
Period : 04-Jan-2012 to 30-Apr-2022

Date	Open	High	Low	Close
4/01/2012	15,967.49	16,004.69	15,822.32	15,882.64
5/01/2012	15,893.07	15,980.17	15,809.31	15,857.08
6/01/2012	15,789.08	16,001.31	15,664.91	15,867.73
7/01/2012	15,893.30	15,900.30	15,835.29	15,848.80
9/01/2012	15,840.22	15,871.51	15,678.30	15,814.72
10/01/2012	15,898.32	16,180.97	15,898.32	16,165.09
11/01/2012	16,222.37	16,244.70	16,127.77	16,175.86
12/01/2012	16,117.19	16,178.58	15,962.59	16,037.51
13/01/2012	16,144.57	16,257.34	16,049.78	16,154.62
16/01/2012	16,086.74	16,214.36	16,037.60	16,189.36
17/01/2012	16,270.87	16,501.38	16,270.87	16,466.05
18/01/2012	16,502.42	16,517.96	16,384.48	16,451.47
19/01/2012	16,573.87	16,662.06	16,572.10	16,643.74
20/01/2012	16,745.01	16,788.48	16,611.71	16,739.01
23/01/2012	16,667.02	16,784.00	16,659.32	16,751.73
24/01/2012	16,806.72	17,050.32	16,770.01	16,995.77
25/01/2012	17,068.85	17,130.24	17,016.69	17,077.18
27/01/2012	17,201.33	17,258.97	17,106.57	17,233.98
30/01/2012	17,138.04	17,138.04	16,828.33	16,863.30
31/01/2012	16,965.58	17,238.99	16,965.58	17,193.55

4)Data Description:

It contains 5 columns and 2557 rows in total.
*Date-from 4/1/12 to 29/4/22
*Open-How much is the SENSEX value opened when the market opened.
*High-The highest SENSEX value reached during the market hours on that day.
* Low-The least SENSEX value reached on that day.
*Close-The SENSEX value when the market closed.

5)Research Method:

*ARIMA - Autoregressive Integrated Moving Average

*ARIMA models are generally denoted as ARIMA(p,d,q)

*The acronym ARIMA, which stands for "Auto Regressive Integrated Moving Average," refers to a class of models that uses a time series' own previous values specifically, its own lags and lagged prediction errors to "explain" the time series in order to predict future values.

*ARIMA models are generally denoted as ARIMA (p,d,q) where p is the order of autoregressive model, d is the degree of differencing, and q is the order of moving-average model. ARIMA models use differencing to convert a non-stationary time series into a stationary one, and then predict future values from historical data. These models use “auto” correlations and moving averages over residual errors in the data to forecast future values.

*An ARIMA model is basically an ARIMA model fitted on d-th order differenced time series such that the final differenced time series is stationary.

*Here, p = the number of lagged observations in the model, ε is white noise at time t, c is a constant and φs are parameters.

6)Steps:

1)Extract the dataset and plot the open data.

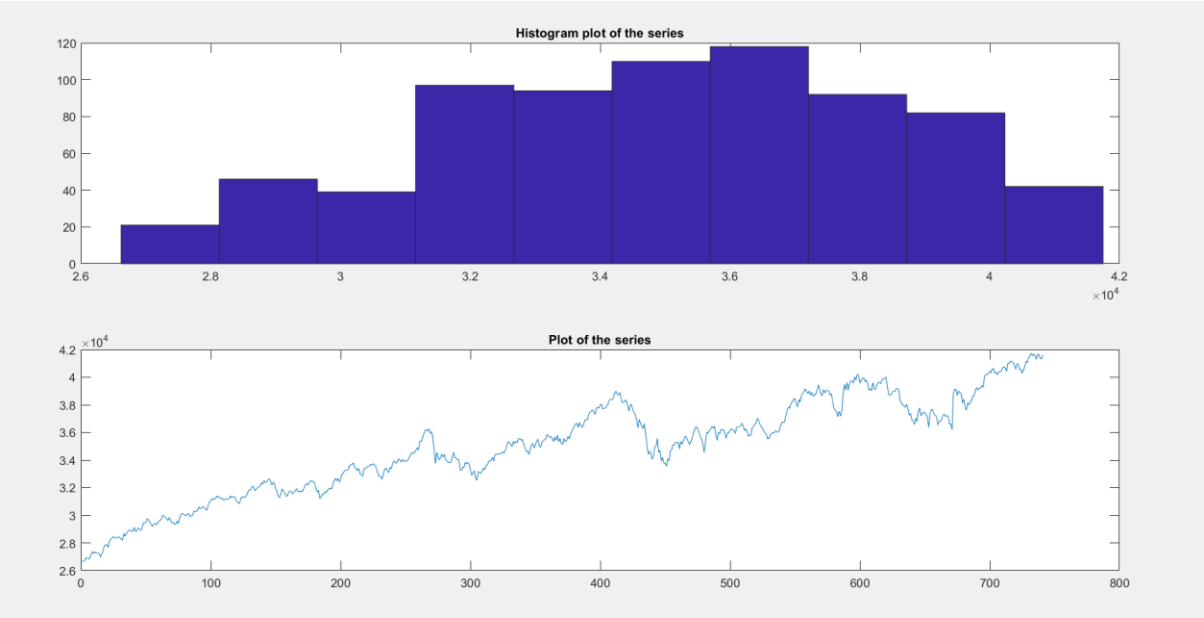


Figure 1

2) Evaluated and plotted the autocorrelation and partial correlation of the open data.

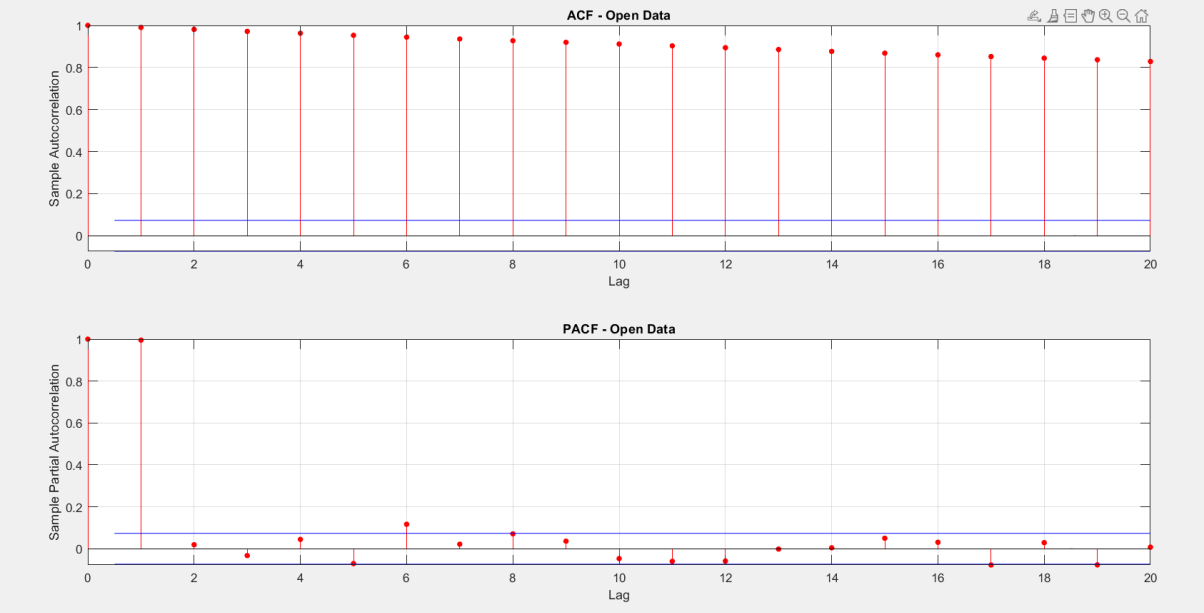


Figure 2

3) Plotting the differences:

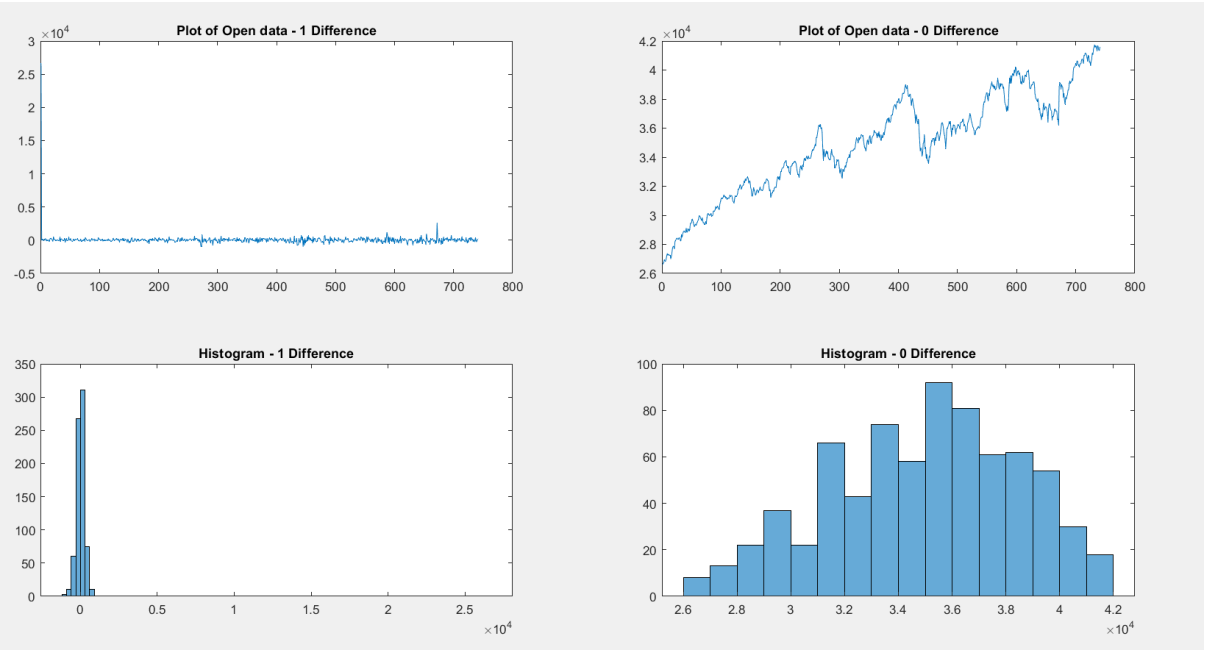


Figure 3

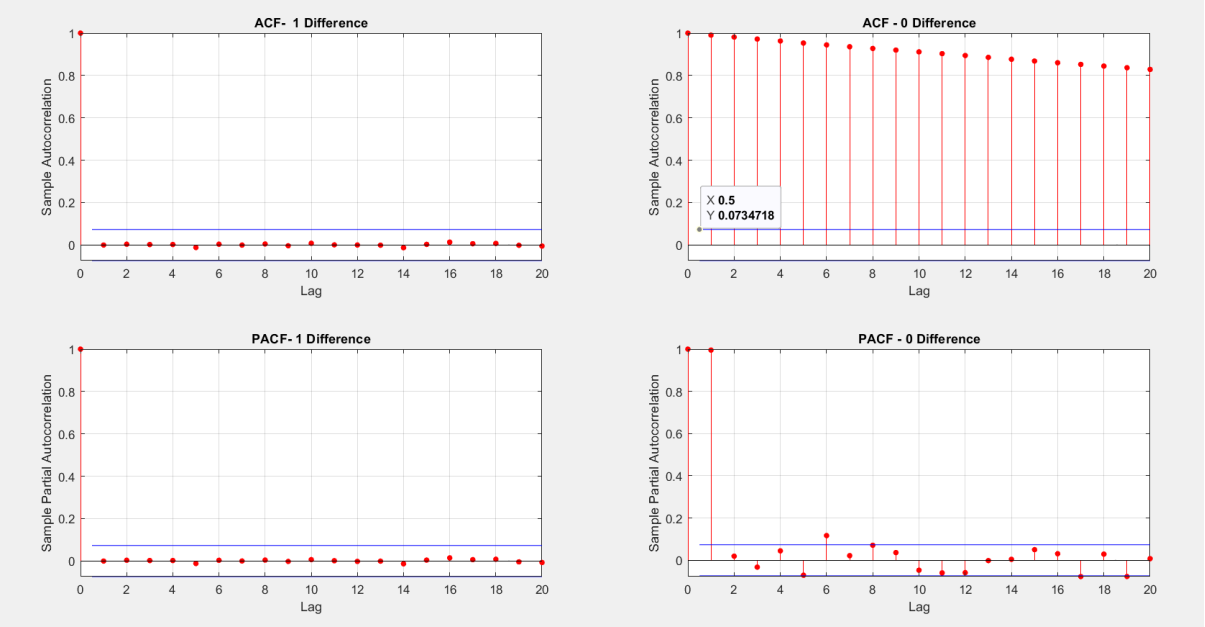


Figure 4

4)Fitting the ARIMA model.

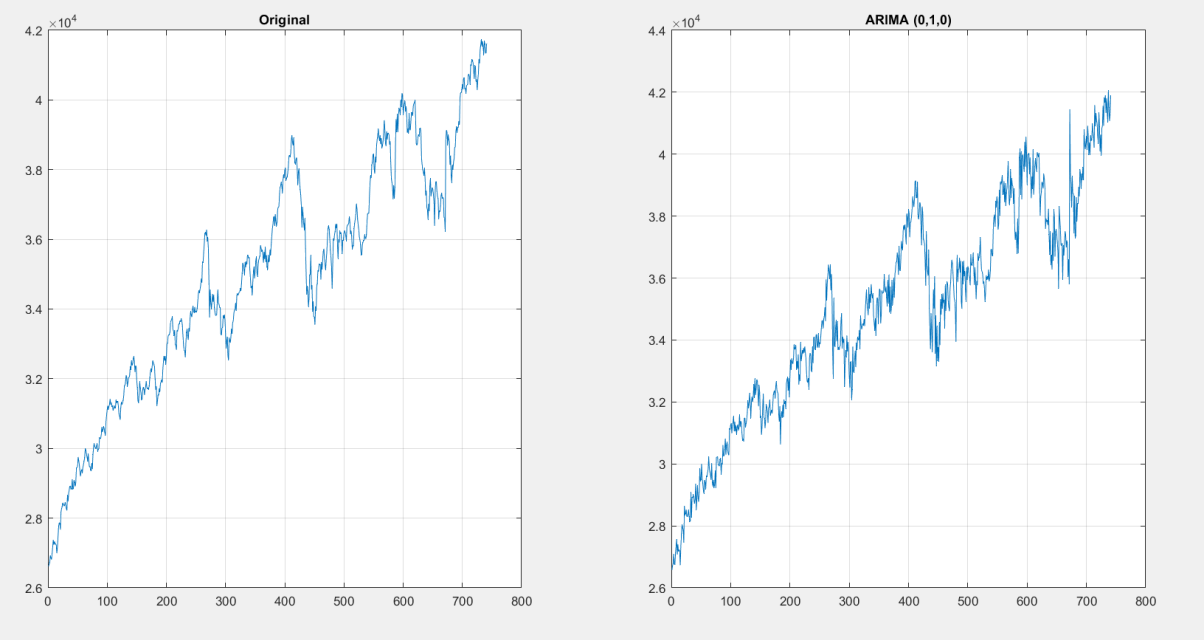


Figure 5

5)Histogram of residuals.

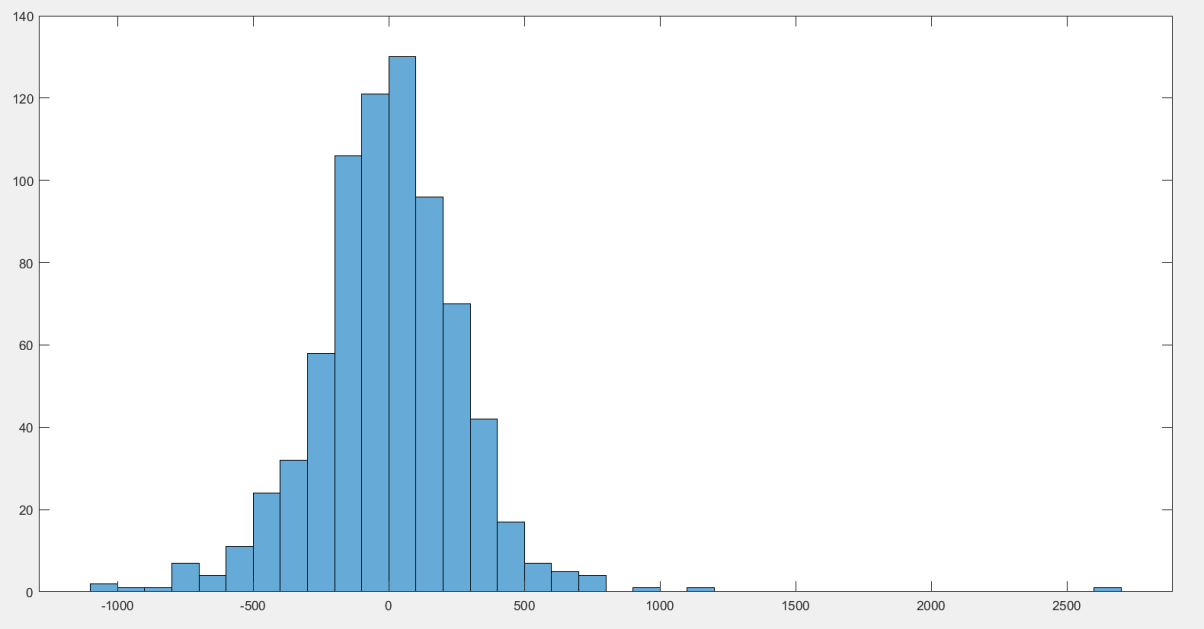


Figure 6

6)Best fit ARIMA model.

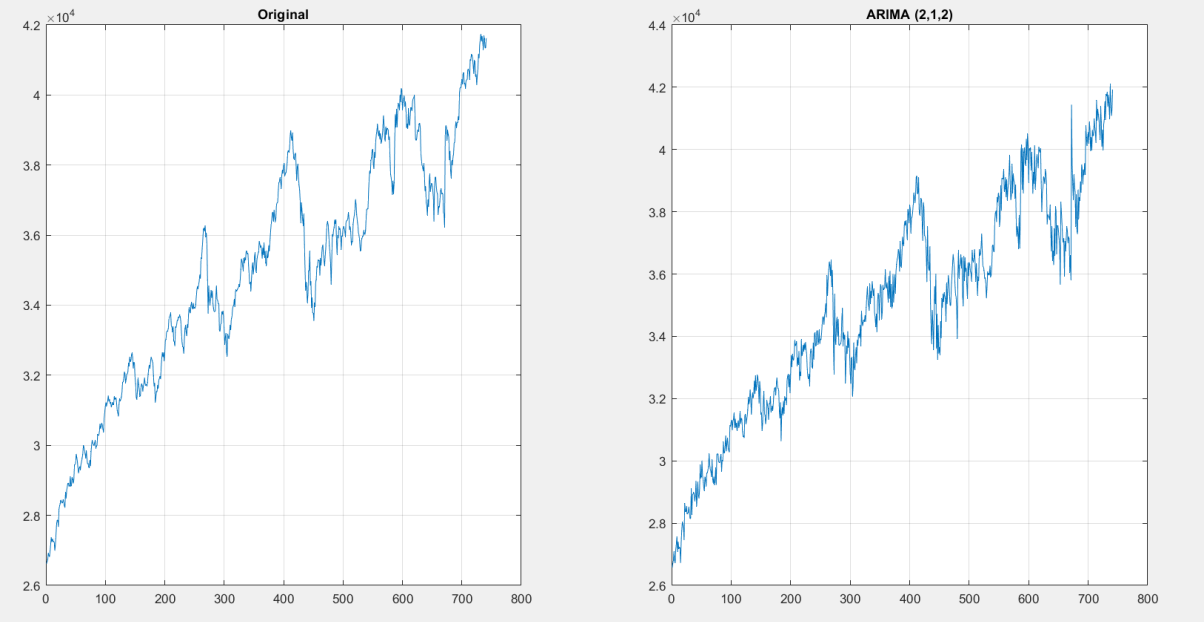


Figure 7

7)Output:

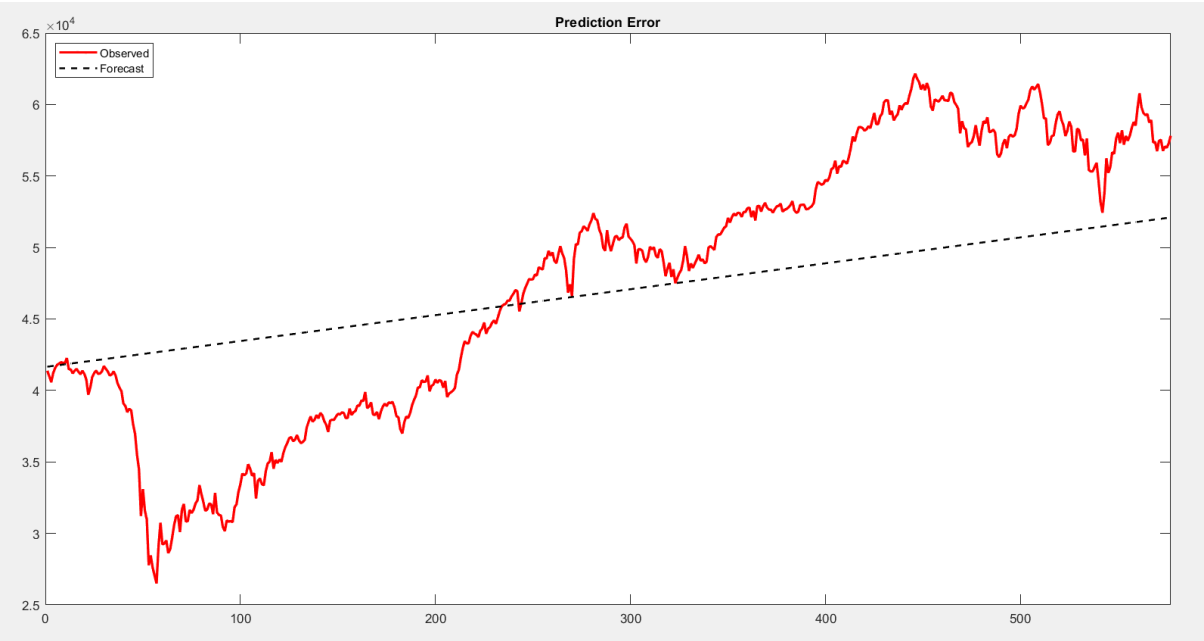


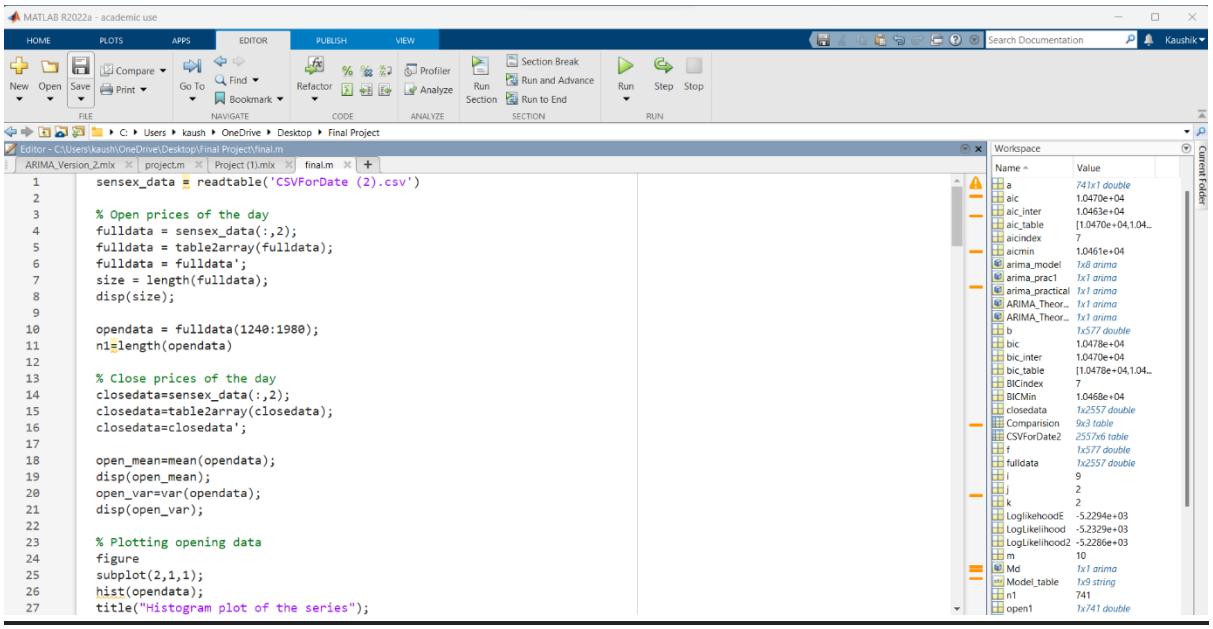
Figure 8

8)Conclusion:

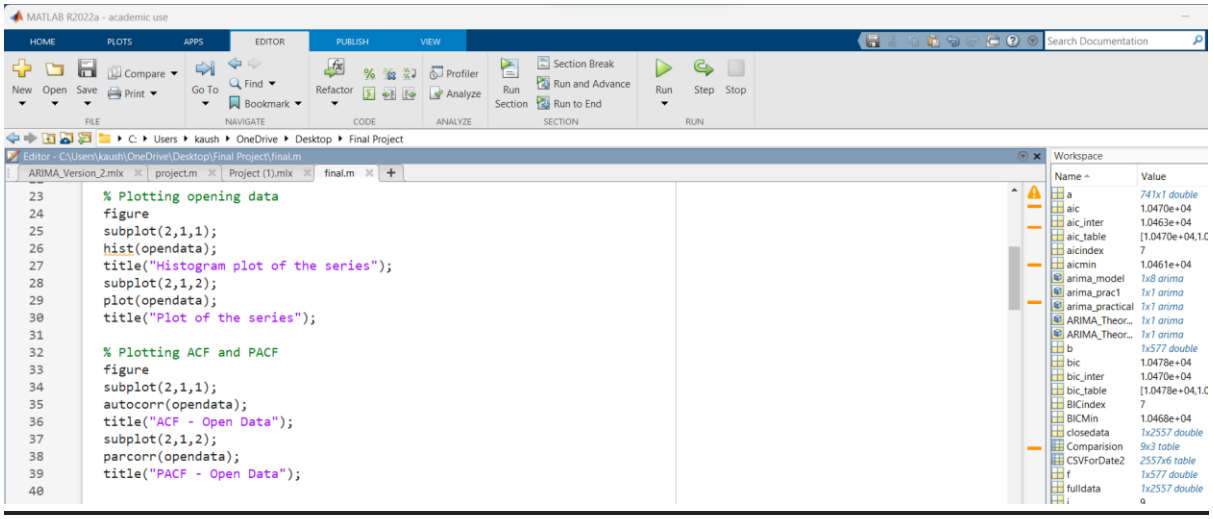
- So we got ARIMA(2,1,2) as the best fit.
- To determine the best fit for the model we performed AIC and BIC.
- When you are dealing with any financial markets whether it be Stocks or Cryptocurrency you always need to be careful as the forecasts are not accurate and anything can happen in the Future.

9)Appendix:

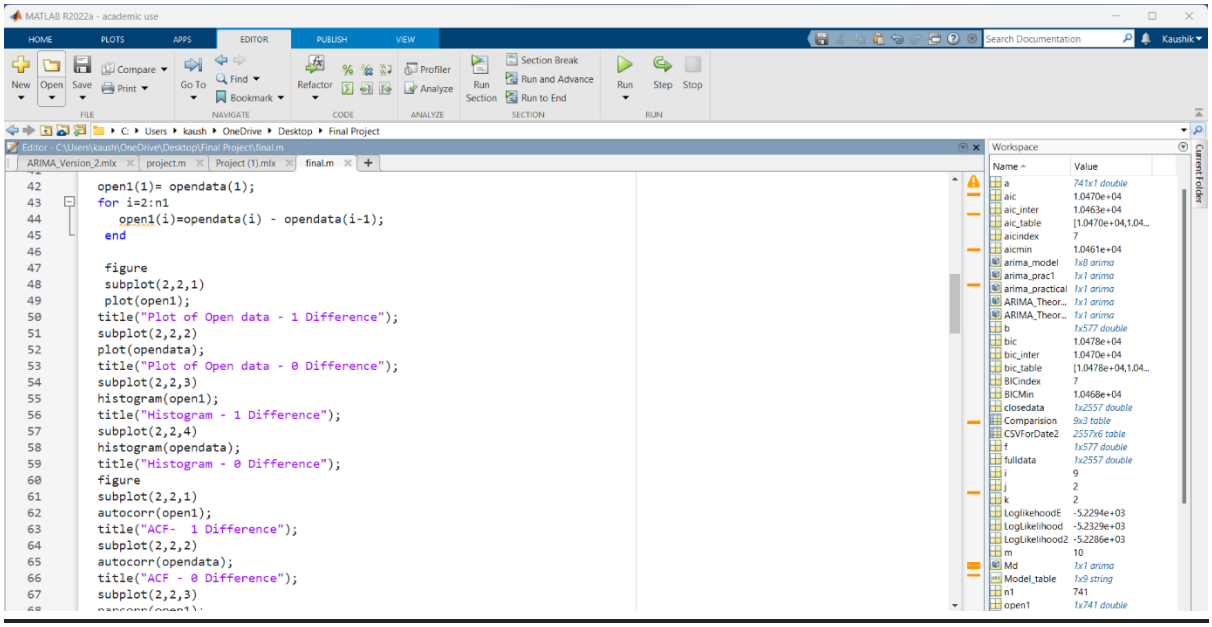
MATLAB Code:



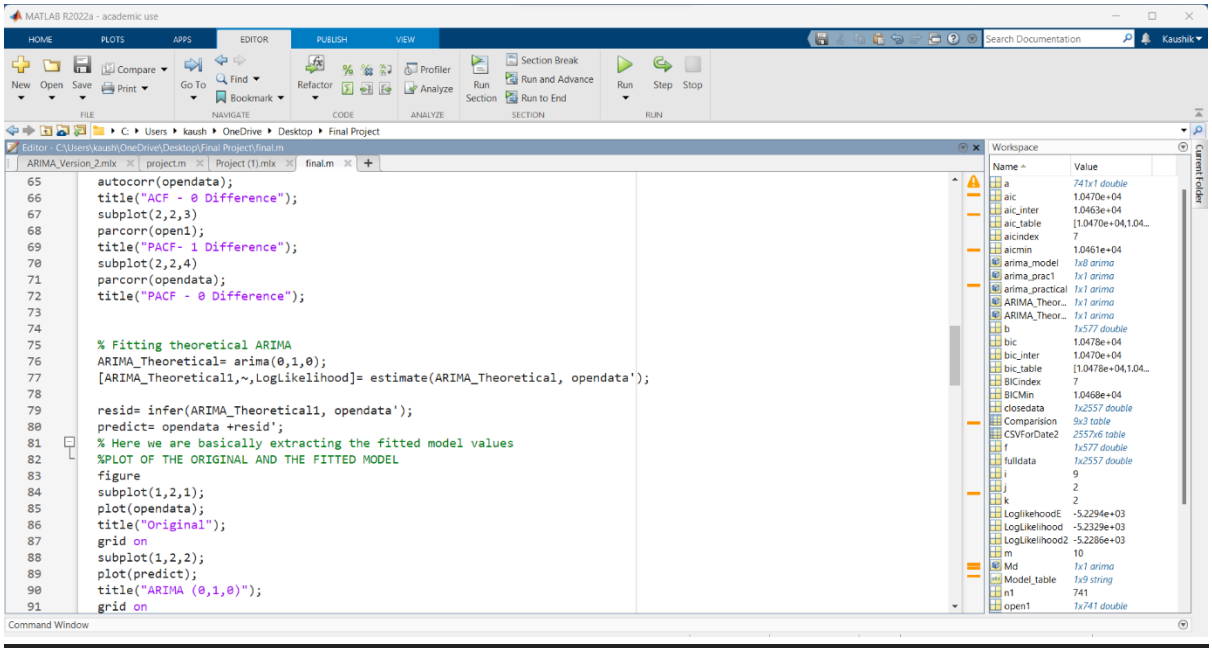
```
sensex_data = readtable('CSVForDate (2).csv')

% Open prices of the day
fulldata = sensex_data(:,2);
fulldata = table2array(fulldata);
fulldata = fulldata';
size = length(fulldata);
disp(size);

opendata = fulldata(1240:1980);
n1=length(opendata)

% Close prices of the day
closedata=sensex_data(:,2);
closedata=table2array(closedata);
closedata=closedata';

open_mean=mean(opendata);
disp(open_mean);
open_var=var(opendata);
disp(open_var);

% Plotting opening data
figure
subplot(2,1,1);
hist(opendata);
title("Histogram plot of the series");
```



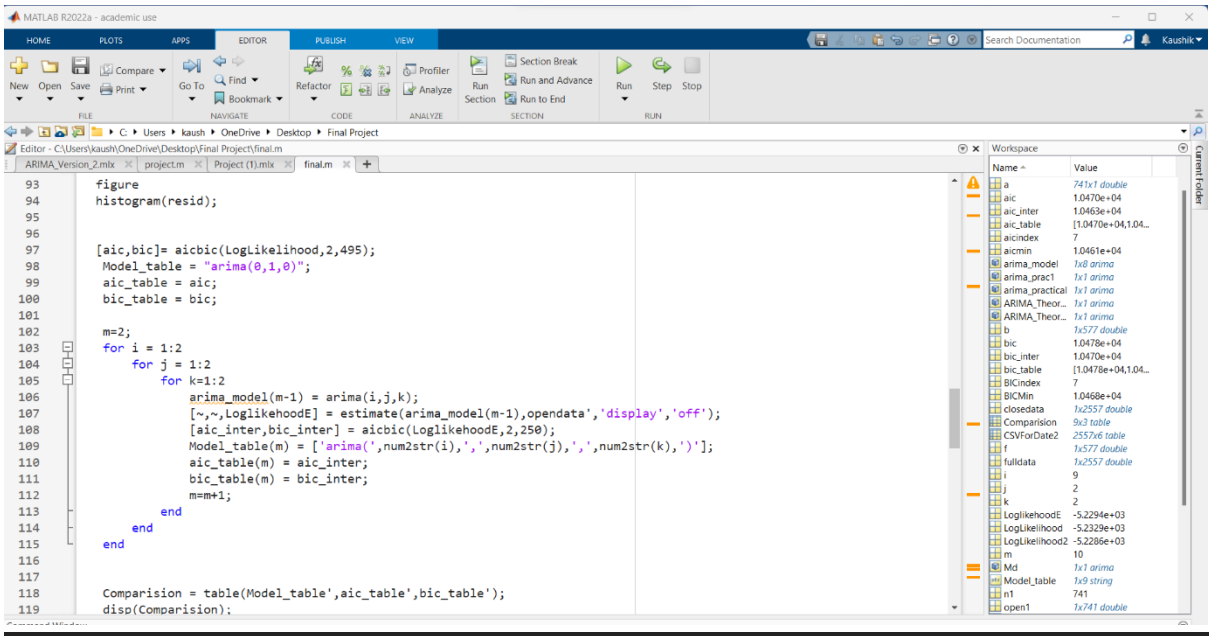
```
% Plotting opening data
figure
subplot(2,1,1);
hist(opendata);
title("Histogram plot of the series");
subplot(2,1,2);
plot(opendata);
title("Plot of the series");

% Plotting ACF and PACF
figure
subplot(2,1,1);
autocorr(opendata);
title("ACF - Open Data");
subplot(2,1,2);
parcorr(opendata);
title("PACF - Open Data");
```



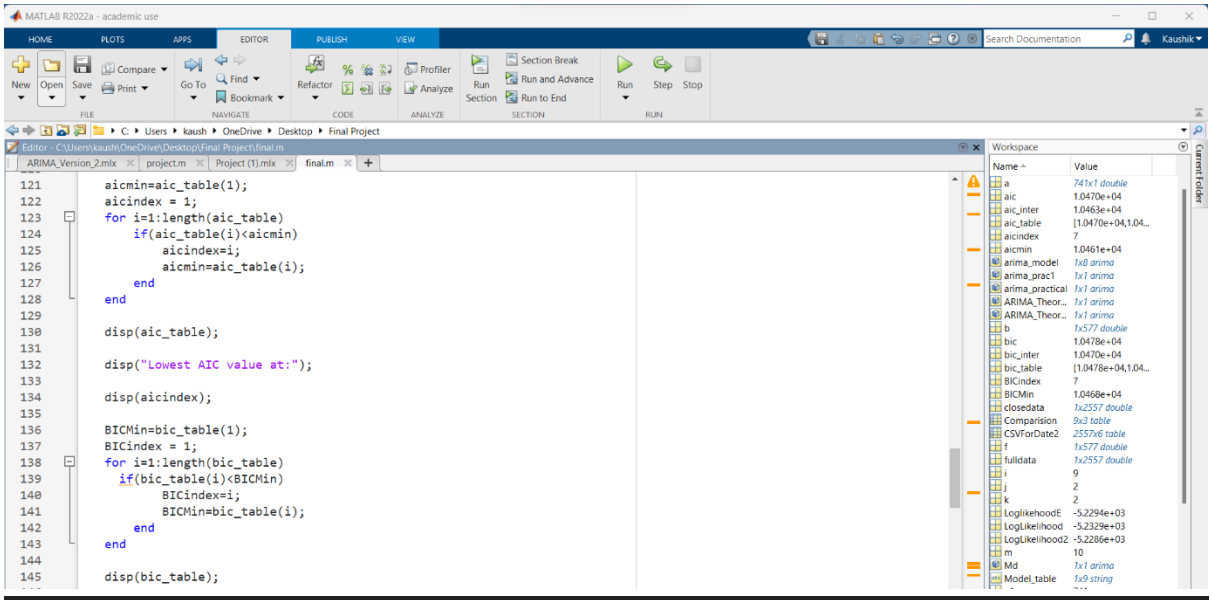
```
open1(1)= opendata(1);
for i=2:n1
   open1(i)=opendata(i) - opendata(i-1);
 end

 figure
 subplot(2,2,1)
 plot(open1);
title("Plot of Open data - 1 Difference");
subplot(2,2,2)
plot(opendata);
title("Plot of Open data - 0 Difference");
subplot(2,2,3)
histogram(open1);
title("Histogram - 1 Difference");
subplot(2,2,4)
histogram(opendata);
title("Histogram - 0 Difference");
figure
subplot(2,2,1)
autocorr(open1);
title("ACF-  1 Difference");
subplot(2,2,2)
autocorr(opendata);
title("ACF - 0 Difference");
subplot(2,2,3)
```



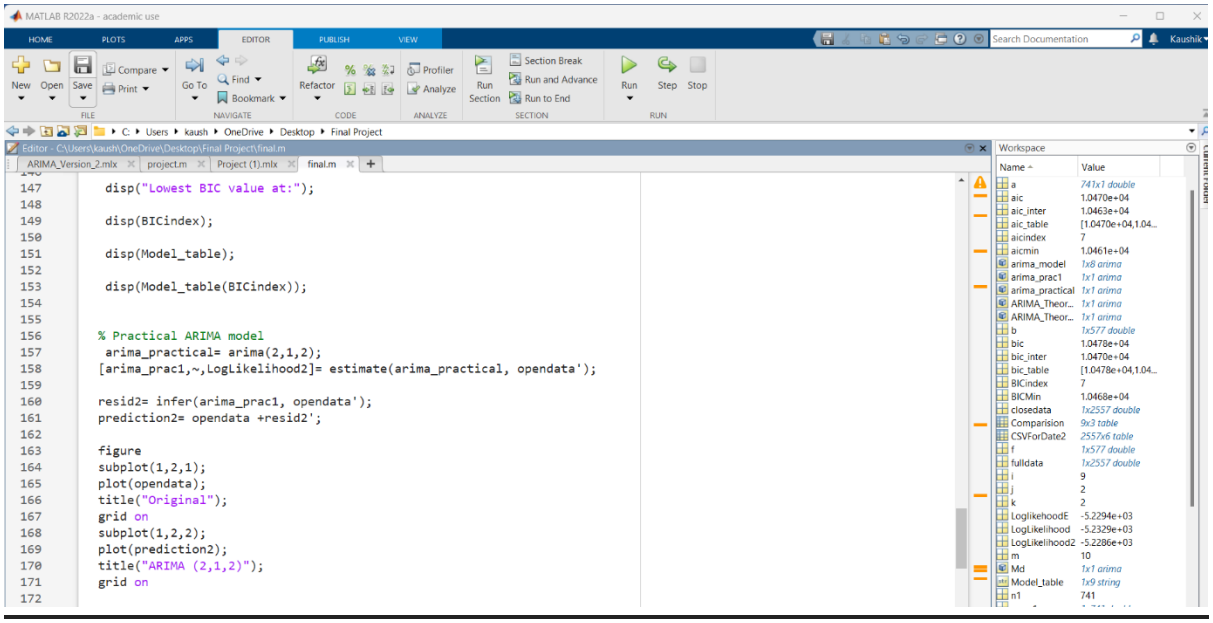
```
autocorr(opendata);
title("ACF - 0 Difference");
subplot(2,2,3)
parcorr(open1);
title("PACF- 1 Difference");
subplot(2,2,4)
parcorr(opendata);
title("PACF - 0 Difference");


% Fitting theoretical ARIMA
ARIMA_Theoretical= arima(0,1,0);
[ARIMA_Theoretical1,~,LogLikelihood]= estimate(ARIMA_Theoretical, opendata');

resid= infer(ARIMA_Theoretical1, opendata');
predict= opendata +resid';
% Here we are basically extracting the fitted model values
%PLOT OF THE ORIGINAL AND THE FITTED MODEL
figure
subplot(1,2,1);
plot(opendata);
title("Original");
grid on
subplot(1,2,2);
plot(predict);
title("ARIMA (0,1,0)");
grid on
```



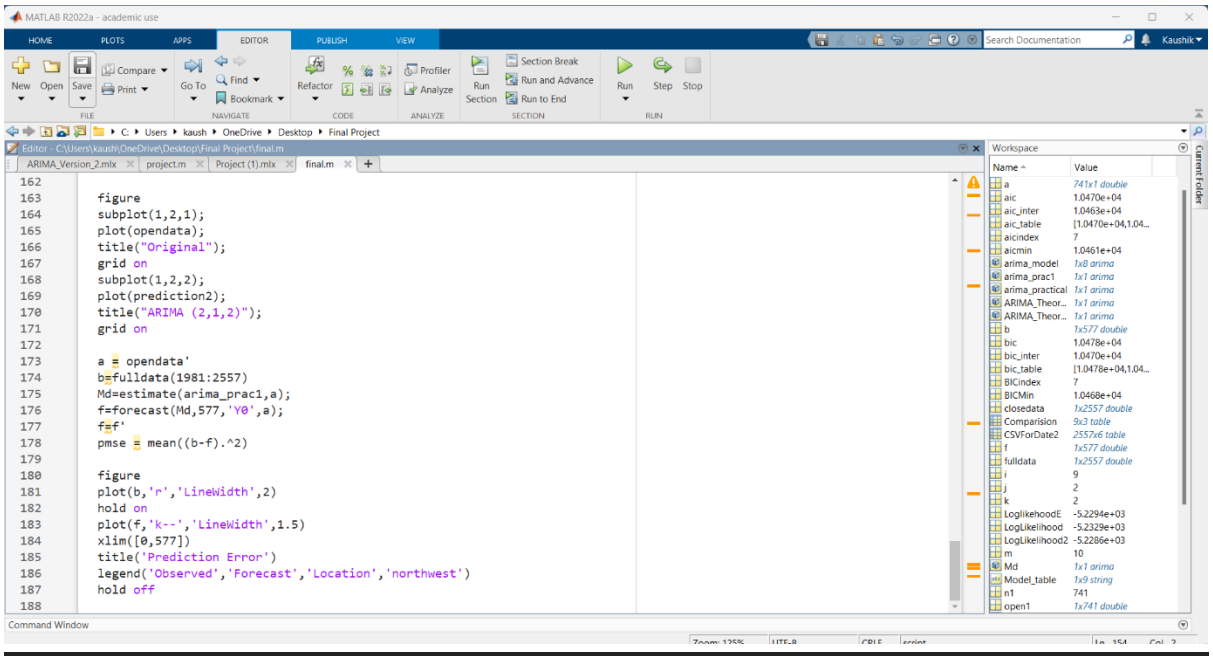
```
figure
histogram(resid);


[aic,bic]= aicbic(LogLikelihood,2,495);
 Model_table = "arima(0,1,0)";
 aic_table = aic;
 bic_table = bic;

 m=2;
 for i = 1:2
     for j = 1:2
         for k=1:2
             arima_model(m-1) = arima(i,j,k);
             [~,~,LoglikehoodE] = estimate(arima_model(m-1),opendata','display','off');
             [aic_inter,bic_inter] = aicbic(LoglikehoodE,2,250);
             Model_table(m) = ['arima(',num2str(i),',',num2str(j),',',num2str(k),')'];
             aic_table(m) = aic_inter;
             bic_table(m) = bic_inter;
             m=m+1;
         end
     end
 end


 Comparision = table(Model_table',aic_table',bic_table');
 disp(Comparision);
```



```
aicmin=aic_table(1);
aicindex = 1;
for i=1:length(aic_table)
    if(aic_table(i)<aicmin)
        aicindex=i;
        aicmin=aic_table(i);
    end
end

disp(aic_table);

disp("Lowest AIC value at:");

disp(aicindex);

BICMin=bic_table(1);
BICindex = 1;
for i=1:length(bic_table)
  if(bic_table(i)<BICMin)
        BICindex=i;
        BICMin=bic_table(i);
    end
end

disp(bic_table);
```



```
147     disp("Lowest BIC value at:");
148
149     disp(BICindex);
150
151     disp(Model_table);
152
153     disp(Model_table(BICindex));
154
155
156    % Practical ARIMA model
157     arima_practical= arima(2,1,2);
158    [arima_prac1,~,LogLikelihood2]= estimate(arima_practical, opendata');
159
160    resid2= infer(arima_prac1, opendata');
161    prediction2= opendata +resid2';
162
163    figure
164    subplot(1,2,1);
165    plot(opendata);
166    title("Original");
167    grid on
168    subplot(1,2,2);
169    plot(prediction2);
170    title("ARIMA (2,1,2)");
171    grid on
172
```



```
162
163    figure
164    subplot(1,2,1);
165    plot(opendata);
166    title("Original");
167    grid on
168    subplot(1,2,2);
169    plot(prediction2);
170    title("ARIMA (2,1,2)");
171    grid on
172
173    a = opendata'
174    b=fulldata(1981:2557)
175    Md=estimate(arima_prac1,a);
176    f=forecast(Md,577,'Y0',a);
177    f=f'
178    pmse = mean((b-f).^2)
179
180    figure
181    plot(b,'r','LineWidth',2)
182    hold on
183    plot(f,'k--','LineWidth',1.5)
184    xlim([0,577])
185    title('Prediction Error')
186    legend('Observed','Forecast','Location','northwest')
187    hold off
188
```

Outputs from Command Window:

```
n1 =

   741

   3.4961e+04

   1.2276e+07


    ARIMA(0,1,0) Model (Gaussian Distribution):

                Value      StandardError    TStatistic      PValue
                _____      _____________    __________      _______

    Constant    20.185         10.764          1.8753       0.060753
    Variance     79905         1733.4          46.098              0

        Var1          Var2    Var3
    ______________    ____    ____

    "arima(0,1,0)"    10470   10478
    "arima(1,1,1)"    10462   10469
    "arima(1,1,2)"    10461   10468
    "arima(1,2,1)"    10472   10479
    "arima(1,2,2)"    10464   10471
    "arima(2,1,1)"    10461   10468
    "arima(2,1,2)"    10461   10468
    "arima(2,2,1)"    10470   10477
    "arima(2,2,2)"    10463   10470
```

```
Command Window

   1.0e+04 *

    1.0470    1.0462    1.0461    1.0472    1.0464    1.0461    1.0461    1.0470    1.0463

Lowest AIC value at:
     7

   1.0e+04 *

    1.0478    1.0469    1.0468    1.0479    1.0471    1.0468    1.0468    1.0477    1.0470

Lowest BIC value at:
     7

    "arima(0,1,0)"    "arima(1,1,1)"    "arima(1,1,2)"    "arima(1,2,1)"    "arima(1,2,2)"    "arima(2,1,1)"    "arima(2,1,2)"    "arima(2,2,1)"    "arima(2,2,2)"

arima(2,1,2)
Warning: Upper bound constraints are active; standard errors may be inaccurate.
> In arima/estimate (line 1070)
In final (line 158)

    ARIMA(2,1,2) Model (Gaussian Distribution):

                Value      StandardError    TStatistic      PValue
                _____      _____________    __________      _______

    Constant    24.353         48.447          0.50267       0.6152
```

ARIMA(2,1,2) Model (Gaussian Distribution):

	Value	StandardError	TStatistic	PValue
Constant	24.353	48.447	0.50267	0.6152
AR{1}	-0.62794	1.3715	-0.45785	0.64706
AR{2}	0.28414	1.3016	0.2183	0.8272
MA{1}	0.61205	1.3812	0.44312	0.65768
MA{2}	-0.24881	1.2661	-0.19651	0.84421
Variance	78772	1846.6	42.659	0

ARIMA(2,1,2) Model (Gaussian Distribution):

	Value	StandardError	TStatistic	PValue
Constant	24.353	0	Inf	0
Variance	78772	0	Inf	0

```
pmse =

   4.6697e+07
```