Assignment 2

Title: Univariate Time-Series Analysis

Name: - Dristanta Nirola

Roll No: - DST-19/20-010

Univariate analysis :

Univariate analysis is a quantitative analysis of only one variable. When we model univariate time series, we are modelling time-series changes that represent changes in a single variable over time. Time series forecasting models use the past movements of variables in order to predict their future behaviour.

Foreign Trade of India: Prior to the 1991 economic liberalisation, India was a closed economy due to which there were extensive quantitative restrictions on imports. Foreign investment was strictly restricted to only allow Indian ownership of businesses. Since the liberalisation, the Indian economy has improved mainly due to increased foreign trade. Indian foreign trade includes both Imports and exports of goods and services.

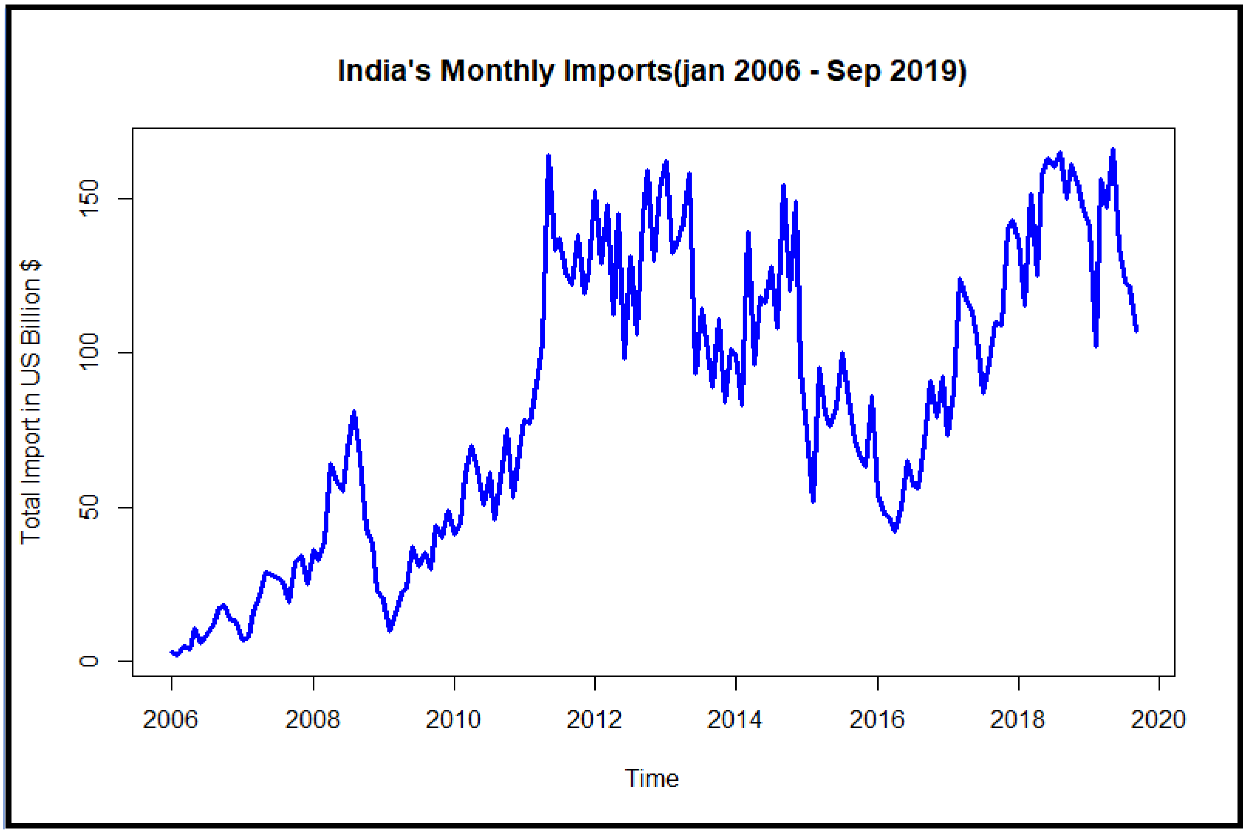
Export involves the selling of good and services from the domestic country to a foreign country. At the same time, **Import** refers to the purchase of foreign products and bringing into one's home country. A country importing more than it's export runs a **trade deficit**. Whereas, a country importing less than it exports creates a **trade surplus**. India suffers a considerable trade deficit when it comes to global trade. Though in the recent report, the monthly trade data for June 2020 shows that India records the first trade surplus in 18 Years; But the bigger questions to be answered is "Will It Last?". According to a report from "FocusEconomy" which forecast "that the exports and imports of India will decrease 3.7% and 8.2% in FY 2020, respectively, bringing the merchandise trade deficit to USD 166.0 billion. In FY 2021, their panel projects exports and imports to increase by 7.4% and 14.6%, respectively, resulting in a merchandise trade deficit of USD 167.0 billion.

Short-term forecasting of key economic variables, such as GDP, Foreign Trade and inflation, has a long tradition, as many prominent business and investment decisions are based on forecasts for the outlook of the economy. In a globalized world characterized over the last several decades by increasing interdependency among nations involving intensive political, social and economic interaction, the need for predictions of critical economic parameters both at home and abroad has increased. With the vital role that trade has played on this broader process of globalization, it comes as no surprise that forecasts of exports and imports of major trading nations and regional blocs have become a central feature of providers of economic forecasts. Here in this analysis, we have used data on Indian imports from January 2006 to September 2019 as our variable for further analysis to forecast the Indian trade from the month Oct 2019 to Oct 2020.
Note: For making the analysis simple, I have ignored the effect caused due to COVID-19 in the report. Thus, our analysis differs much from the original figures.

Data Structure:

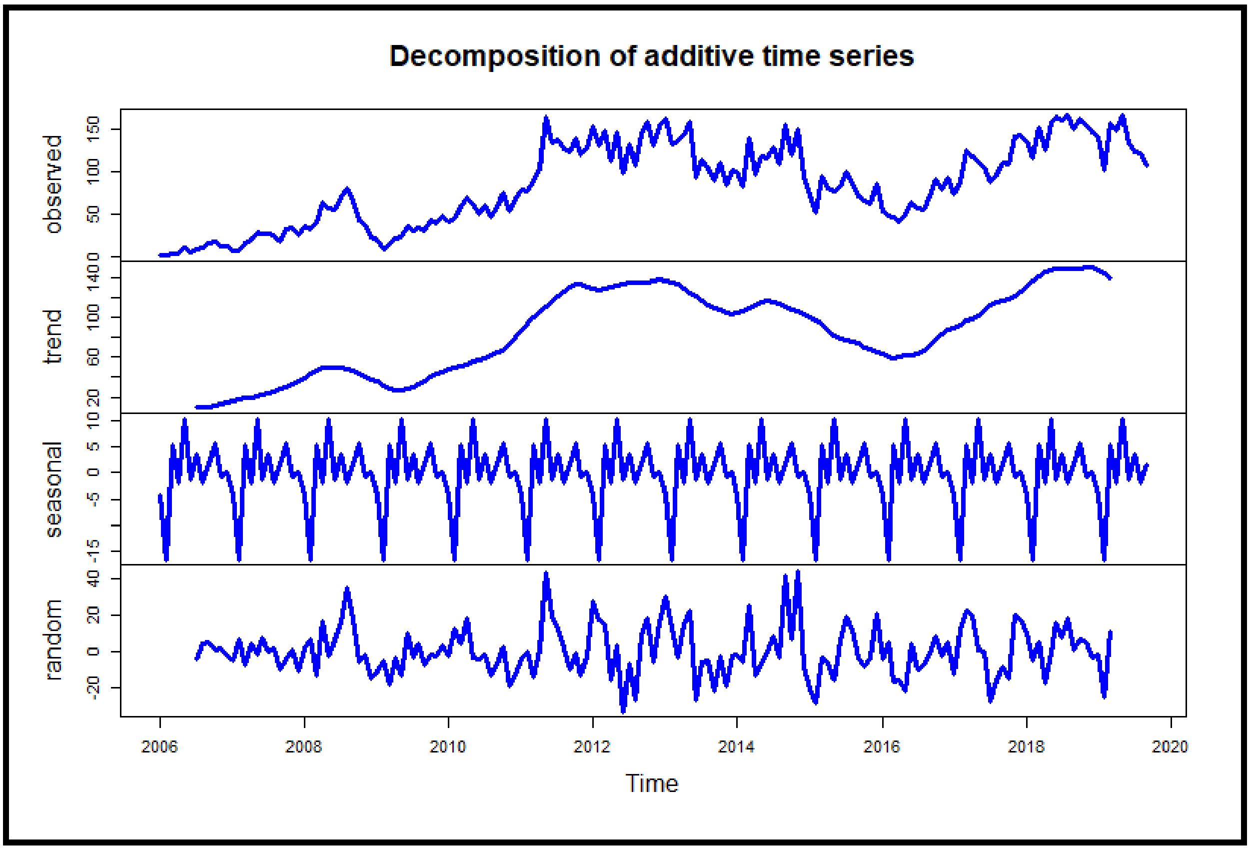
Variable	Indian Imports
Start Date	January 2006
End Date	September 2019
Frequency	Monthly
No. of observations	165
Data Source	https://www.kaggle.com/gauraviiitian/indian-trade-data-monthly

The plot of X(t):



Conclusion: From the plot, it is seen that the value of x(t) increases with the pass of time; thus, we can conclude that x(t) follows a deterministic trend. Also, there is some repeating pattern shown in the data, which indicates that the data contains seasonality. This series is thus further decomposed to plot the different time series components of x(t).

Decomposition Function: Decompose function allows us to extract different components of a time series data.



Conclusion: From the plot, we can find the trend component which concludes that x(t) follows a deterministic trend. Also, there is a seasonal component in the data, which indicates that the data also contains seasonality.

ADF Test: Applying the Augmented Dickey-Fuller Test for checking the stationarity of x(t).

Hypothesis:

H0: Non- stationary

H1: Stationary

Test 1: Applying ADF function on the deseasonalized x(t) series

Summary Test 1:

Value of test-statistic is: -2.2176 1.9312 2.6963

Critical values for test statistics:

	1pct	5pct	10pct
tau3	-3.99	-3.43	-3.13
phi2	6.22	4.75	4.07
phi3	8.43	6.49	5.47

Hypothesis: Since the test statistics is -2.217 and the critical value at 5% is -3.43, so we accept the null hypothesis and conclude that our series is non-stationary ## Cancer is found

Test 2: Applying ADF function on the 1st difference series yt= x(t)-x(t-1)

Summary Test 2:

Value of test-statistic is: -10.0701; 33.8104; 50.7122

Critical values for test statistics:

	1pct	5pct	10pct
tau3	-3.99	-3.43	-3.13
phi2	6.22	4.75	4.07
phi3	8.43	6.49	5.47

Hypothesis: Since, the test statistics is -10.0701, and the critical value at 5% is -3.43, so we reject the null hypothesis and conclude that our series is stationary ##Cancer is removed

Test 3: Checking for deterministic Trend in the yt series. lm(yt~t)

Summary Test 3:

Residuals:

Min	1Q	Median	3Q	Max
-1.04095	-0.20926	0.00952	0.18820	1.16173

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-5.371e-03	2.411e-02	-0.223	0.824
t	1.102e-05	5.607e-05	0.197	0.844

Hypothesis: Since, the p-value 0.844 is greater than 0.05, i.e. insignificant we fail to reject the null hypothesis and conclude that there is no trend present in the series thus the series yt is stationary(free from both cancer and fever).

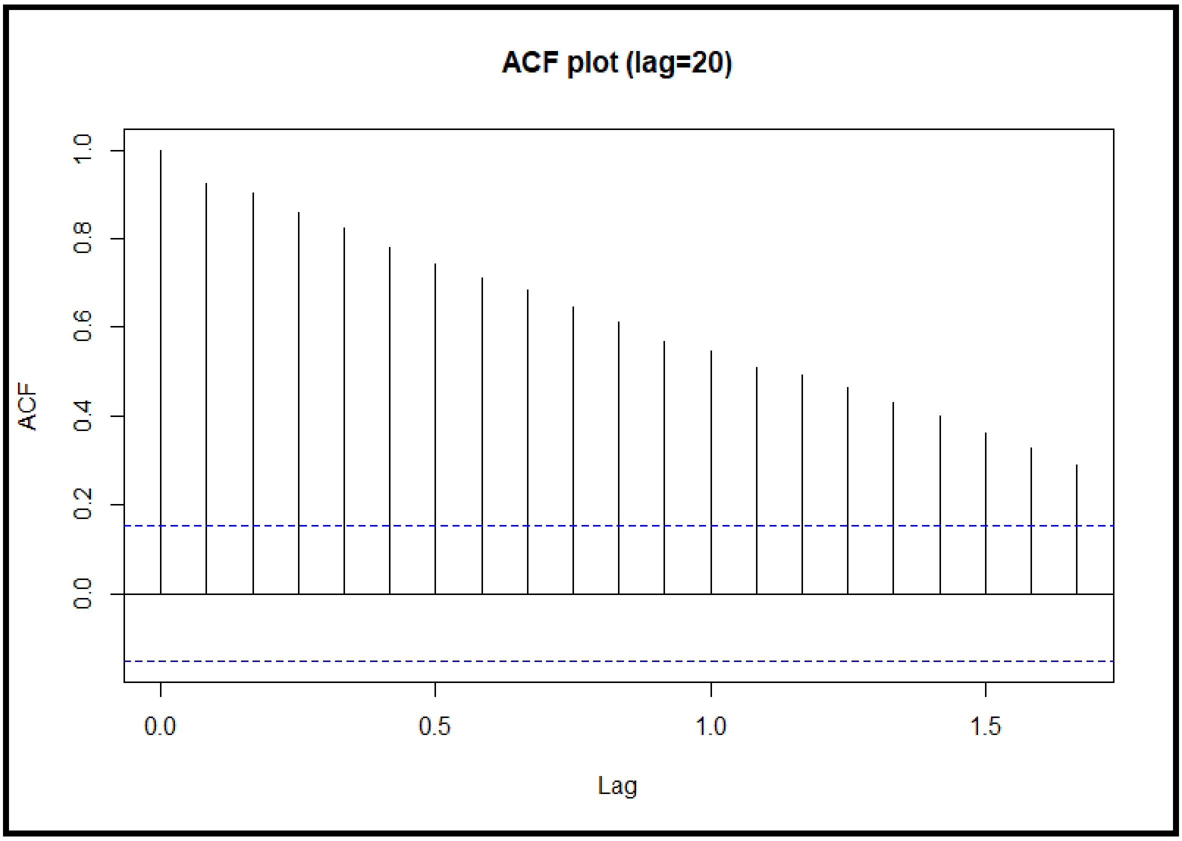
Test for Sample ACF:

Hypothesis:

H0: P1=P2=……..=P20=0

H1: At least their exist one non zero P

Method 1: Using Autocorrelation plot



From the plot, it is seen that the Acf is diminishing. Thus the series may fit in with an AR or an ARMA model.

Method 2: Using L-Jung Box Test

Table1:

Lags(k)	Acf_value	Q(k)	Critical value	Result
1	1	28.65234627	3.841458821	Rejected
2	0.923303979	32.97250095	5.991464547	Rejected
3	0.9040863	33.9552063	7.814727903	Rejected
4	0.859137058	34.45006241	9.487729037	Rejected
5	0.82471152	34.75869058	11.07049769	Rejected
6	0.780391362	34.89532532	12.59158724	Rejected
7	0.743378925	35.13922681	14.06714045	Rejected
8	0.709424482	36.07344048	15.50731306	Rejected
9	0.684079527	36.20334979	16.9189776	Rejected
10	0.647328064	37.29246892	18.30703805	Rejected
11	0.613371092	42.05772549	19.67513757	Rejected
12	0.568424522	43.8167383	21.02606982	Rejected
13	0.54512494	46.94085112	22.36203249	Rejected
14	0.508187534	47.18187629	23.6847913	Rejected
15	0.49060911	48.84804544	24.99579014	Rejected
16	0.465539916	49.12086815	26.2962276	Rejected
17	0.429343156	49.81385833	27.58711164	Rejected
18	0.398273989	50.03859122	28.86929943	Rejected
19	0.362082197	50.33959272	30.14352721	Rejected
20	0.328037049	50.37886791	31.41043284	Rejected

From Table 1, it is clear that the q statistics are grater then the chisqr critical values at df=lag. Thus it is clear that the series best fits the AR or an ARMA model.

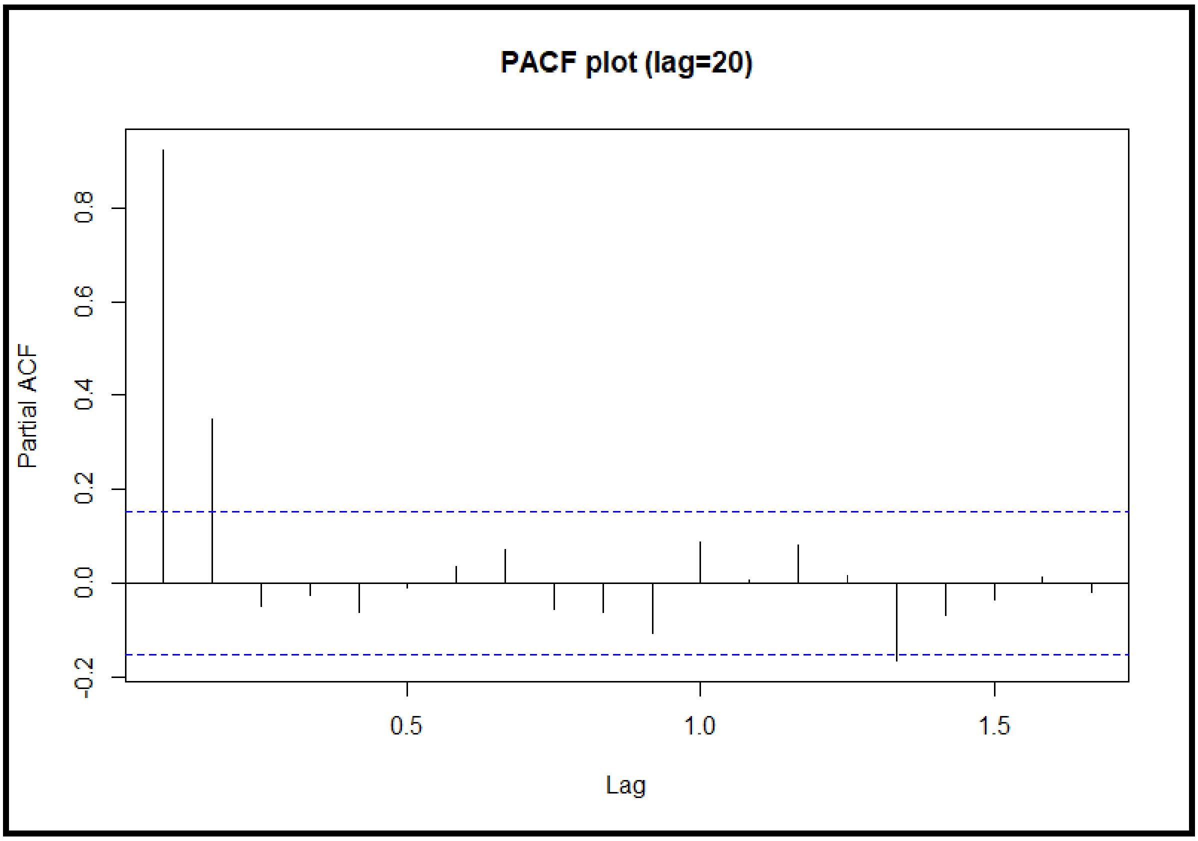
Test for Sample PACF:

Hypothesis:

H0: $\varphi 1 = \varphi 2 = \cdots = \varphi 20 = 0$

H1: At least their exist one non zero φ

Plot 2: Autocorrelation plot



From the plot, it is seen that the PACF plot has a significant spike only at lag 1 and lag 2, meaning that all the higher-order autocorrelations are effectively explained by the lag-1 and 2 autocorrelations. Thus an AR model of order 2 best fits the series.

Method 2: PACF table

Lags(k)	Pacf_value	Test _statistics.	Conclusion
1	0.923304	25.25198404	Rejected
2	0.349781	9.566357227	Rejected
3	-0.05087	1.391139847	Accepted
4	-0.02504	0.684730702	Accepted
5	-0.06324	1.729700192	Accepted
6	-0.01139	0.311518746	Accepted
7	0.034832	0.952630533	Accepted
8	0.070141	1.918317726	Accepted
9	-0.05671	1.551006858	Accepted
10	-0.06297	1.722195802	Accepted
11	-0.10734	2.935750224	Rejected
12	0.085309	2.333158304	Rejected
13	0.004396	0.12023367	Accepted
14	0.081647	2.233022801	Rejected
15	0.016633	0.454909642	Accepted
16	-0.16629	4.547957779	Rejected
17	-0.07012	1.917883498	Accepted
18	-0.03476	0.950776249	Accepted
19	0.012952	0.354239633	Accepted
20	-0.01986	0.543067595	Accepted

From the PACF table, it is seen that from lag 2 onward all of the pacf values are not significant enough; thus, our series follows AR(2) model.

Estimation:

A non-seasonal **ARIMA** model is classified as an "ARIMA(p,d,q)" model, where:

p is the number of autoregressive terms,

d is the number of nonseasonal differences needed for stationarity, and

q is the number of lagged forecast errors in the prediction equation.

Here, previously from the ADF test, we determined that the series needed (at least) one order of nonseasonal differencing to be stationarized. Thus **d=1.**

Also, from the PACF test, we have found that the correlation at lag 1 is significant and positive and shows a sharper "cutoff" than the ACF. In particular, the PACF has only two significant spikes, while the ACF is declining. Thus, the differenced series displays an AR(2) signature. If we, therefore, set the order of the AR term to 2--i.e., we have **p=2** and **q=0;** fit an ARIMA(2,1,0), model. We now obtain the ACF table for the residuals diagnostics.

Residual diagnostics:

To check residuals are white noise or not we perform L Jung Box Test on the residuals.

Table 3:

Lag(k)	res_acf	q_statistics	Critical value df=(k-2)	Conclusion
1	1	0.000362	NA	NA
2	-0.00147	0.009761	NA	NA
3	-0.00746	0.013693	3.841459	Accept
4	-0.00481	0.321357	5.991465	Accept
5	0.042399	1.076062	7.814728	Accept
6	-0.0662	3.703646	9.487729	Accept
7	-0.12313	3.704966	11.0705	Accept
8	0.002752	5.577486	12.59159	Accept
9	0.103291	5.703772	14.06714	Accept
10	-0.02674	6.003624	15.50731	Accept
11	0.041069	7.616461	16.91898	Accept
12	-0.09494	8.595774	18.30704	Accept
13	-0.07374	10.19125	19.67514	Accept
14	-0.09381	10.49006	21.02607	Accept
15	0.040466	15.26191	22.36203	Accept
16	0.161172	15.3646	23.68479	Accept
17	0.023564	15.69339	24.99579	Accept
18	0.042023	15.77739	26.29623	Accept
19	-0.02117	17.1821	27.58711	Accept
20	0.086272	17.20071	28.8693	Accept

Conclusion: From the above table, we see that all the residuals have white noise which means our assumed model, i.e., ARIMA(2,1,0) is correct.

Model Selection:

From the above steps, it is clear that our time series data follows an ARIMA(2,1,0) model. Now, since the next step is forecasting, we fit our original time series data into an ARIMA(2,1,0) model with the help of "***arima()***" function.

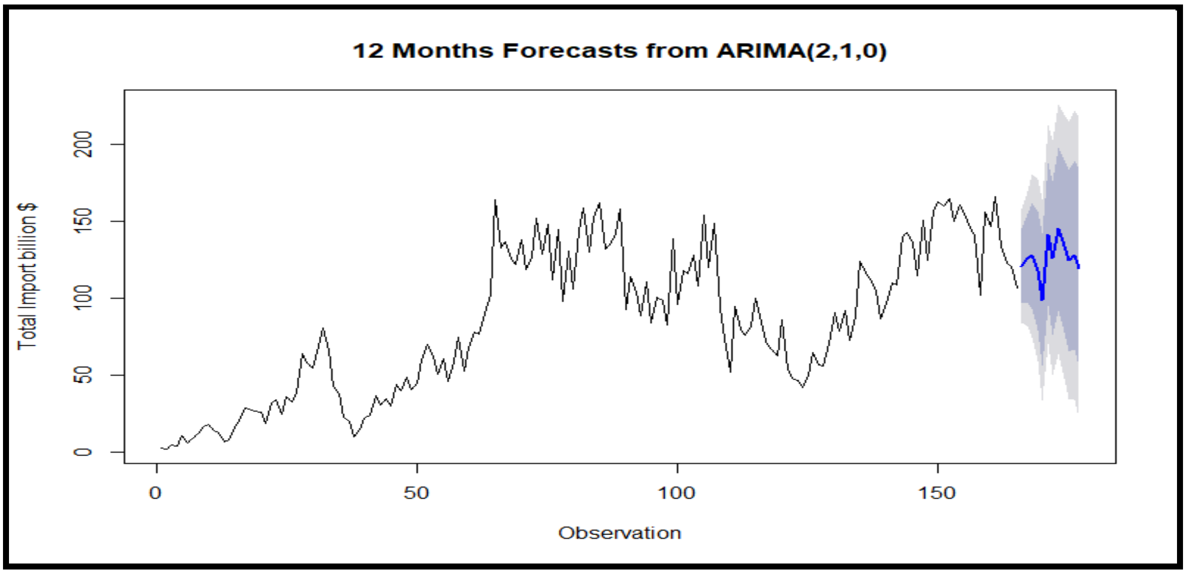
Forecasting:

We have to use the fitted ARIMA(2,1,0) to forecast for the monthly imports to India for the month starting from Sep 2019. The plot of the forecasted values are shown below.

Forecast Table with CI:

Import Month	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Oct-19	121.1844	97.01902	145.3497	84.22667	158.142
Nov-19	125.9792	97.36464	154.5937	82.21702	169.7414
Dec-19	127.9197	93.40255	162.4369	75.13027	180.7092
Jan-20	118.4551	79.72105	157.1891	59.2165	177.6937
Feb-20	99.04734	56.19492	141.8998	33.51022	164.5845
Mar-20	141.6756	95.19592	188.1553	70.59107	212.7601
Apr-20	126.3427	76.44693	176.2385	50.03369	202.6518
May-20	145.3722	92.30101	198.4433	64.20684	226.5375
Jun-20	134.0329	77.9571	190.1087	48.27237	219.7934
Jul-20	124.6674	65.74363	183.5911	34.55127	214.7835
Aug-20	128.1774	66.53558	189.8192	33.90437	222.4504
Sep-20	119.8458	55.60149	184.0901	21.59259	218.099

The plot of the Forecast values:



Conclusion: From the above analysis it was found that if the situation had been normal; between Sep 2019 to Sep 2020, the average Import to India in the year 2020 would have lesser than the average Import of 2019. Moreover, India would have the lowest export of 99 billion US dollar in March 2020 and the highest imports of 145 billion US dollars in March 2020.