Indian Trade Analysis and Forecasting

Project Report

Akarsh Somani (162) and Gaurav Misra (172)

Objective –

Trade is the economic concept which invokes on BUY and Sell of the commodities, or exchanging goods and services. Trade increases competition and decreases overall world wise cost for a product. Here we will be dealing with the Indian trade with the other countries and the impact of the trade. Trade is one of the important factors for the economy of our country and that is why this Indian trade data analysis and forecasting is very important.

Here we will be forecasting using different models and check the prediction with the available data.

Overview about the data –

The data is taken from the MINISTRY OF COMMERCE AND INDUSTRY DEPARTMENT, GOVT. OF INDIA.

The data is monthly data from Jan 2006 to Sept 2019.The value is in million US Dollar $ and the stats are as follows–

	Import	Export
count	165.000000	165.000000
mean	32172.849333	21185.384848
std	9151.546169	5976.205557
min	11479.690000	8624.660000
25%	25868.920000	15757.360000
50%	33772.550000	23012.240000
75%	39966.570000	25949.040000
max	46618.800000	32717.300000

Methods –

We built seven models to forecast –

- Exponential Smoothing
- Auto Regressive
- Moving Average
- Holt-Winters model
- ARIMA Multiplicative
- ARIMA Additive
- Seasonal ARIMA

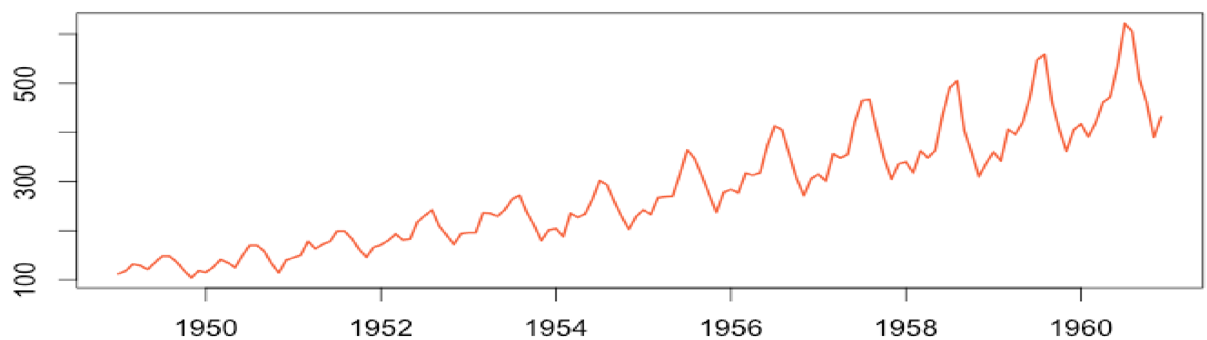
Now we will go one by one with each model and see there methodologies and prediction –

Lets first discuss the types of timeseries –

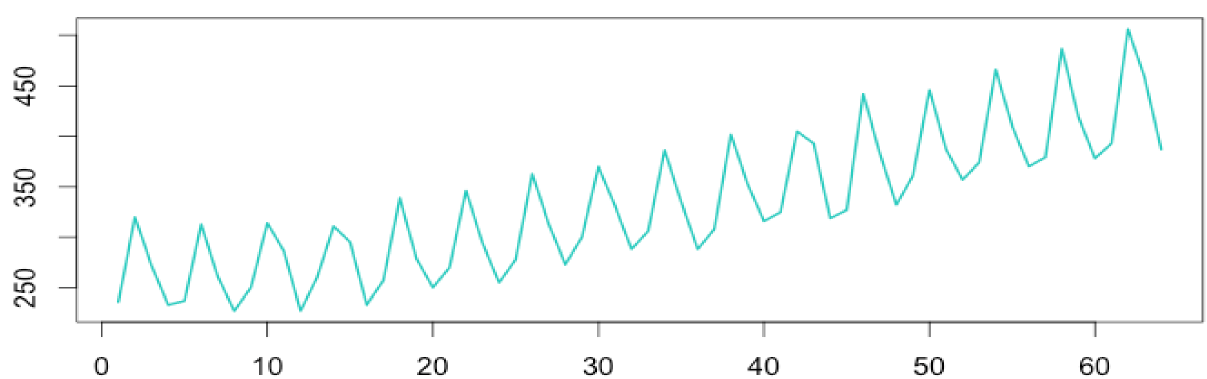
There are basic two kind of timeseries – Multiplicative and Additive

If the amplitude does not vary then that is Additive and if amplitude varies then that is Multiplicative.

Multiplicative Time Series = Trend * Seasonality * Randomness



Additive Time Series = Trend + Seasonality + Randomness



Exponential Smoothing(Exponential Averaging) –

Exponential Smoothing is the technique for smoothing(Averaging) the timeseries using exponential window function. In simple moving average the past observation are equally weighted where as here it exponentially decreases over the time.

In this method we use

$$s_0 = x_0$$
$$s_t = \alpha x_t + (1-\alpha)s_{t-1},\ t > 0$$

where α is the *smoothing factor*, and $0 < \alpha < 1$.

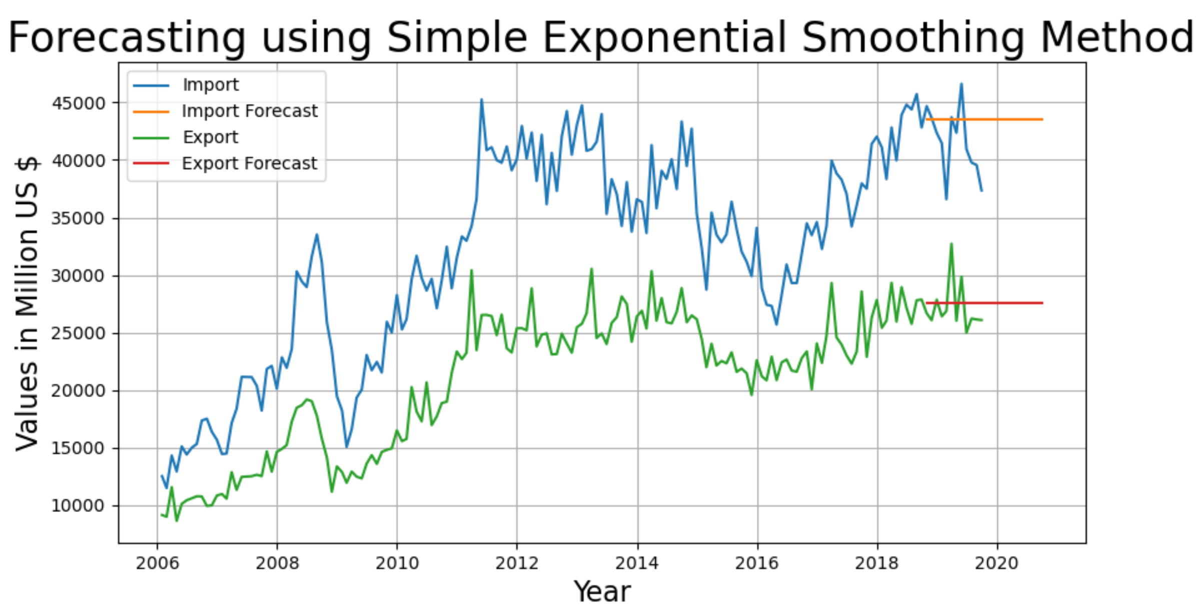
Where S_0 is the value at time t = 0 and the forecast at time t is given as S_t

The term smoothing factor applied to α here is something of a misnomer, as larger values of α actually reduce the level of smoothing, and in the limiting case with α = 1 the output series is just the current observation.

As mentioned, it uses the exponential window function, we substitute the value of the above equation back to itself.

$$\begin{aligned} s_t &= \alpha x_t + (1-\alpha)s_{t-1} \\ &= \alpha x_t + \alpha(1-\alpha)x_{t-1} + (1-\alpha)^2 s_{t-2} \\ &= \alpha\left[x_t + (1-\alpha)x_{t-1} + (1-\alpha)^2 x_{t-2} + (1-\alpha)^3 x_{t-3} + \cdots + (1-\alpha)^{t-1}x_1\right] + (1-\alpha)^t x_0. \end{aligned}$$

In other words, it forms a GP which is a discrete version of an exponential function.



RMSE Import/Export - 3454.44, 2073.22

Auto Regressive Model –

It is used when a value from the time series has dependency on previous values, like Y_t on Y_{t-1} .

Ex - $Y_t = f(Y_{t-1})$

The order of an autoregression is the number of immediately preceding values in the series that are used to predict the value at the present time.

An order of 2 denotes that the time t is predicted based on t-1 and t-2.

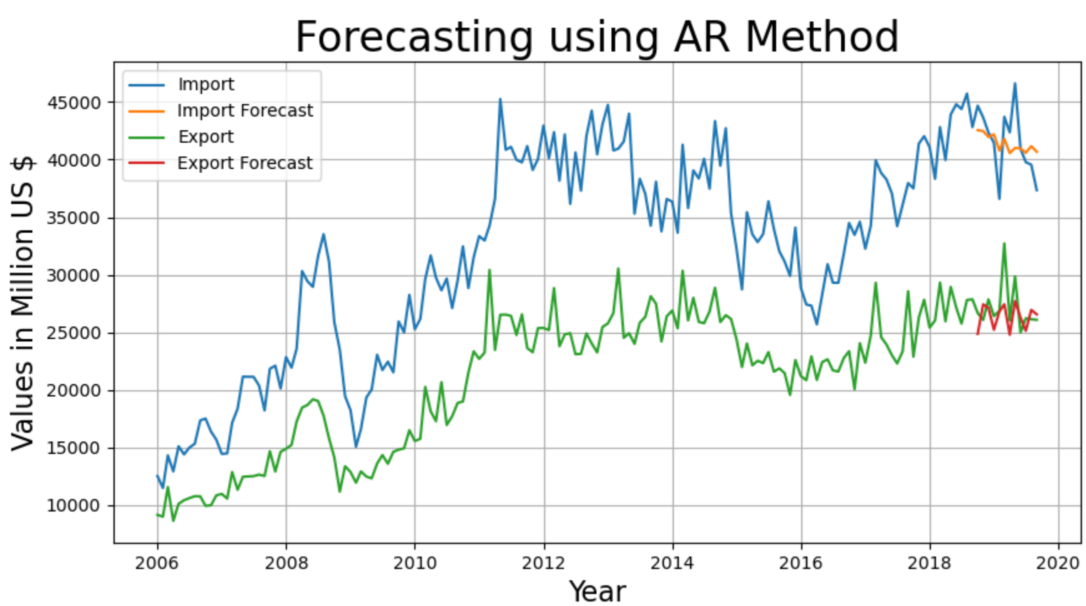
Terminologies - Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF).

ACF = CORR (Y_t , Y_{t-k}) ,

PACF can be calculated as

- Remove the linear dependency from the timeseries
- Calculate the correlation.

PACF is used to find the order of the autoregressive model.

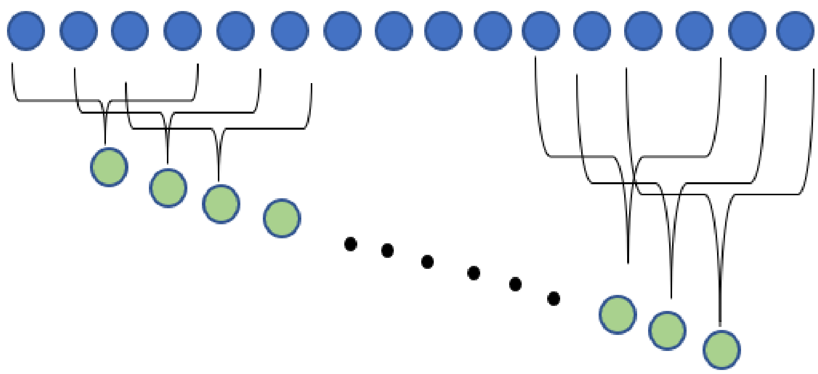


RMSE Import/Export - 2530.75, 1932.5

Moving Average Model –

Moving Average is a technique that calculates the overall trend in the dataset. As the name suggest that we go by taking the Average over a fixed rolling size window. The MA_t is calculated by taking the unweighted mean of the previous window_size (here 4) data.

The moving average is thus calculated as –

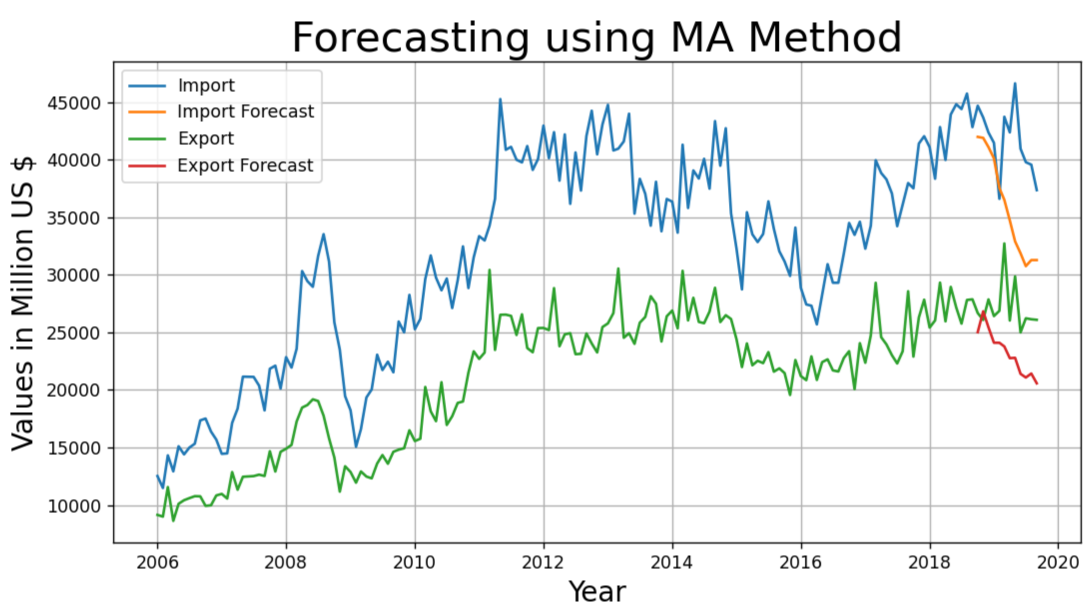


$MA_4 = X_4 + X_3 + X_2 + X_1$

And the successive value can be calculated as –

$MA_n = MA_{n-1} + (X_n – X_{n\text{-window\_size}})$

In our model we have chosen the window of 12



RMSE Import/Export - 6969.83, 4609.61

Holt-Winters Method –

It is the extension over the simple exponential smoothing method. Here we use triple smoothing with the factor - seasonal period, trend type and seasonal type. Here seasonal and trend type means *Multiplicative* or *Additive*.

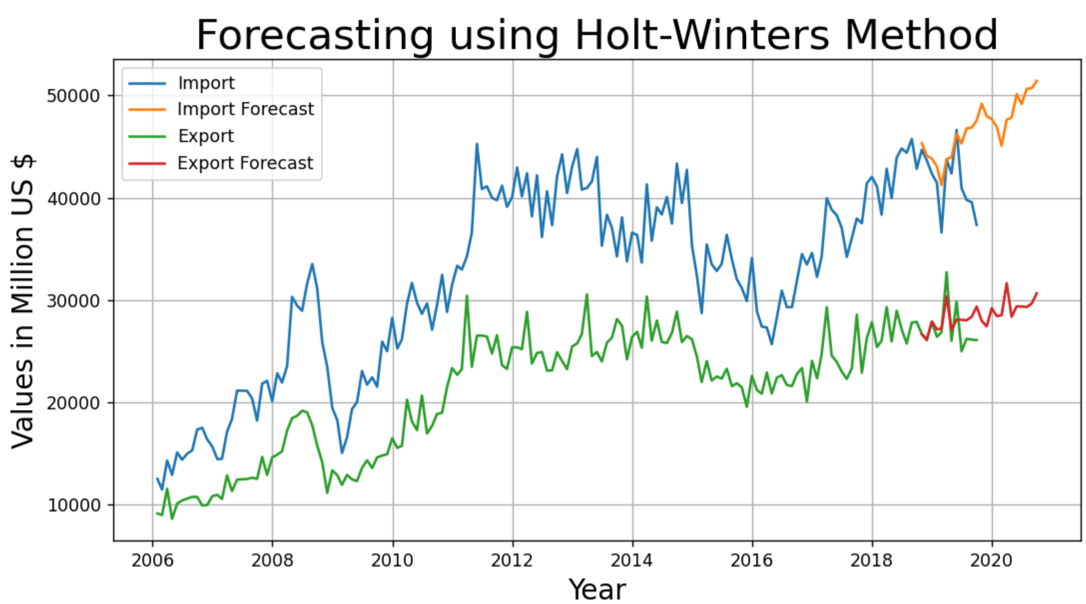
s_t represents the smoothed value of the constant part for time t. b_t represents the sequence of best estimates of the linear trend that are superimposed on the seasonal changes. c_t is the sequence of seasonal correction factors. c_t is the expected proportion of the predicted trend at any time t mod L in the cycle that the observations take on. As a rule of thumb, a minimum of two full seasons (or 2L periods) of historical data is needed to initialize a set of seasonal factors.

The output of the algorithm is again written as F_{t+m}, an estimate of the value of x at time t+m, m>0 based on the raw data up to time t. Triple exponential smoothing with multiplicative seasonality is given by the formulas

$$\begin{aligned}
s_0 &= x_0\\
s_t &= \alpha\frac{x_t}{c_{t-L}} + (1-\alpha)(s_{t-1}+b_{t-1})\\
b_t &= \beta(s_t - s_{t-1}) + (1-\beta)b_{t-1}\\
c_t &= \gamma\frac{x_t}{s_t} + (1-\gamma)c_{t-L}\\
F_{t+m} &= (s_t + mb_t)c_{t-L+1+(m-1) \mod L},
\end{aligned}$$

where α is the data smoothing factor, 0 < α < 1, β is the trend smoothing factor, 0 < β < 1, and γ is the seasonal change smoothing factor, 0 < γ < 1. Reference

In our model we have used seasonal period = 12 and trend type and seasonal type as multiplicative.



RMSE Import/Export - 3768.38, 1642.84

ARIMA Multiplicative –

ARIMA stands for *Autoregressive Integrated Moving Average* which is a combination of three terms –

The AR part of ARIMA indicates the evolving variable of interest is regressed on its own lagged (i.e., prior) values. The MA part indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in the past. The I (for "integrated") indicates that the data values have been replaced with the difference between their values and the previous values (and this differencing process may have been performed more than once). The purpose of each of these features is to make the model fit the data as well as possible.

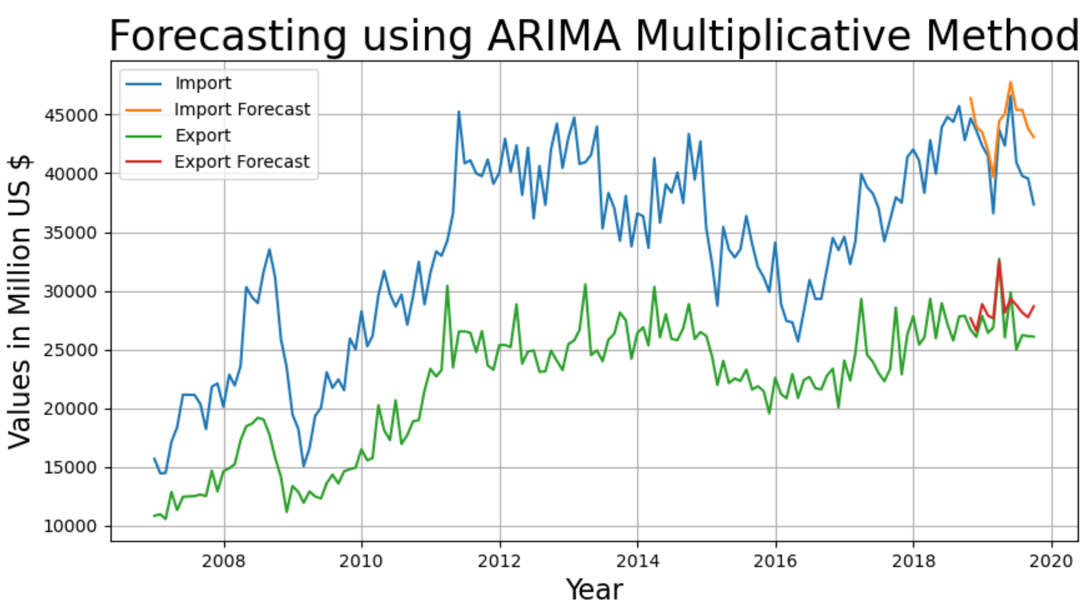
Non-seasonal ARIMA models are generally denoted ARIMA(p,d,q) where parameters p, d, and q are non-negative integers, p is the order (number of time lags) of the autoregressive model, d is the degree of differencing (the number of times the data have had past values subtracted), and q is the order of the moving-average model.

We have already seen the Autoregressive and Moving average individually.

Multiplicative Time Series = Trend * Seasonality * Randomness, As explained previously.

For predicting the p and q we use partial autocorrelation and autocorrelation function. We have already discussed about them earlier. We used plots to find the optimal value of p and q graphically.

Here we are taking the (p,d,q) = (1,0,5). d as 0 because we do not require the differencing as our randomness is stationary.



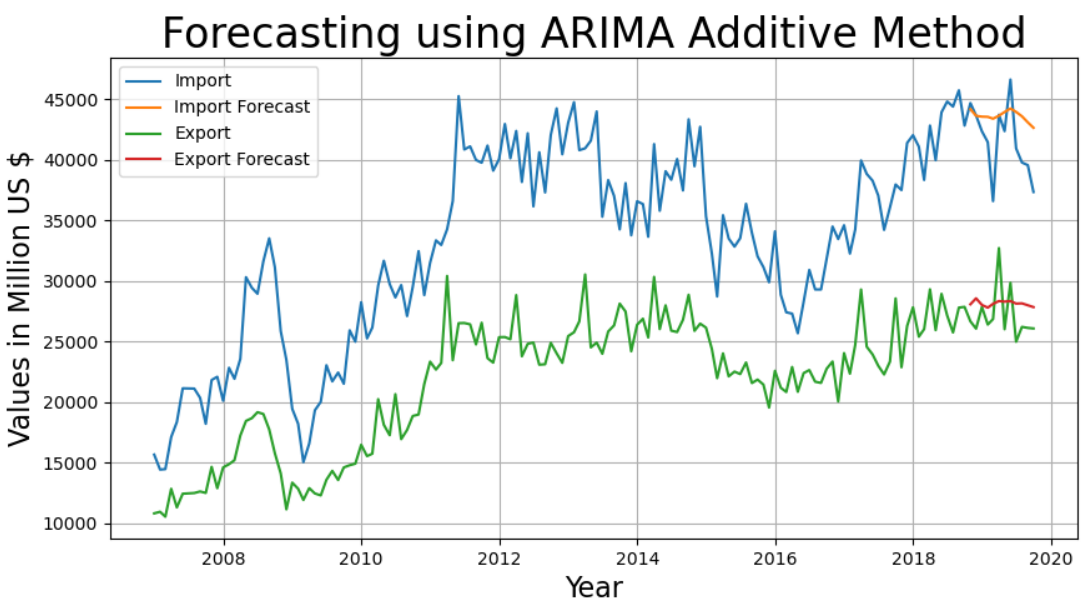
RMSE Import/Export - 3238.24, 1762.79

ARIMA Additive –

This is similar to the above method except for the fact that we have considered our time series as Additive (just for comparison).

Additive Time Series = Trend + Seasonality + Randomness

Here we are taking the (p,d,q) = (1,0,5). d as 0 because we do not require the differencing as our randomness is stationary.

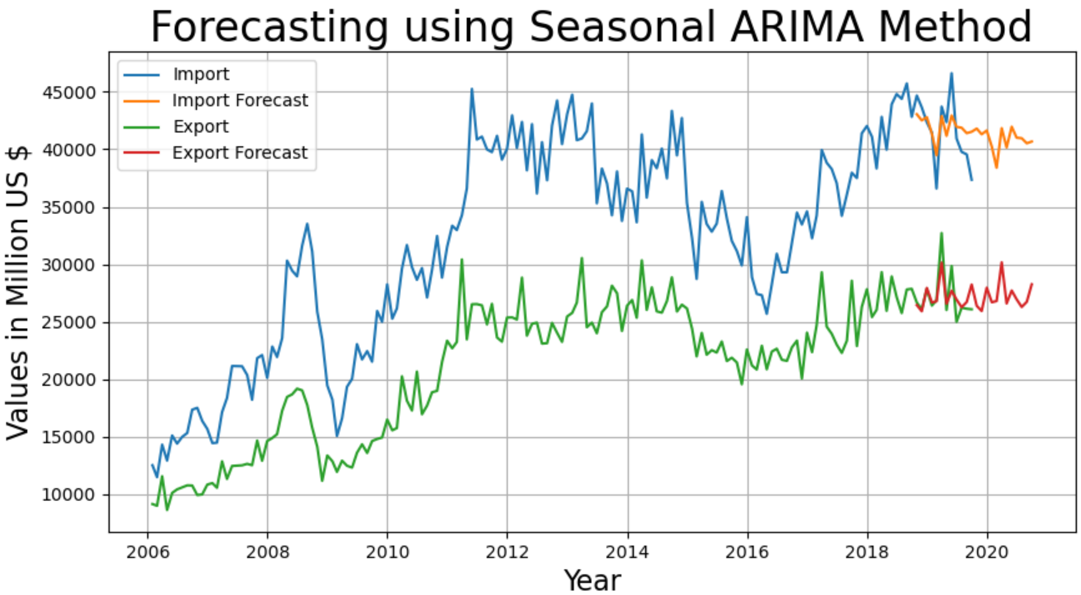


RMSE Import/Export - 3221.36, 2210.54, we can see that the forecast does not fits right.

Seasonal ARIMA –

Seasonal ARIMA here is ARIMA Multiplicative method with a seasonality factor m. Here we have chosen the multiplicative factor of 12 as ours is a monthly data.

Also we used brute force to find the best p,d,q values which reduces the RMSE. Import (1,0,0) and Export (1,0,2). This model has performed quite well and thus below is the plot.



RMSE Import/Export - 2130.2, 1305.91

Conclusion –

All the models perform quite well but the best observation which we have observed is of Seasonal ARIMA.

Below is the RMSE Comparisons –

Model\Error	Import RMSE	Export RMSE
Exponential Smoothing	3549.73	2090.32
Auto Regressive(AR) model	2530.75	1932.50
Moving Average(MA) model	6969.83	4609.61
Holt-Winters Method	3768.38	1642.84
ARIMA Multiplicative	3238.24	1762.79
ARIMA Additive	3221.36	2210.54
Seasonal ARIMA	2130.20	1305.91

The comparisons are so close because one model is the extension of the other with some advancement.

Holt-winters is the advancement of the Simple exponential smoothing. ARIMA is the combination of AR and MA model, Seasonal ARIMA is the advancement of the ARIMA model.

The results are quite satisfactory and we see that the there is a downward trend in Import forecast and upward trend in the Export forecast. The more we export and the less we import is beneficial for the Indian economy. This shows that India will be playing pretty good in terms of trade.