

Report

Problem statement

Forecasting revenue of us air traffic for the year from 2001 to 2024 based on data available from 2000 to 2020

Data collection:

Data collected from this site <https://www.transtats.bts.gov/TRAFFIC/>

Preprocessing

As the date is given in month and year we had to convert it in a proper date format which is suitable for time series

By plotting a graph we check the trend in the data

Checking the stationarity:

We check the Stationarity using adfuller test where we check T statistic against the critical value for rejecting the null hypothesis

Checking the Seasonality:

By using Seasonal decompose we check whether there is any seasonality present in the data by plotting we come to a conclusion that there is a seasonality in the data

Model Selection:

As our data is seasonal we tend to use SARIMA and SARIMAX with order as (1,1,1) and seasonal order as (2,0,1,12) and we get the best possible outcome as log likelihood is Maximum with 348.401 and AIC with lowest(-684.801) and BIC(-665.257) and all the p-value less than 0.05

Inspecting the model by checking residuals:

We check the residuals and as a result are residuals are normally distributed

Testing the Model:

- Forecasted values for 4 years and as a result are model captures the data very well with lowest MSE of our forecast has a value of 0.012, which is very low as it is close to 0.
- An MSE of 0 would that the estimator is predicting observations of the parameter with perfect accuracy, which would be an ideal scenario but it not typically possible.

Conclusion

The best Season model is SARIMAX with order and best seasonal order having the log likelihood as 348.401 and AIC and BIC as this with order as (1,1,1) and seasonal order as (2,0,1,12) and we get the best possible outcome as log likelihood is Maximum with 348.401 and AIC with lowest(-684.801) and BIC(-665.257) and p value less than 0.05