TIME SERIES FORECASTING

PRESENTED BY
Tanukrishna Chetia

Executive Summary

In this assignment we are given the Sales of souvenir data. We have arrived at the 5-year predictions through the below mentioned methods.

Part A)
Using the Winter-Holts methods, data prediction for the next 5 years will be provided.
Our output contains the complete modeling steps with explanations. R-code where applicable is shown and images included appropriately.

Part B)
Using the ARIMA method, data prediction for the next 5 years will be provided. Our output contains the complete modeling steps with explanations. R-code where applicable is shown and images included appropriately.

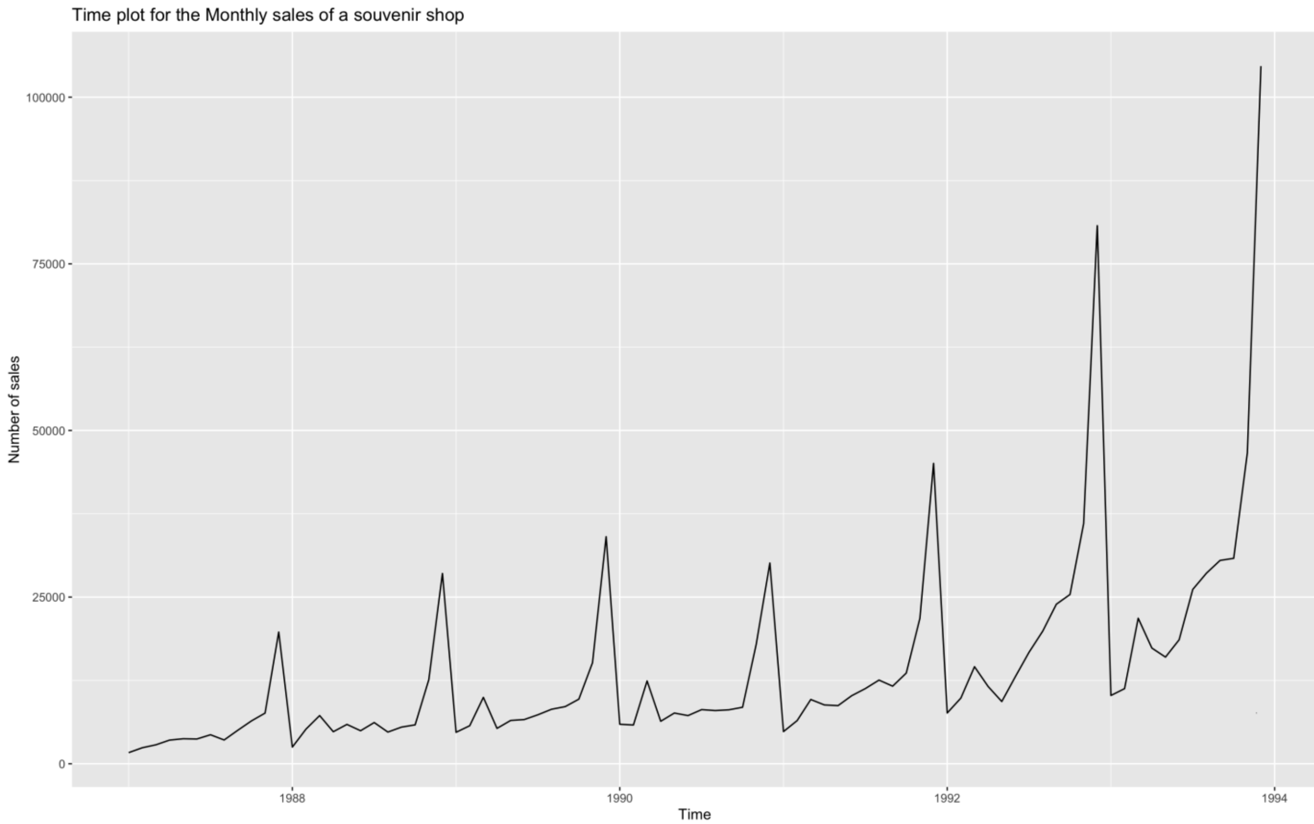
Criteria

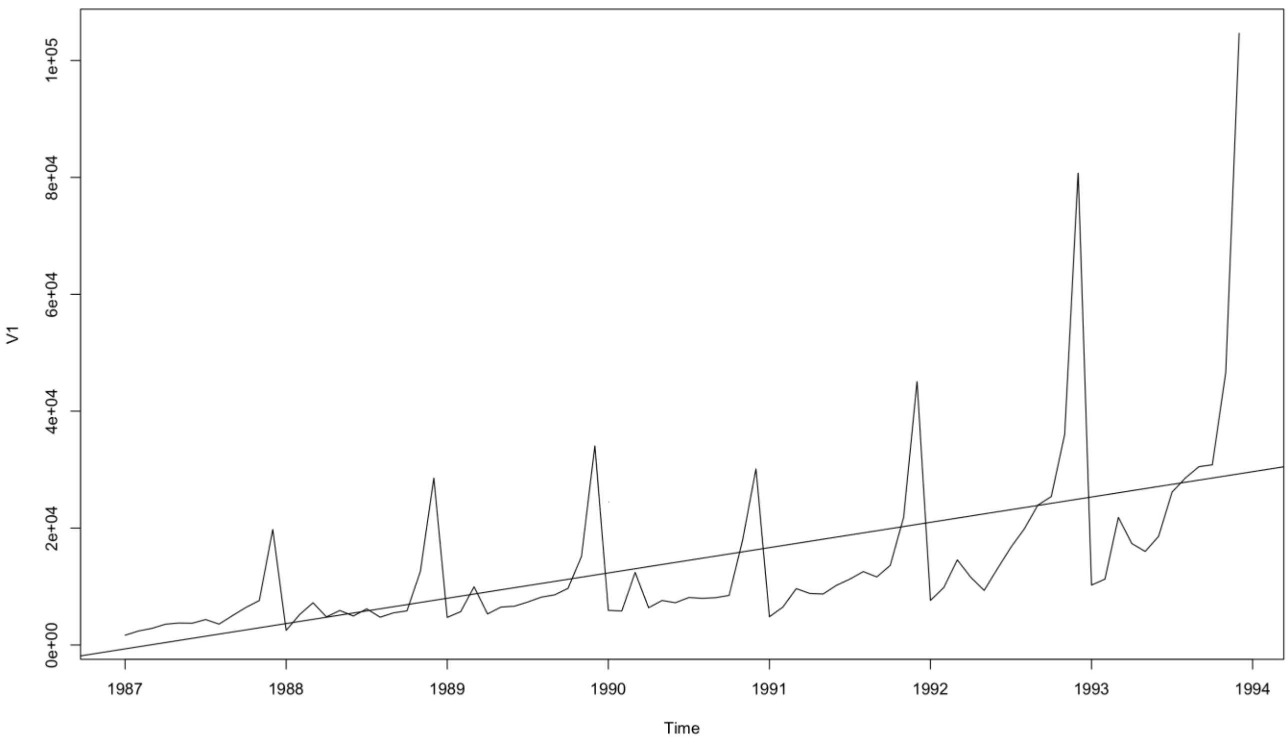
1.a) Visualizing the whole data, dividing the time series data into different components and plotting the different components individually

The given data was in data.frame format and we have converted the same into Time series.

Initial ObservationsEvery year there has been a gradual increase in the sales from the month of January to December. The sales is at its peak during the month of December every year.

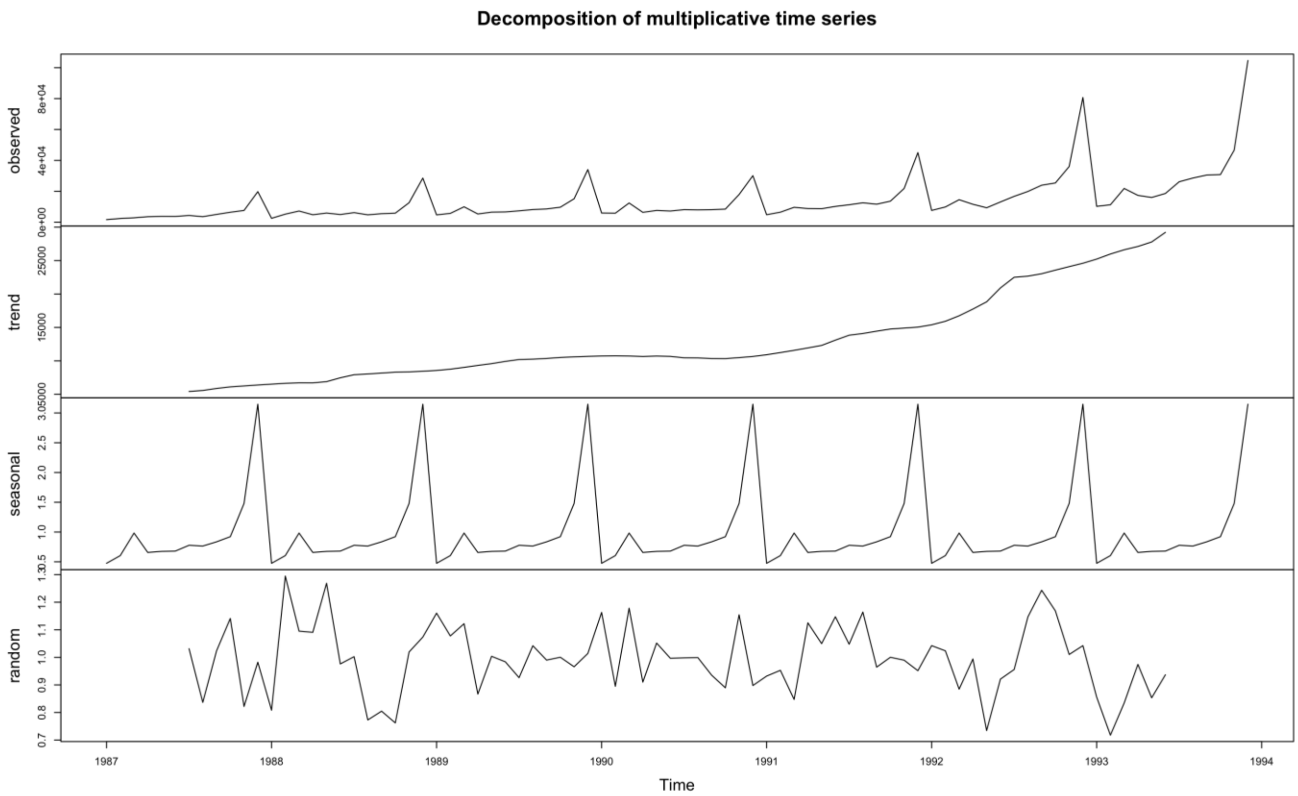
1. Drop in sales is observed in April and August from their previous months in majority of the years.
2. Overall there is an increase in the sales from 1987 to 1993.
3. Sudden drop in sales is observed in January (every year).


Time plot for the Monthly sales of a souvenir shop
Number of sales
100000
75000
50000
25000
0
1988
1990
1992
1994
Time

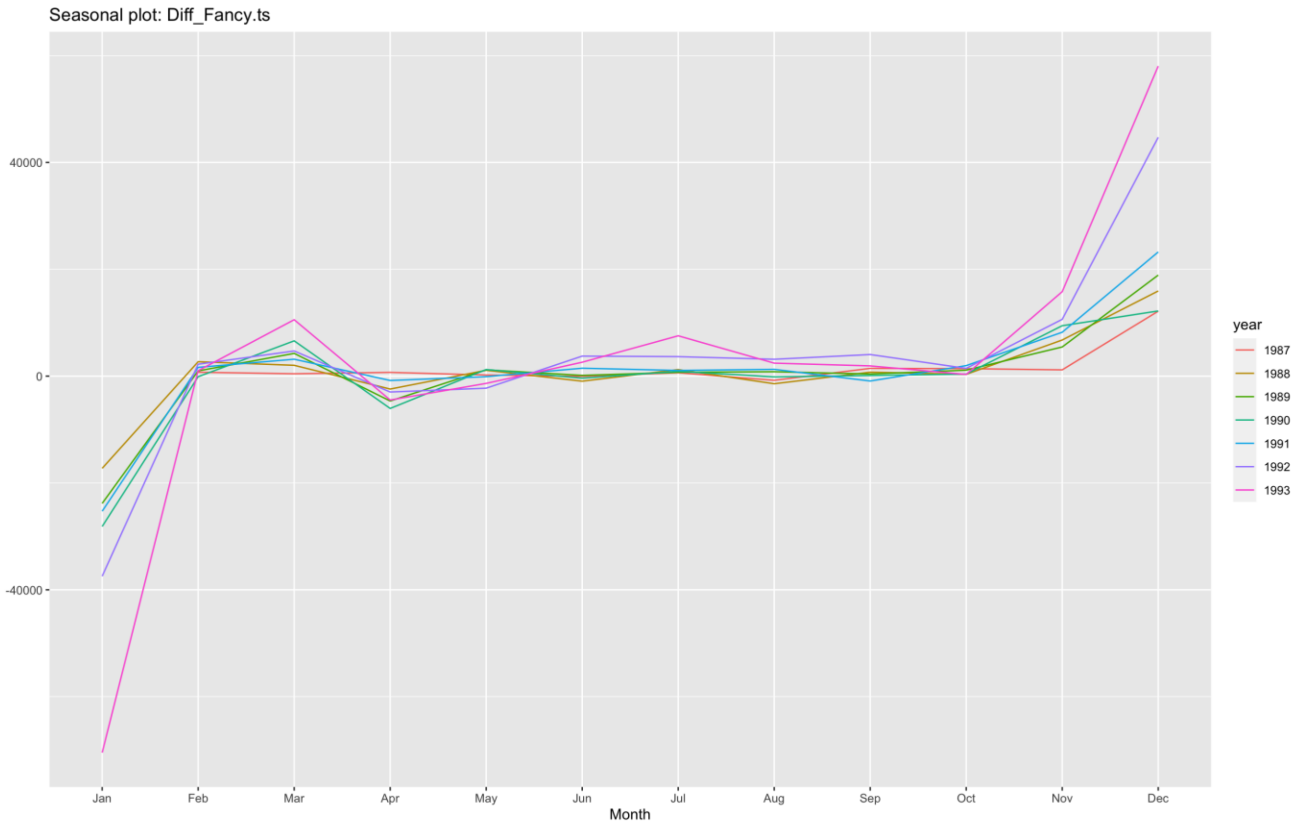

V1
1e+05
8e+04
6e+04
4e+04
2e+04
0e+00
1987
1988
1989
1990
1991
1992
1993
1994
Time

ObservationsWe can observe that there is an increasing trend in sales over the years. Hence the dominant component in our time-series is Trend.
Check Presence of seasonality by decomposing data

1. Decomposing the time-series into components
2. We go for the multiplicative decomposition because it is observed that the variation of data increases with every year.



3. From the decomposition plot it can be seen that the variation in the time-series data repeats itself every year.
4. The data includes both seasonality and trend which is depicted in the above graph.
5. The seasonality can be better understood when we overlap the seasonality curves for each year. The graph for the same is given below.



6. Seasonality is clear in the above plot where there is a big drop in January every year and big rise in the months of November and December.

Note:
Our series has trend and seasonality.

1.b) Time series stationarity check

Check -1: Visuality
From the plots above it can be observed that the Fancy dataset time-series exhibits non-stationarity because of the non-constant mean (increasing best fit line) and non-constant variance (increasing from left to right).
Check -2: Using KPSS test and ADF test The p-value from KPSS test is 0.01 which is less than 0.05

1. The p-value from ADF test is 0.54 which is greater than 0.05

The above observations show that the data series is non-stationarity. This requires us to perform log transformation. Log transform is probably the most commonly used transformation, if you are seeing a diverging time series. However, it is normally suggested that we use transformation only in case differencing is not working.

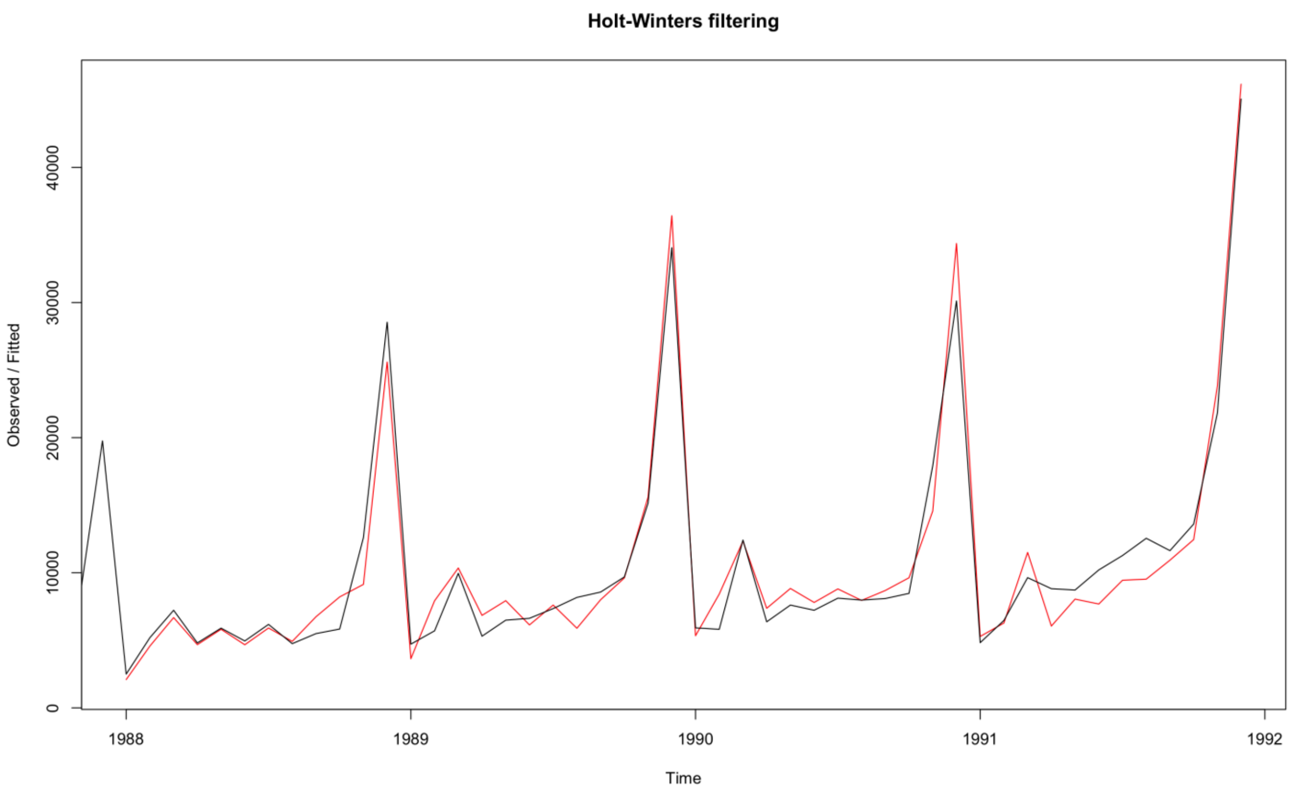
2. Splitting data into Test/Train or Dev/holdout

The data has been split into 5 and 2. Where the first 5 years have been taken for training and the last 2 have been taken for data testing.

- 1987 Jan – 1991 Dec is our training data.
- 1992 Jan – 1993 Dec is our testing data.

3.a) Holt Winter Model creation

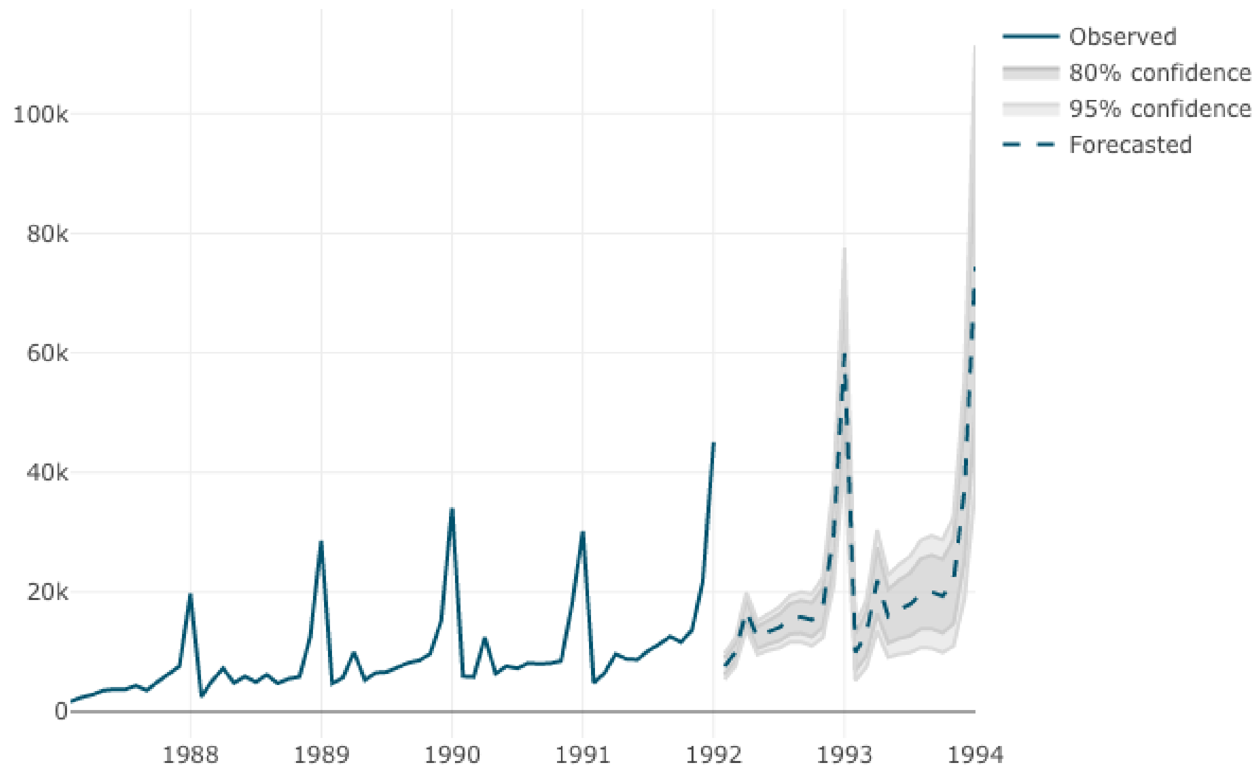
Since we have both trend and seasonality we are good to go with the Holt Winter Model. Also, the variance is higher when the time series is higher, hence we chose the multiplicative model.



Our graph here shows that the data captured from training years and that from model is quite similar. We shall draw an inference that the model and training data fit is apt.

3.b) Predicting the values for test dataset using HW model

While forecasting for the next two years based on the model, we have the below graph.



To check the stability of the model we are using the Ljung-box test. This tests the correlation of the residuals. The p-value from the box test is 0.75 which is more than 0.05, this shows that there is enough evidence of non-zero co-efficient between the residuals. Stability of the model hence proved.

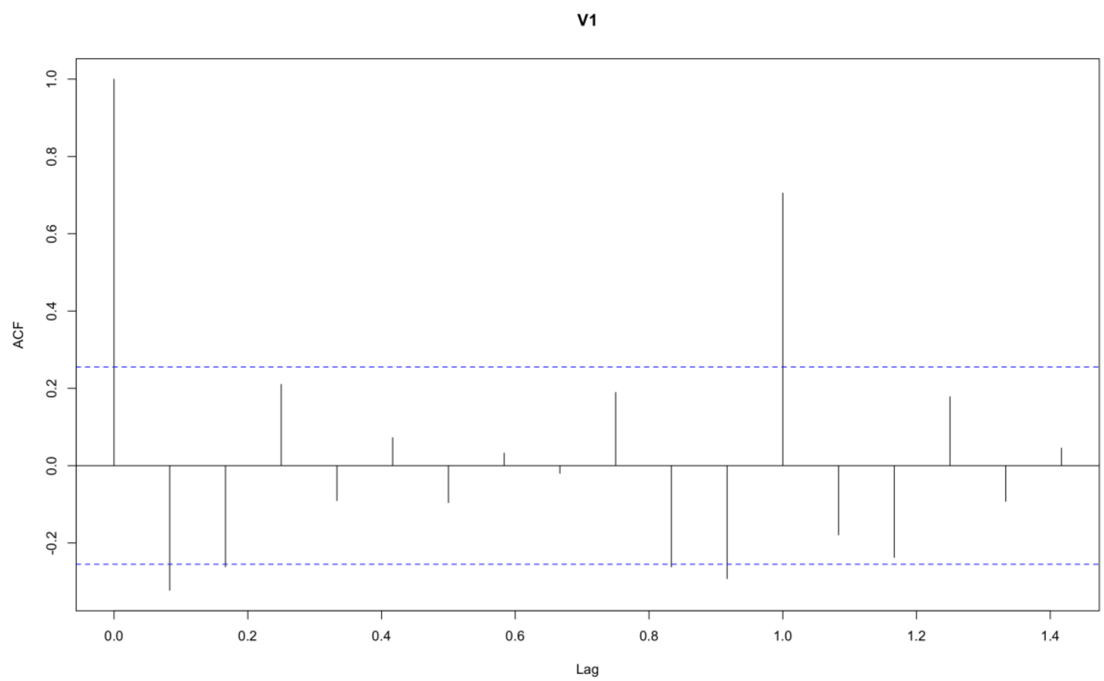
3.c) Validation against actual values using MAPE

By validating against the actual values with our forecasted values using HW, the MAPE gives accuracy of 13.49. General notion of the accuracy being 10 when the model fits is tested and since our output is closer, we are having a fair model.

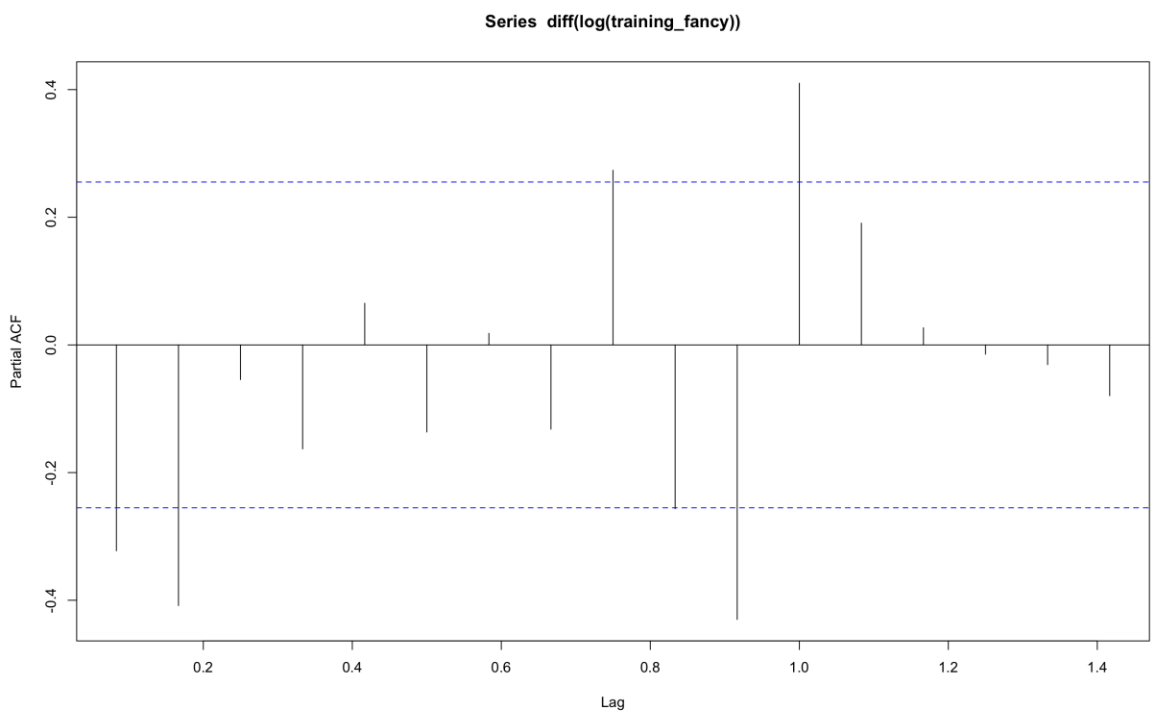
4.a) ARIMA Model creation

To create an ARIMA model we need the P, D, Q values.

1. Q value is the moving average value, we have arrived at it from the ACF function. Q=2



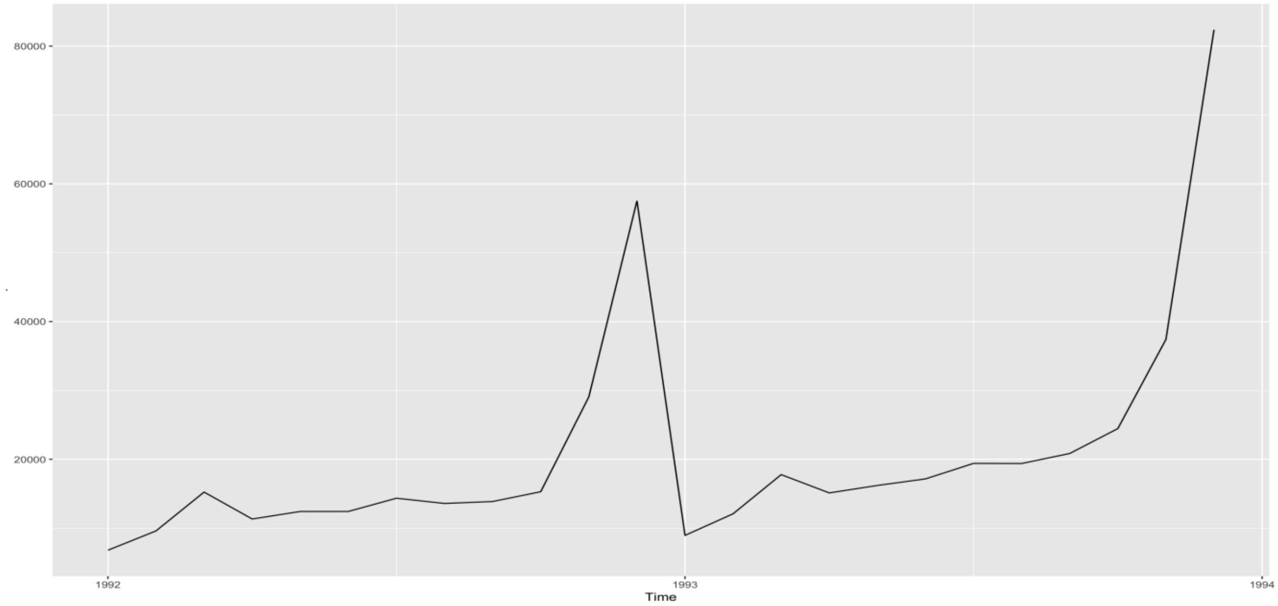
2. P is the number of autoregressive terms. P = 2



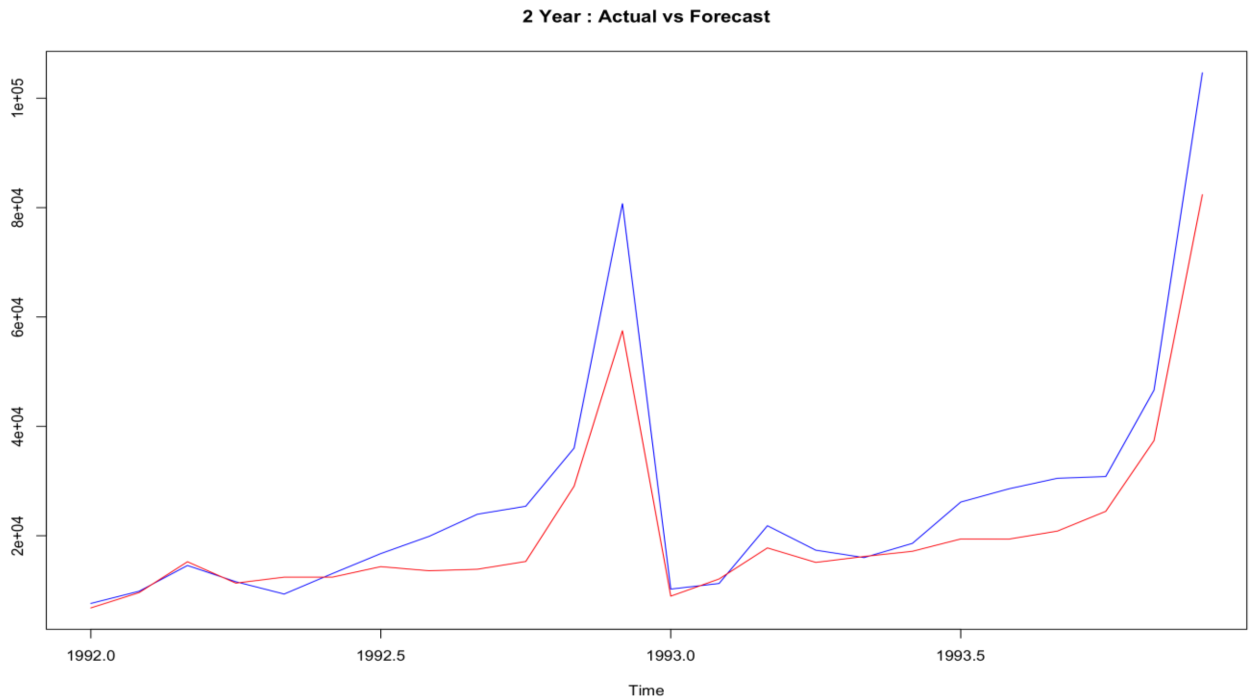
3. D is the number of differences needed for stationarity. D = 1

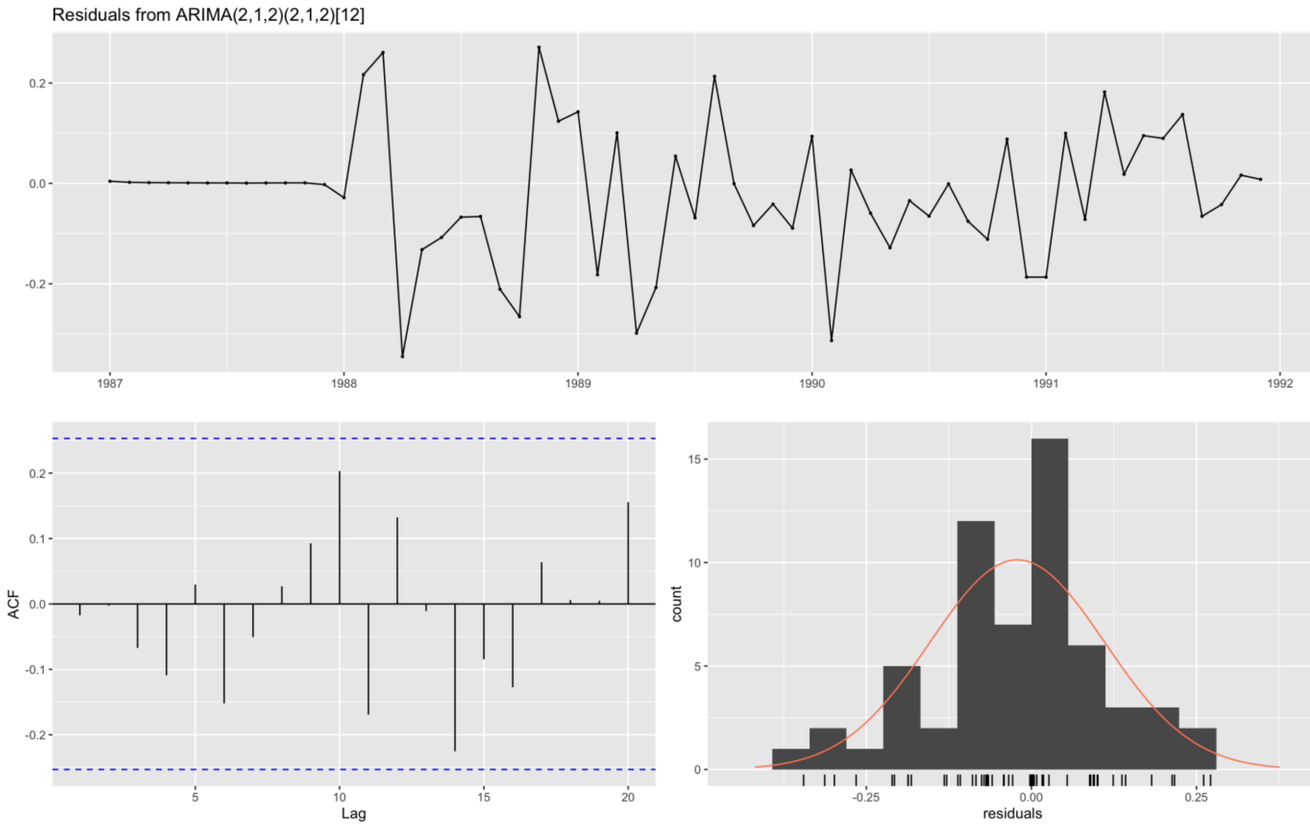
4.b) Predicting the values for test dataset using ARIMA model

With these values and the training data, we created the ARIMA model through which we predicted for 2 years. The below plot for reference:



EVALUATING OUR MODEL WITH TEST DATA





Residual Histogram:
From the above graph, we see that residuals of the ARIMA model seems to better follow a normal distribution.

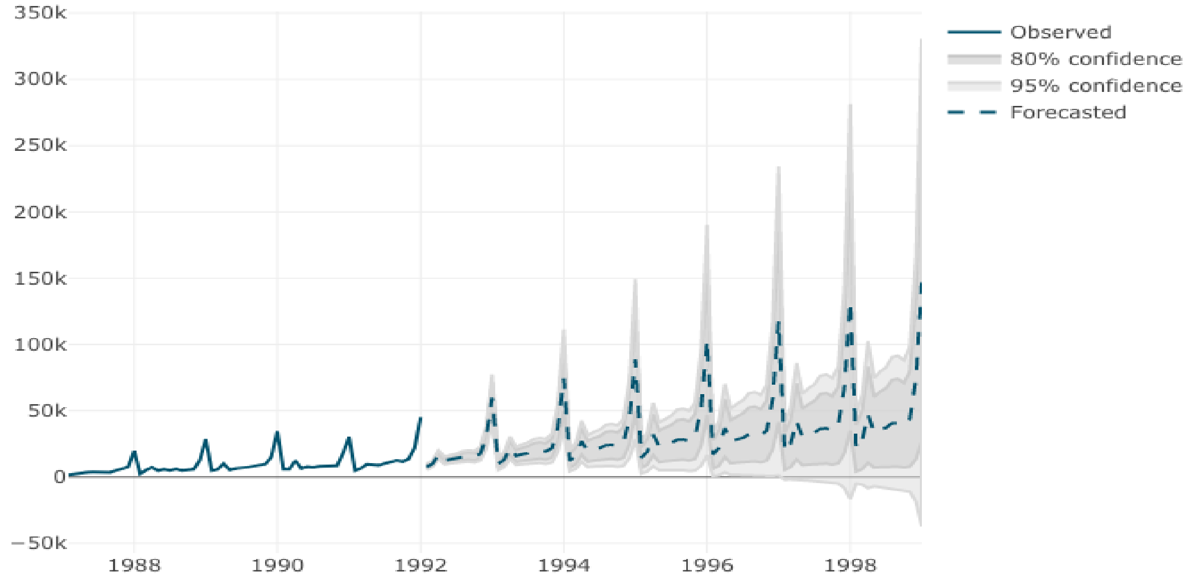
ACF Plot:
The residuals of the model are not correlated.

4.c) Validation against actual values using MAPE

By validating against the actual values with our forecasted values from ARIMA, the MAPE gives accuracy of 28.18.

5.a) Use both the models to predict the values for next 5 years using HW model

Based on our HW model, we have predicted for 5 years and the below graph depicts the plot.



5.b) Use both the models to predict the values for next 5 years using ARIMA model

The below graph uses both the models to predict the values for the further 5 years using ARIMA model.