**Team Name:** **MASK**

**Team Members:**

1. Chinnasubbareddygari Mohan Reddy (A0163433L)
2. Anusuya Manickavasagam (A0163300Y)
3. Sridharan Kesavan (A0163207M)

**Dataset used:** ActualRatings\_weeklyGRP.xls which is the TV rating data of an Indian network.

**Problem statement:** To try the following methods: Exponential Smoothing, ARIMA, Decomposition methods, Time series Regression on the dataset and to select the best technique.

**Tool used:** **“Rstudio” and “Excel”**

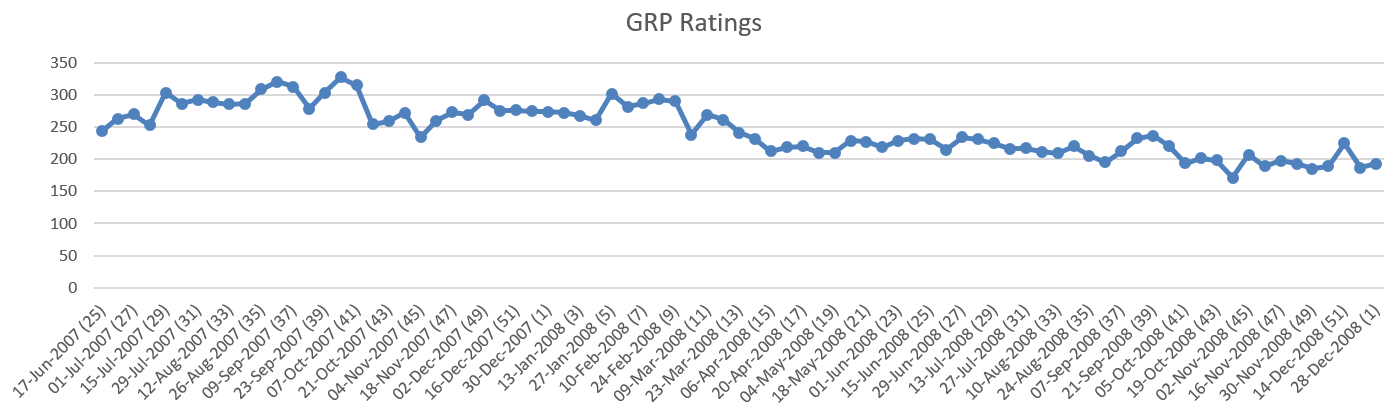
**Summary:** 1. The GRP ratings dataset was taken and the training data was taken from 17-Jun-2007 to 28-Dec-2008 and the graph was analyzed. There was no seasonal effect found on data, only a downward trend was observed.

2. The various Time series methods were applied on the data like Holt’s exponential smoothing, ARIMA, Simple moving average, Time series regression were performed on the data.

3. Forecasting was done for the various methods and the various errors were calculated.

**Steps:**

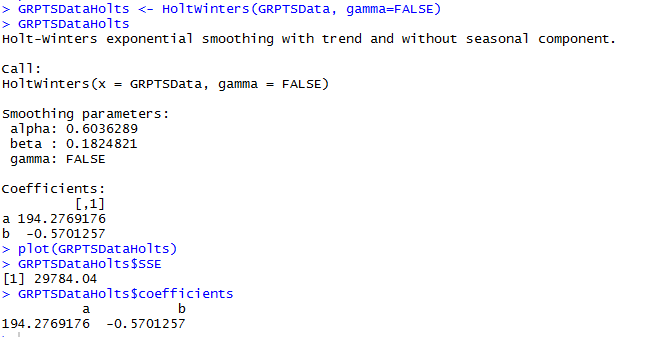
We plot the graph till 28th Dec 2008 from the dataset and observe the graph.



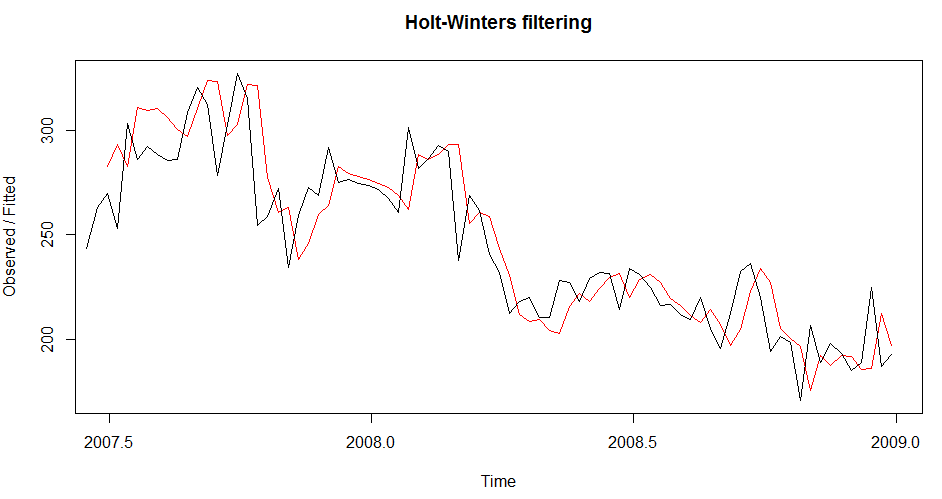
We observe the graph and see that it follows a somewhat downward trend, and to find seasonality we have done the “**Decomposition by BATS model”** test using **“tbats()”** in r and found seasonality to be false.

|  |  |
| --- | --- |
|  |  |

1. **Exponential smoothing:** Since we have a model with decreasing trend and no seasonality, we use **Holt’s exponential smoothing** to make short term forecasts.

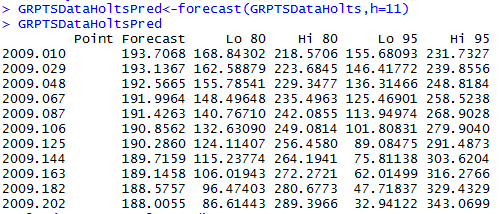


Predicted vs observed values by Holts exponential smoothing

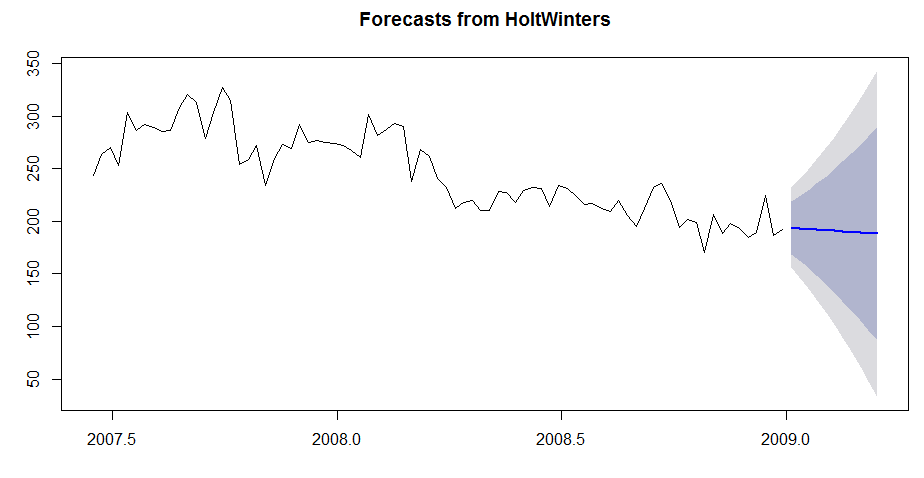


**Forecasting next 11 weeks GRP ratings using Holts Exponential smoothing**

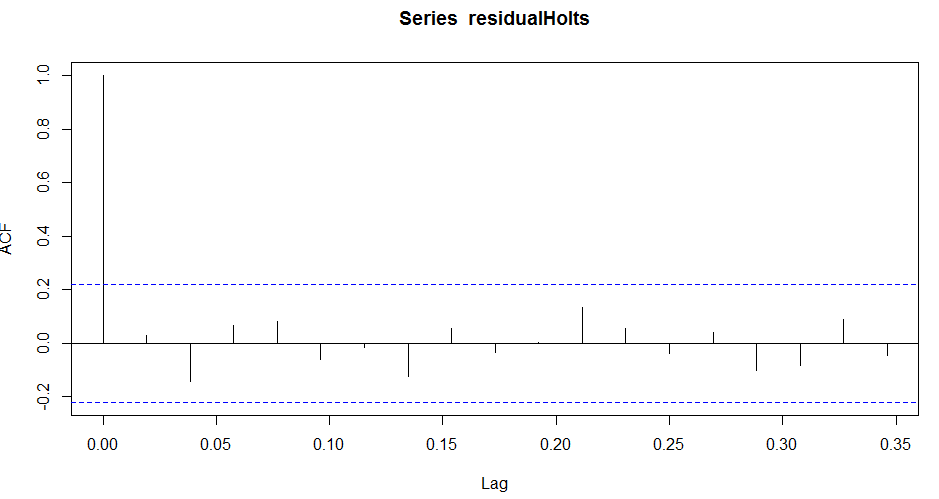
**Predicted value**



The forecasts are shown as a blue line with 80% prediction intervals in the dark grey area and the 95% prediction intervals in the light grey area.

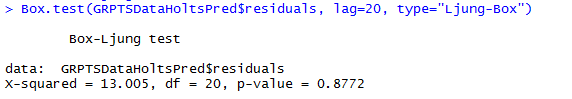


The **residuals ACF values** are all with in the significant line.



**Ljung-Box test**

The Ljung-Box test statistic is 13.005, and the p-value is 0.8772, so there is little evidence of non-zero autocorrelations in the in-sample forecast errors at lags 1-20.



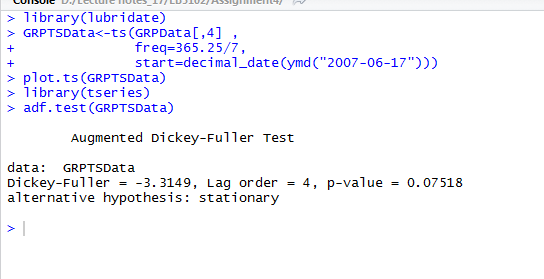
1. **ARIMA:** To fit the ARIMA model we first see the stationarity of the graph. We see that the initial graph does not look stationary.

**ADF Test:** Null hypothesis: There is a unit root

Alternate hypothesis: The data is stationary

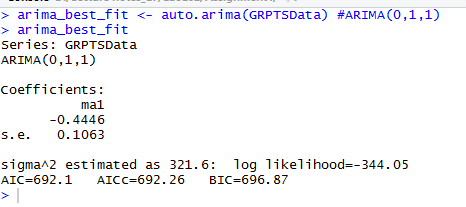
We then do the ADF test. We get the p value > critical value. So, we do the first order differencing and then observe the graph. We also do the ADF test after first order differencing and see that p value < critical value. So, we reject the null hypothesis and conclude the data is stationary after the first differencing.

**r code for ADF test**



|  |  |
| --- | --- |
| **Initial dataset** | **After first differencing** |
|  |  |

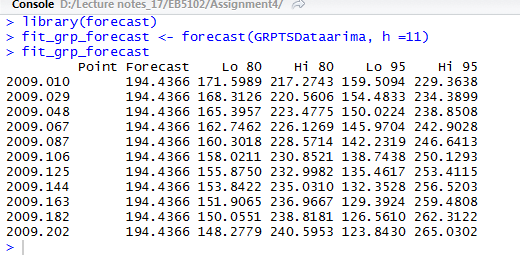
Next, we fit the best model through the r command auto.arima(). We get the best fit at ARIMA (0,1,1).



We check the residuals (ACF and PACF) of the best model ARIMA(0,1,1)

|  |  |
| --- | --- |
| **ACF** | **PACF** |
|  |  |

We see both the ACF and PACF are well within the significant limits. Hence, this further proves that ARIMA(0,1,1) is the best model for this dataset, Next **we forecast the values for the next 11 weeks.**

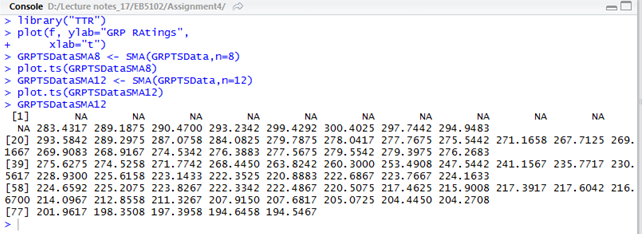


1. **Decomposition methods**: To estimate the trend component of a **non-seasonal time** series that can be described as an additive model we use **simple moving averages.** We take the moving average of order 8 and observe the graph.

|  |  |
| --- | --- |
| **8 Period moving average** | **12 period moving average** |
|  |  |

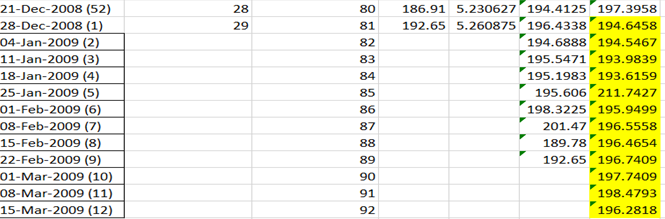
We observe that the in 8 period moving average the graph has not smoothed out fully. So, we try different periods and finally arrive at 12 period moving average where we can see a downward trend.

**r code used for calculating simple moving averages for period 8 and period 12**

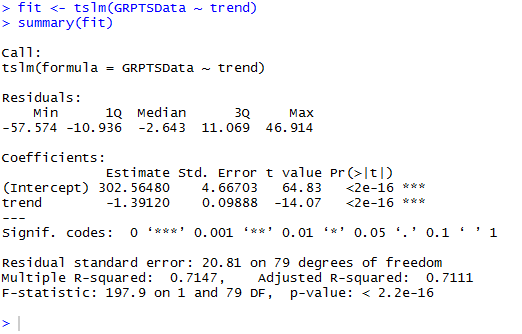


Forecasts for simple moving average was calculated by taking the average of the actual data of previous 12 weeks and in cases where the actual data was not sufficient the forecasted data was added with the actual data of previous weeks and the forecast was made.

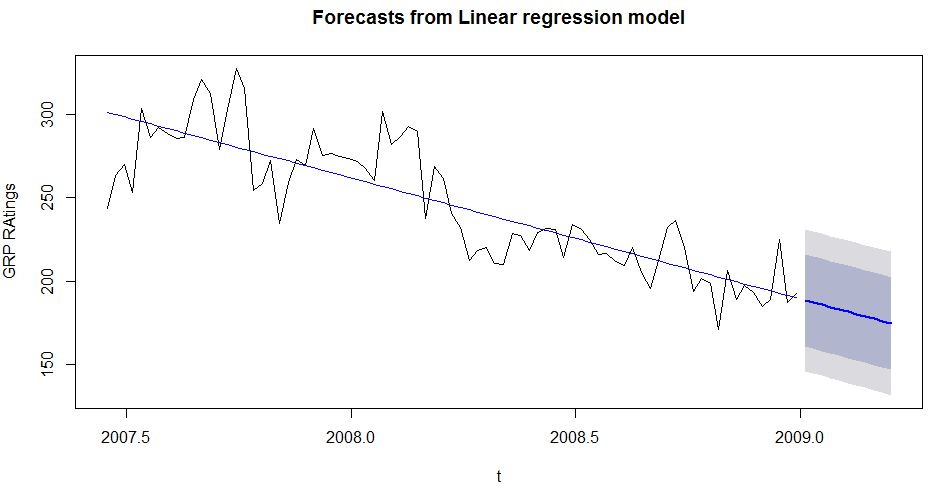
**Forecast using simple moving average for the next 11 weeks**

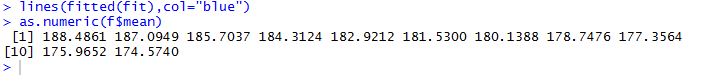


1. **Time series Regression:** We can fit linear regression for the GRP ratings over trend. We can either use “lm(GRP~TimePeriod,data=GRP)” to fit the regression or “tslm(GRPTSData ~ trend)”. Both gives same output.

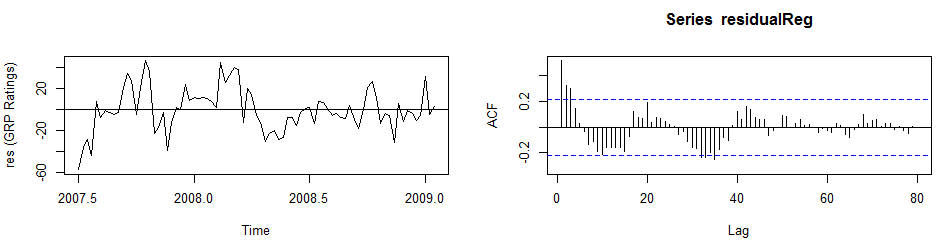


We can then **forecast for next 11 weeks from Jan-15th March**. Below is the plotted value



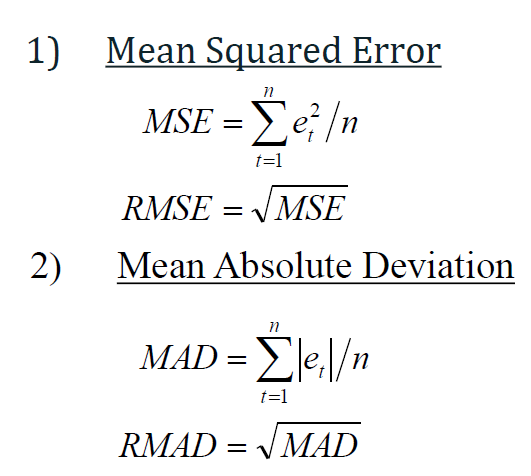


We can plot the residuals that gives insight around lag 3-4 and lag 38 the ACF value exceeds the significant line.

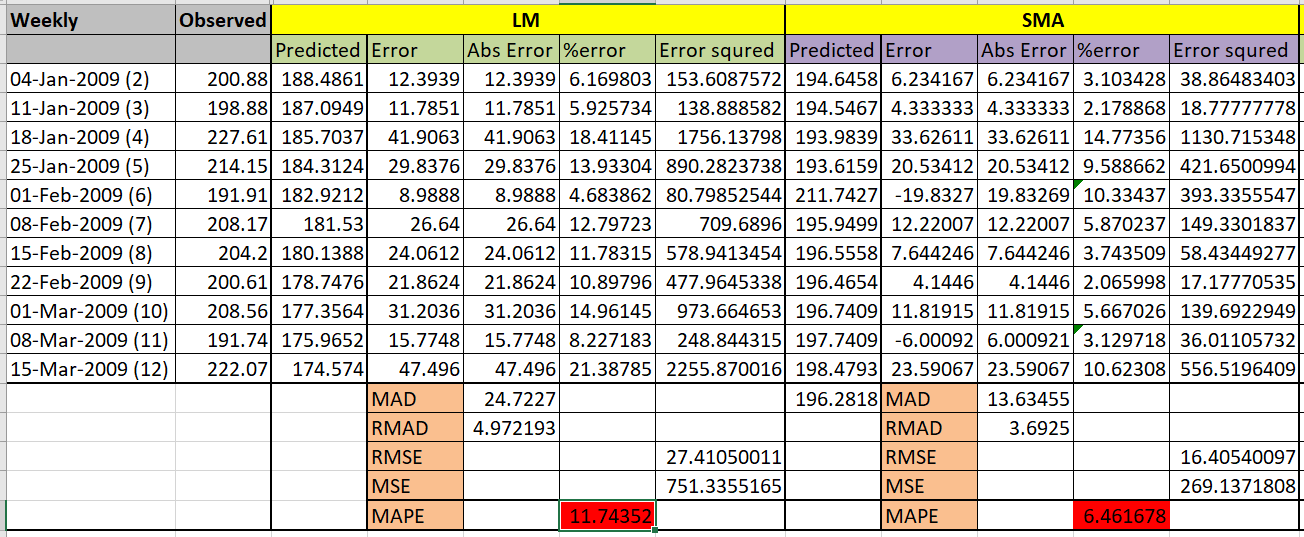


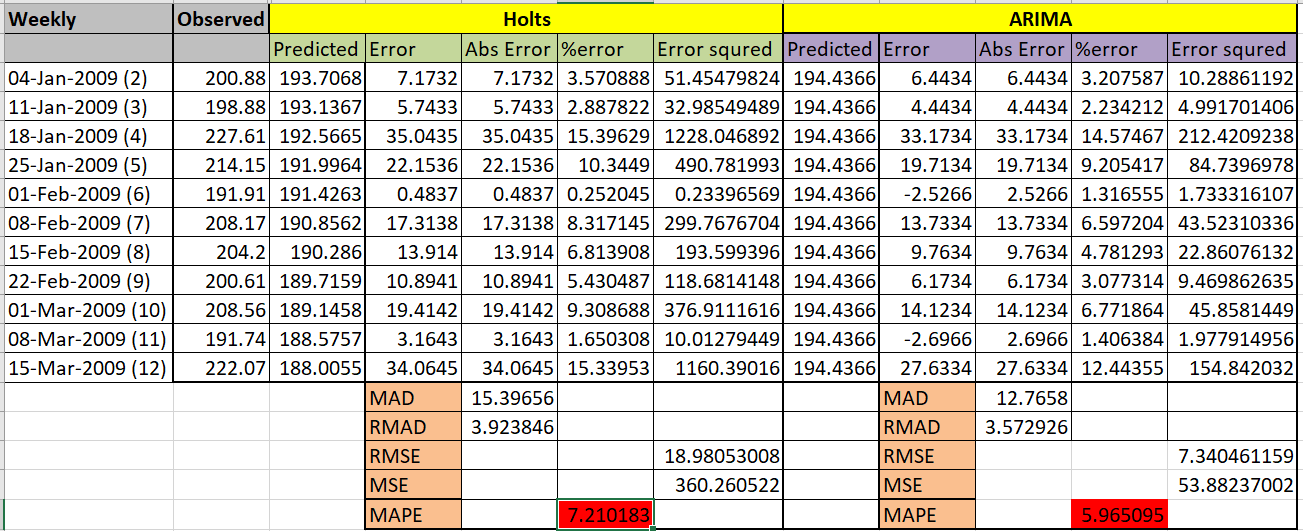
**Error calculation:**

The following errors were calculated for **LM-Time series Regression, SMA- Simple Moving average, Holts- Holt’s exponential smoothing, ARIMA** methods



3) And MAPE-Mean absolute percentage error





**We see that ARIMA model has the least MAPE(5.96%),MSE(53.88),RMSE(7.34),RMAD(3.57),MAD(12.76) among all four techniques used.**

**Conclusion:**

1. On closely examining the graph of our GRP dataset we see that our data doesn’t have any seasonal pattern but there is only a downward plunge. On further doing a test in r for **“Decomposition by BATS model”** we find that the test for seasonality comes as false. Also, on analyzing the ACF graph of ARIMA(0,1,1) we see there are parallel lags at 3,11 and 17 lags, though not at equal intervals to conclude seasonality. So, based on the above points we conclude there is no seasonality and use “Simple Moving averages” to do the forecasting, since we cannot do the traditional decomposition.
2. On applying the different techniques, we found that we were able to do forecasting using all four methods, i.e. Time series regression, Simple moving average, ARIMA and exponential decomposition. However, based on the error calculation like **MAPE, MAD, RMAD, RMSE and MSE,** we came to the conclusion that ARIMA **is the best model.** ARIMA has the least error in all four error calculations (MAPE, MAD, RMAD, RMSE and MSE) and hence is the best model.
3. We find that the MAPE is > 10% for Time series regression which has plotted a downward trend straight line. On observing the data and the regression line closer we find that there is an increase in GRP from July 2007 to September 2007. Compared to the year 2007, in the year 2008 there is a dip in GRP. However, during the months July 2008 there is an increase in ratings and a slow dip in August and an increase in September 2008. The increase in ratings is consistent in September month in both years. This might be due to the Indian festivals like Ganesh Chaturthi celebrated during that month.