Univariate Time Series Analysis and Forecasting GRP of an Indian TV Channel

Team Name: Ensemble

EBAC 4 (Mixed Group), Institute of Systems Science, NUS

[Abhilasha: A0163276B, Muni: A0163382E, Pradeep: A0163453H, Sambit: A0163285B]

# Abstract

*This work is part of assignment to apply different time series forecasting methods a.k.a* ***Exponential Smoothing****,* ***Decomposition Methods****,* ***ARIMA****, and* ***Time Series Regression.*** *We will analyze each method output and compare their performances. Also, based on the nature of time series, we would like to evaluate if one or other methods are not suitable for analysis. Different model assumptions, model performances will be the key indicators to decide upon that.*

# Executive Summary

Our business problem is to forecast the GRP of an Indian TV channel. Proliferation of TV channels started to happen in 90s with socio and economic reforms under that time government of [India [1]](https://en.wikipedia.org/wiki/Television_in_India#Phase_I). Currently, there are total 857 permitted private satellite channels in India and more than 190 government [channels [2]](https://en.wikipedia.org/wiki/Lists_of_television_stations_in_India). Nowadays, Indian TV channels face stiff competition to maintain high rankings. Hence it is critical for them to be able to forecast the accurate GRP to maintain profits. We used the weekly data provided as part of assignment and applied following methods to build and compare performances:

* Exponential Smoothing,
* Decomposition Methods
* ARIMA
* Trend Models (Time Series Regression)

As part of this exercise, we found that, ARIMA(0,1,1) performance was best with **RMSE** 13.98 and **MAE** 9.5 followed by Exponential Smoothing pattern with **RMSE** 17.67 and **MAE** 13.78. Detailed results are show in the model performance later in the document. [[ 3]](#_Model_Performance)

Based on the model performance, we found following methods to be least performer:

1. Decomposition
2. Time Series Regression (Trend Models)

# Data Selection, Partition & Preprocessing

The data provided is a weekly time series of GRP ratings of an Indian TV channel. The data is between **17 June 2007** to **15 March 2009**.

## 3.1 Data Partition

As per the guidance, data partition was performed as below:

|  |  |
| --- | --- |
| Train Set | 17 Jun 2007 – 28 Dec 2008 |
| Forecast Period | 04 Jan 2009 – 15 Mar 2009 |

## 3.2 Tools

We used the following tools for our analysis:

* Excel/R (Smoothing Models, Decomposition, Time Series Regression, Residual analysis),
* Jmp (ARIMA)

## Data Preprocessing

GRP time series plot with trend line is as show below. Clearly, we can see there is a downtrend in the GRP of this Indian channel. The visual inspection indicates it’s a linear kind of downtrend.

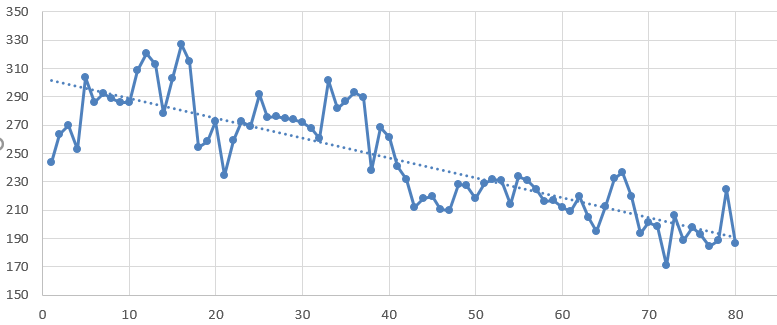
* There was no missing data and the time series was in right order. For the time variable (t), we reset the counter.
* For ARIMA, after analyzing the residual plot with actual GRP variable, we found the high variance in the residuals and hence we took the log and performed the ARIMA again. Taking log of variable helped in model performance.

**Fig (1) GRP Time Series Plot**

# Model Building

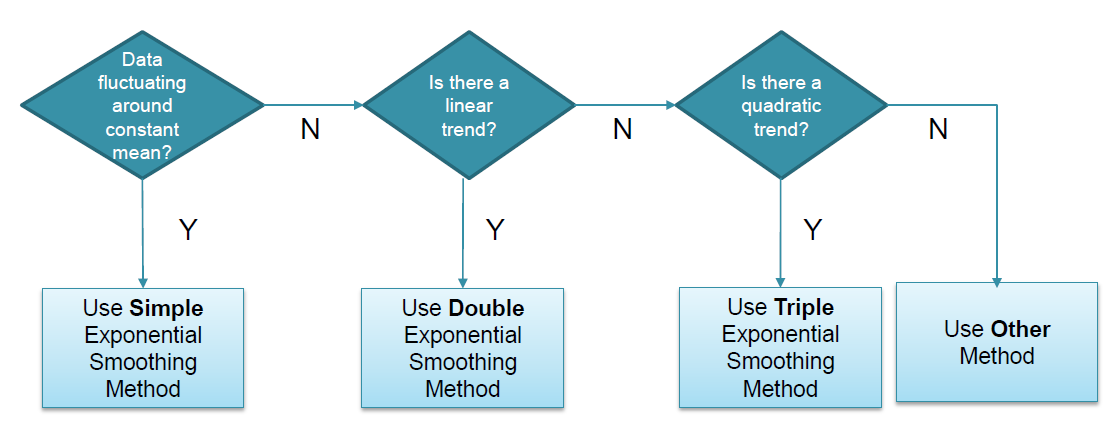
## Exponential Smoothing

The data shows a distinct downward trend as per below chart:



The Moving Average and Simple Exponential Smoothing methods don’t adequately model this, but **Holt’s Linear Trend Method** (aka **Double Exponential Smoothing**) does. This is accomplished by adding a second single exponential smoothing model to capture the trend (either upwards or downwards).

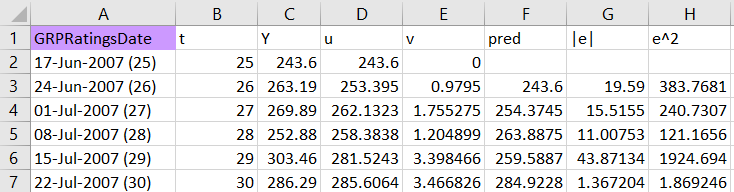
**Fig 2 Exponential Smoothing Deciding chart**



### Formulas in different cells below

* D2 = C2,
* E2 = 0, cell
* D3 = (C$96 \* C3 + (1-C$96) \* (D2+E2)),
* E3 = (D$96 \* (D3-D2) + (1-D$96) \* E2)
* F3 = D2+E2.

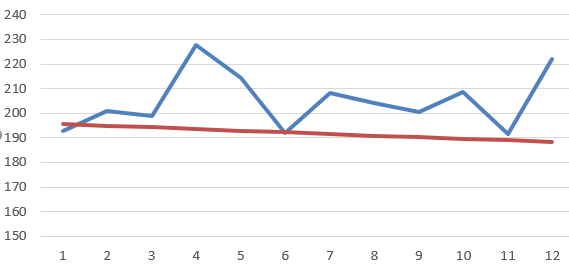
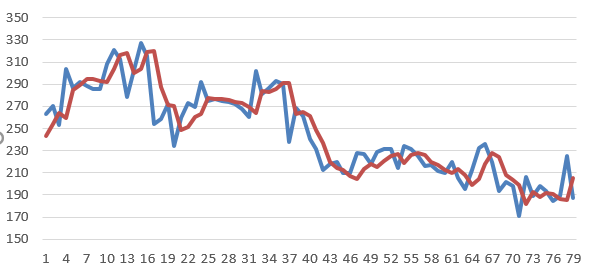
We tried minimizing the value of MAE by changing the value of alpha and beta subject to the constraint that alpha <= 1.0 and beta <= 1.0. The result shown below is at α = .5 and β = .1, with MAE = 13.2. [[4](http://www.real-statistics.com/time-series-analysis/basic-time-series-forecasting/holt-linear-trend/)]



Next step is to forecast the next 12 weeks. The actual Y (blue) and predicted Y(red) values are plotted below. The straight line (in red) on **Fig 4 plot** shows the forecasting using smoothening while the blue curve shows the actual value.

**Note:** This method is appropriate since the data is getting converged as clearly evident from the train data plot.

**Fig 3 Train Data Plot Fig 4 Test Data Plot**



## Decomposition

Decomposition methods are based on an analysis of the individual components of a time series. The strength of each component is estimated separately and then substituted into a model that explains the behavior of the time series. Two of the most important decomposition methods are:

* Multiplicative decomposition
* Additive decomposition

Here based on the visual inspection of the time series data, we found that overall fluctuation is getting reduced and not constant width. Hence we chose the the multiplicative decomposition model which is expressed as the product of the four components of a time series: ***Yt* = *TRt \* SNt \* CLt \* IRt***

These variables are defined as follows:

* *Yt* = Value of the time series at time t
* *TRt* = Trend at time t
* *SNt* = Seasonal component at time t
* *CLt* = Cyclical component at time t
* *IRt* = Irregular component at time t

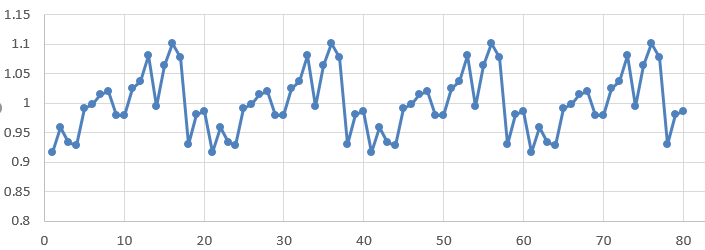
### **Output** **(Overall Trend):**

As shown in Fig 5 below, Time series plot shows overall linear downward trend. However, we used degree 2 and degree 3 trend lines and found linear trend was fitting the best possible way.

**Fig 5 Time Series plot**

#### **Seasonality (SN):**

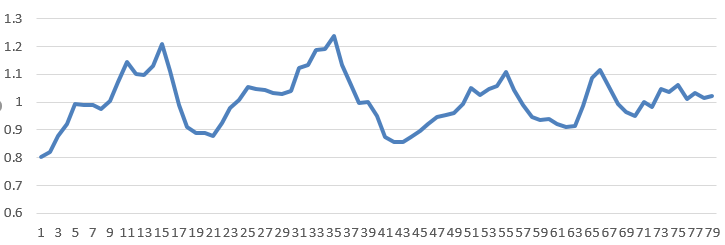
After decomposing the data in to different parameters using Excel. Seasonality chart was found to be as show below **Fig 6**. Near every 20th week, we see a drop in the GRP ratings. It will be worth to investigate the possible reasons for this trend and introduce the appropriate intervention to maintain the GRP rating for a healthy business.



#### **Cycle (CL):**

Cyclic plot shown below is similar to seasonality. However to understand the cyclic pattern, we would need at least 3 years of data to understand patterns.

**Fig 7. Decomposed Cyclic plot**



#### **Irregularity (IR):**

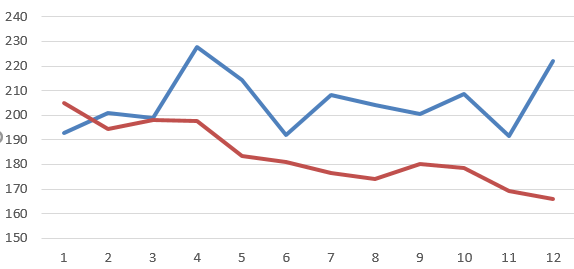
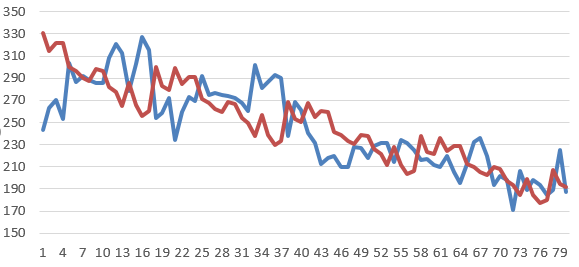
Irregularity was found to be constant around mean value of 1.

#### **Forecast:**

Next, we forecasted the last 12 values in the times series.

The actual Y (blue) and predicted Y(red) values are plotted below. Based on model performance parameters like MAE, RMSE, this model wasn’t found to be performing well and hence not recommended for this data forecast.

**Fig 8 Train Data Plot** **Fig 9 Test Data Plot**



## Time Series Regression

Time Series Regression is based on an analysis of the trend (TR) components of a time series. The strength of this component is estimated as per below:

*Yt = TRt + ԑt where,*

*Yt = value of the time series in period t*

*TRt = the trend in time period t*

*ԑt = the error term in time period t*

*No Trend: TRt = β0*

*Linear Trend: TRt = β0+ β1t*

*Quadratic Trend: TRt = β0+ β1t + β2t2*

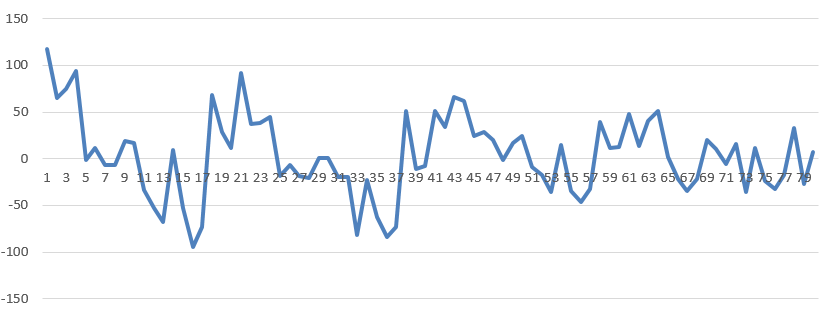
As we have already observed in the time series plot that there is a linear downward trend in the data.

### **Process**

When we performed the linear regression as per the equation above. Following coefficient and constant was found: *TRt = 313.336 – 1.594t*

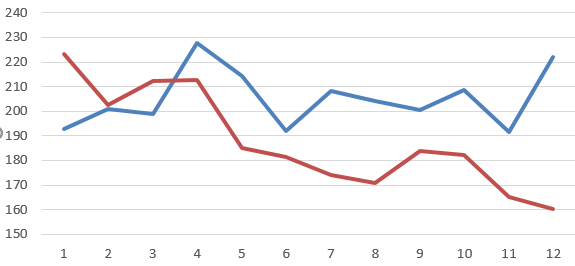
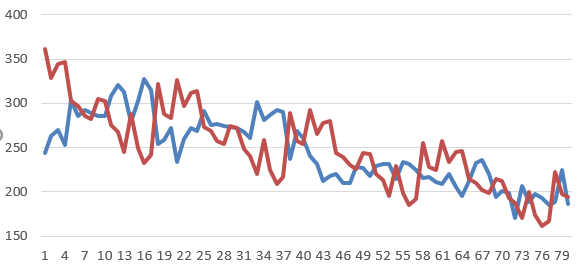
Please find below the ԑt (Error) plot:

**Fig 10 Error plot**



As we can see that error term is reducing over time. It shows that trend variable could be used to forecast provided we have large training set. Again, we evaluated the mode performance based on RMSE, MAE. It was found to be comparatively low as to other models.

**Fig 11 Train Data Plot** **Fig 12 Test Data Plot**



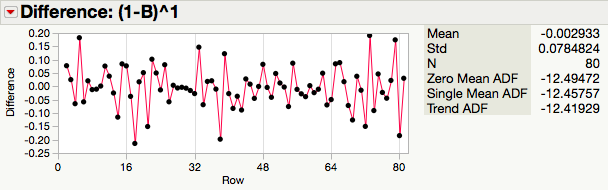
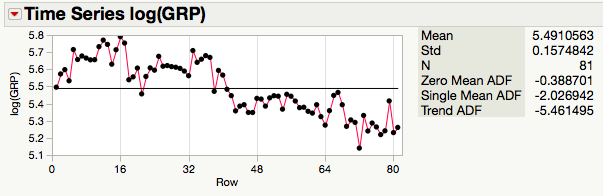
## ARIMA

Since ARIMA models help in forecasting nonstationary series and provide better formulation for incremental rather than structural changes, its well suited for the given data.

### Steps Followed:

* The given data was plotted and observed to be non-stationary with a downward trend
* Unit root test (Augmented Dickey fuller) < -4, further suggests the data to be non-stationary and hence needs to be differenced to make it stationary which is an assumption to run Box & Jenkins method.
* Differencing order 1 was found to be sufficient to make data stationary.

**Fig 13. Time Series Plot** **Fig 14. Difference Order 1 Plot**



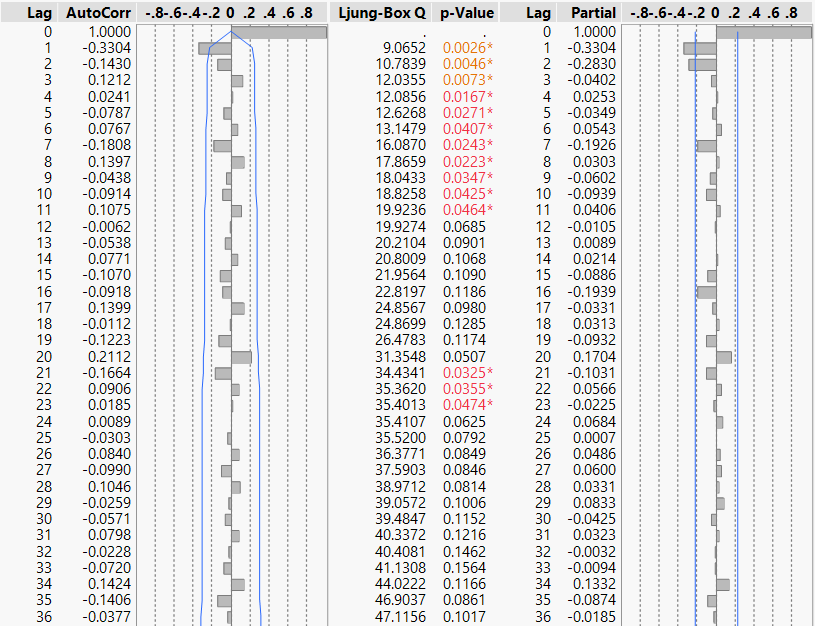
* The parameters autoregressive and moving‐average portions of the ARIMA

model, the p, d and q terms, where, p refers to the AR part, q refers to

MA part and d is the order of differencing were determined with the help of **ACF and PACF** plots as shown below **Fig 15**. The Plots suggested the model to be of AR (2), MA (1) , with no significant seasonality observed given the length of data.

However, to build a better fit model, other models were also tested by running ***“ARIMA model group”*** option in Jmp. The different model performance is compared in the model performance section below.[[5]](#_Model_Performance)

**Fig 15 ACF and PACF Plot**

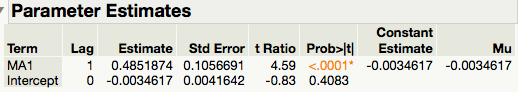


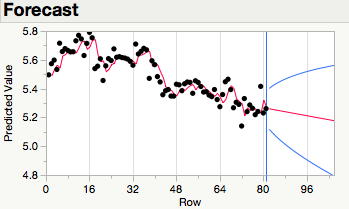
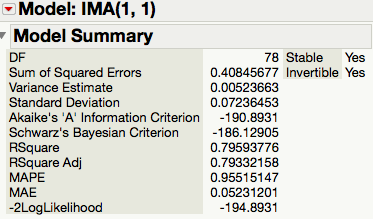
* After approximation of ARIMA model, estimated model parameters were

checked to be significant and Ljung Box test was observed to confirm that the white noise was achieved, Model Residuals were checked to achieve white noise (constant mean and constant variance),

* Upon analyzing the residuals, it was found that the variance is not uniform, so data was transformed with the help of natural log and steps were redone from the beginning and again model group was run which also suggested the similar result. However, in terms of performance model with log(GRP) was found to be performing better. Fig 16 below shows the Parameter estimates and model summary for IMA (1,1)

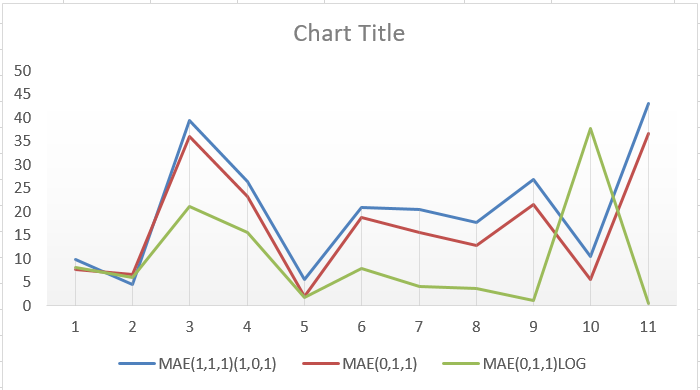
**Fig 16 Parameter Estimates/Model Summary**



******

* Residual plot was done for test set for different ARIMA models we ran. Below charts indicates the performances. IMA (0,1,1) after taking log(GRP) variable shows the lowest absolute error.

**Fig 4. *Residual plot for ARIMA Models***



# Model Performance

Below is shown the different method performances and comparison for our Test Data (04 Jan 2009 – 15 Mar 2009)

**Table 1**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | RMSE | MAE | MAD | RMAD |
| Exponential Smoothing | 17.67 | 13.78 | 9.24 | 3.04 |
| ARIMA(0,1,1) log | 13.98 | 9.50 | 9.0 | 3.0 |
| ARIMA(0,1,1) | 20.32 | 16.94 | 16.9582 | 4.11 |
| SARIMA (1,1,1)(1,0,1)4 | 23.86 | 20.47 | 9.08 | 3.01 |
| Decomposition | 27.41 | 23.47 | 11.24 | 3.35 |
| Time Series Regression | 29.01 | 24.92 | 11.05 | 3.32 |