Time Series Forecasting

Assignment

By,

Team Uplift

Balagopal Unnikrishnan (A0178398E)

Maddi Kamal Manisha (A0178259M)

Nivedita Valluru Lakshmi (A0178253Y)

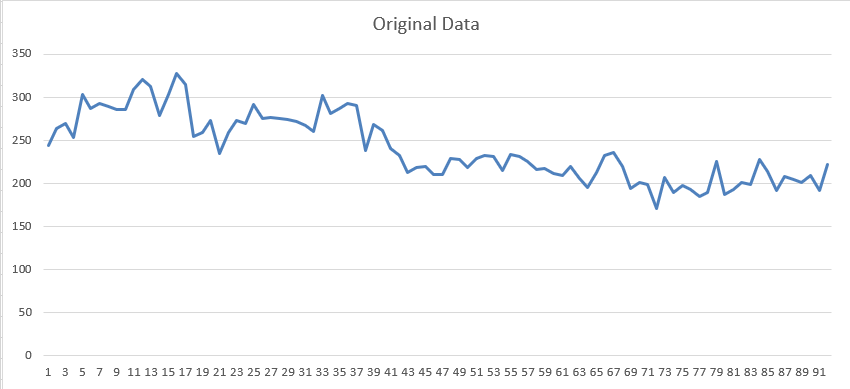
Savita Rani Ravichandran (A0178327W)

Pedapudi Venkata Sai Vijay Kumar (A0178190Y)

**Introduction**

The main objective of this assignment is to perform time series analysis on an Indian TV channel rating dataset containing weekly stats of GRP(Gross rating point) from 17-Jun-2007 till 15-March-2009. We have considered data till October end 2008 as our training set and the rest as testing set. We have tried several models ranging from simple time series regression to exponential smoothing to decomposition to ARIMA models for forecasting the target values.

**Data**

The given data was plotted to check for its general properties like trend and seasonality. As seen in the graph below, the data shows a downward trend. Since the data represents tv channel ratings, it is recommended to analyse the kind of channel we are dealing with in order to make better predictions.

On analyzing the peaks and troughs in the graph, it seems like the channel under study could be a sports channel. On days that the GRP crossed 300, there was a cricket match in which the Indian team participated. Particularly the peaks in points 12,13,16 referring to the first weeks of September and the last week of September coincide with the inaugural ICCTwenty20 WorldCup in which India won the tournament. The final match was against Pakistan and owing to Indias rivalry against Pakistan, the match was viewed by almost every household in India and this could coincided with the highest point in the graph. We also observe that the points from 38(March-2008) are showing a decreasing trend. On checking the Indian cricket team schedule, we observe that there wasn’t much activity between March-June and post that Indian team took part in Asian cup and visited Sri Lanka and won the ODI series and this corresponded to the small high region in the latter part of the graph. From this we understand that the graphs show highs and lows based on Indian cricket team schedules which donot follow any seasonality as such. The only matches which would surely be watched are the World Cups which occur once in 4 years.

**Time series Regression**

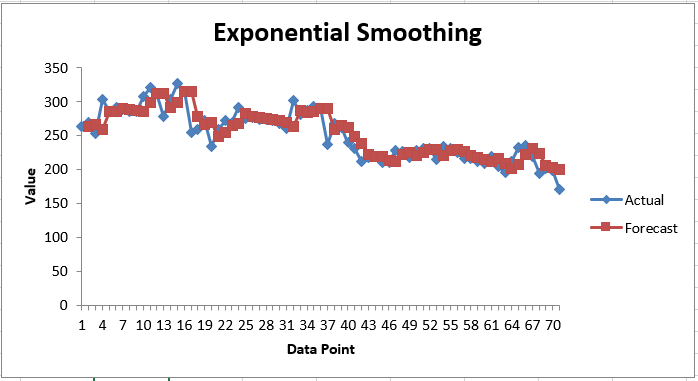
**Exponential Smoothing**

Smoothing involves some form of local averaging of data such that the nonsystematic components of individual observations cancel each other out. Moving averages are the simplest form of smoothing. For a time series that can be described using an additive model with constant level and no seasonality, simple exponential smoothing can be used to make short-term forecasts. For a time series that can be described using an additive model with increasing or decreasing trend and no seasonality, Holt’s exponential smoothing can be used to make short-term forecasts.

**Simple Exponential Smoothing**

The simple exponential smoothing method provides a way of estimating the level at the current time point. Smoothing is controlled by the parameter alpha; for the estimate of the level at the current time point. We tried different values for alpha on the entire dataset and their results are shown below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| alpha | 0.30 | 0.35 | 0.4 | 0.5 | 0.6 |
| RMSE | 17.70063 | 17.51168 | 17.511684 | 17.49591 | 17.731407 |
| Sum of standard error | 1344.263 | 1330.524 | 1330.524 | 1332.23 | 1354.00741 |

We decided to build our final model using alpha=0.5 since it gave good results for both the RMSE and Sum of standard error. Logically since our data represents tv ratings of a sports channel it is only right that more weightage be given to the previous week since a cricket series is bound to last a few weeks or sometimes even months. Hence the model built needs to give more weightage to the previous week as compared to the other weeks.

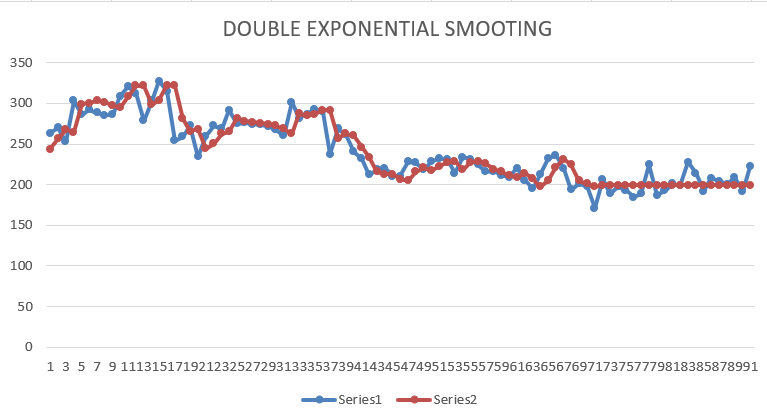
The final model was built on the training set and the results were computed on the test set. The graph above shows the fitting line on the training set. The metrics evaluated on the test set are shown below.

|  |  |  |  |
| --- | --- | --- | --- |
| MAD | MAPE | MSE | RMSE |
| 10.51424 | 5.191431 | 156.596 | 12.51 |

**Holts exponential smoothing**

Holt’s exponential smoothing estimates the level and slope at the current time point. Smoothing is controlled by two parameters, alpha, for the estimate of the level at the current time point, and beta for the estimate of the slope b of the trend component at the current time point. The model has been tested with alpha and beta values ranging from 0.1 to 0.9 and the table below displays the SSE values for the entire data.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **beta alpha** | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
| 0.1 | 51405.64 | 49623.95 | 52385.33 | 58385.11 | 65204.49 | 70593.37 | 73981.95 | 75353.18 | 74219.58 |
| 0.2 | 38236.36 | 39767.31 | 41633.14 | 42340.13 | 42008.77 | 41306.77 | 40942.58 | 41182.28 | 41840.44 |
| 0.3 | 32575.18 | 33390.51 | 33916.13 | 34168.57 | 34481.84 | 34960.34 | 35538.19 | 36133.41 | 36668.51 |
| 0.4 | 29638.22 | 30259.41 | 30792.73 | 31315.9 | 31887.39 | 32478.48 | 33049.44 | 33586.48 | 34115.47 |
| 0.5 | 28205.49 | 28898.27 | 29597.07 | 30328.67 | 31107.9 | 31938.55 | 32844.49 | 33870.25 | 35065.8 |
| 0.6 | 27688.58 | 28561.52 | 29500.5 | 30526.02 | 31665.96 | 32952.02 | 34421.97 | 36108.48 | 38029.41 |
| 0.7 | 27787.72 | 28911.13 | 30160.64 | 31562.57 | 33151.14 | 34956.54 | 37000.09 | 39288.87 | 41816.08 |
| 0.8 | 28335.64 | 29760.01 | 31369.08 | 33190.64 | 35253.67 | 37576.07 | 40163.09 | 43008.87 | 46101.35 |
| 0.9 | 29240.97 | 31007.55 | 33014.95 | 35289.1 | 37852.24 | 40715.21 | 43880.53 | 47348.23 | 51122.95 |

From the table, it is apparent that the best results are obtained for alpha=0.6 and beta=0.1. The fitting line for these values is displayed below. 

The metrics associated with this model(test data) are displayed below.

|  |  |  |  |
| --- | --- | --- | --- |
| MAD | MAPE | MSE | RMSE |
| 10.9197145 | 5.428803 | 186.6712 | 13.66277 |

It can be seen that the simple exponential smoothing produced better results as compared to holtz double exponential smoothing.

ARIMA

Stationarity

ACF, PACF

ARIMA(p,d,q) model

Decomposition

Why additive