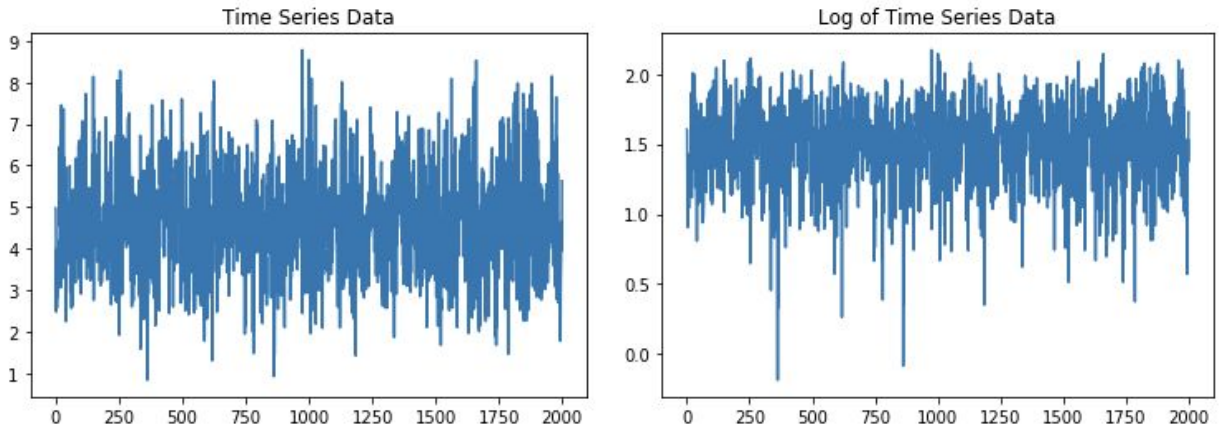
**CSC 591 IOT Analytics**
**Project 3 Forecasting**
**Name: Vidhisha Jaswani**
**Unity ID: vjaswan**

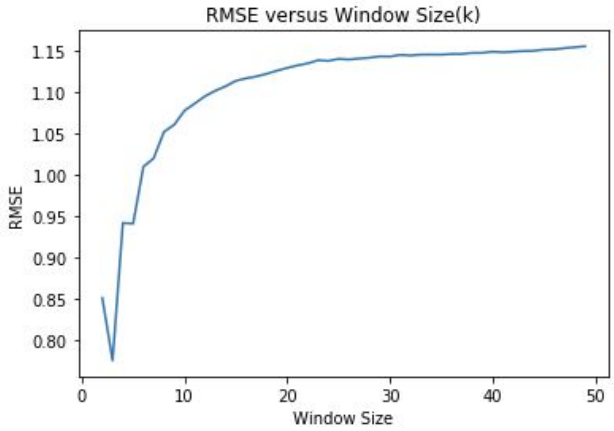
## Task 1. Check for stationarity

If we plot the time series data, the following plot is obtained.

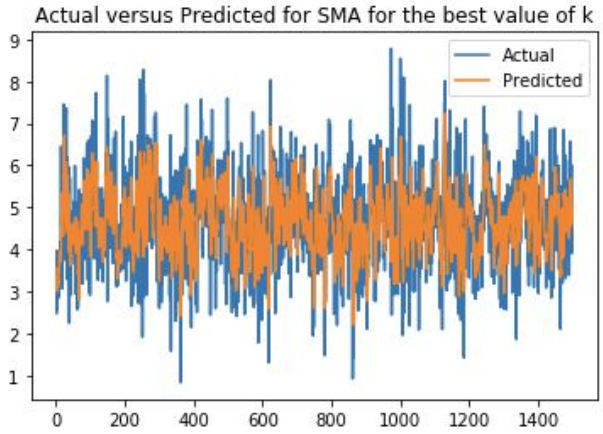


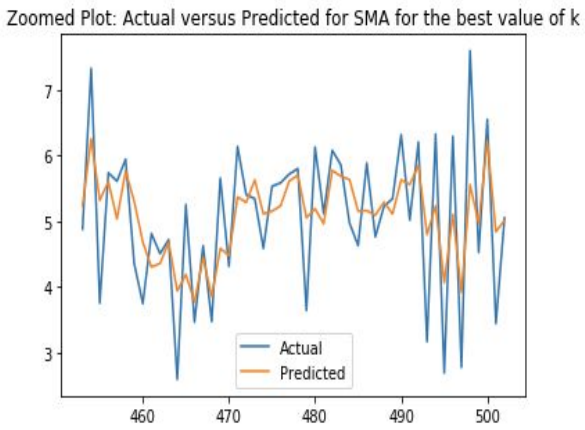

The plot indicates that the data is stationary since the mean seems to be constant throughout, there is not much variance or seasonality. If we take log of each data point, a similar plot is obtained.

## Task 2. Fit a simple moving average model (use the training set)



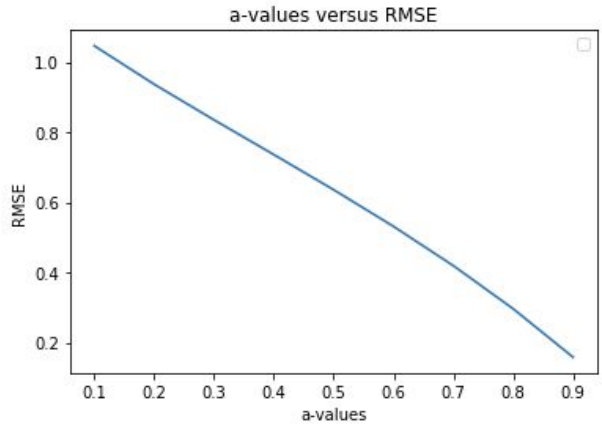

Above is the plot for RMSE versus Window Size (k). Also, Minimum RMSE is 0.775 and is obtained at k= 3 which is also visible from the graph.



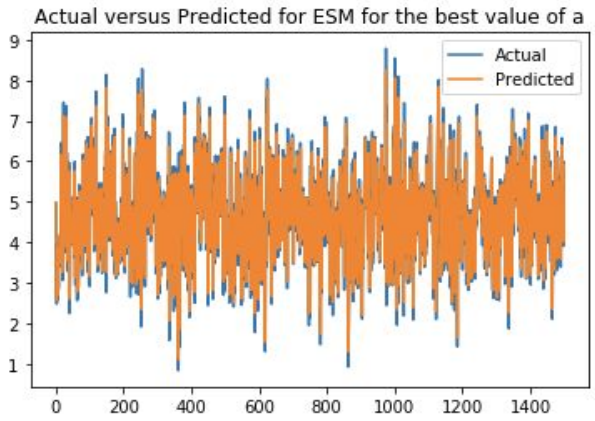



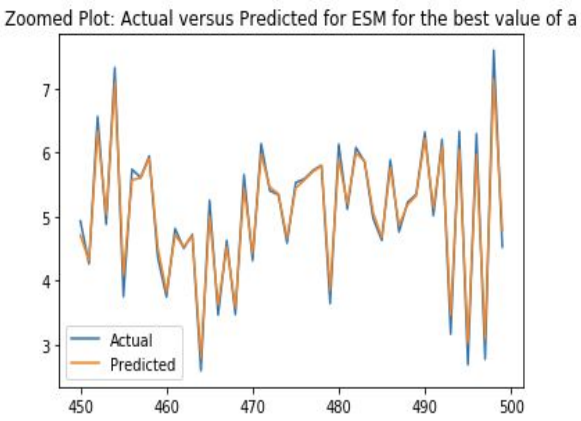

The above shows the actual and predicted values. It shows that the predicted values from the simple moving average model try to follow the actual values but there is still error. We can also make comparison when we obtain the RMSE for all three models.

**Task 3. Fit an exponential smoothing model (use the training set)**




Above is the plot for a-values between 0.1 and 0.9 versus RMSE. It is observed that minimum RMSE is 0.158 at a=0.9 which is also evident from the plot.

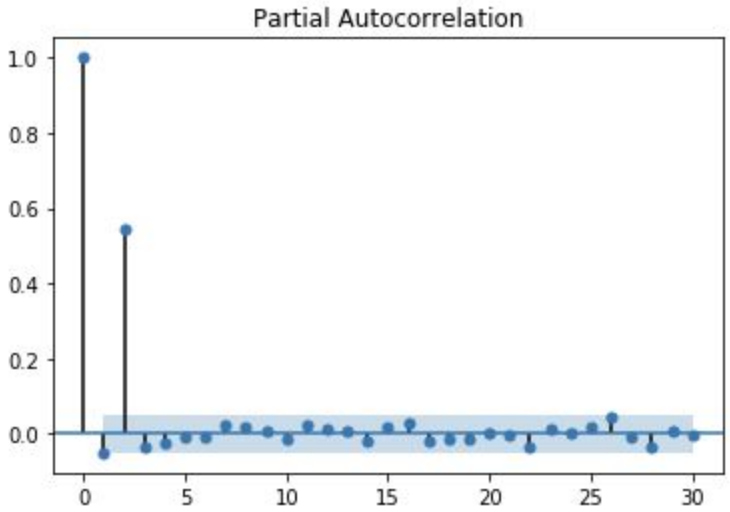


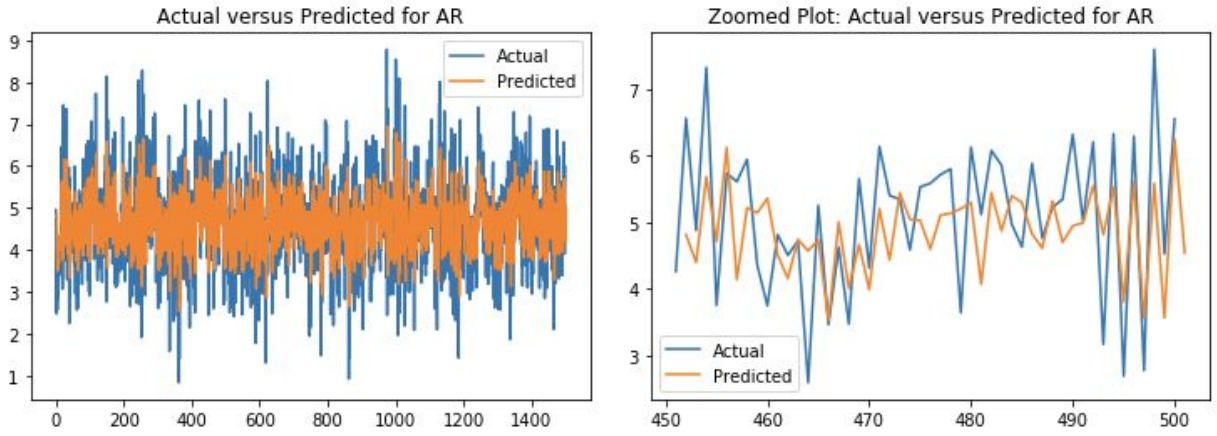

The above are plots for actual versus predicted values for the exponential smoothing model for the best value of a i.e 0.9. This model does a better prediction than the simple average model as the accuracy seems higher here. Also the RMSE for Exponential model is much lower than RMSE for the Simple Moving Average Model.

## Task 4. Fit an AR(p) model (use the training set)

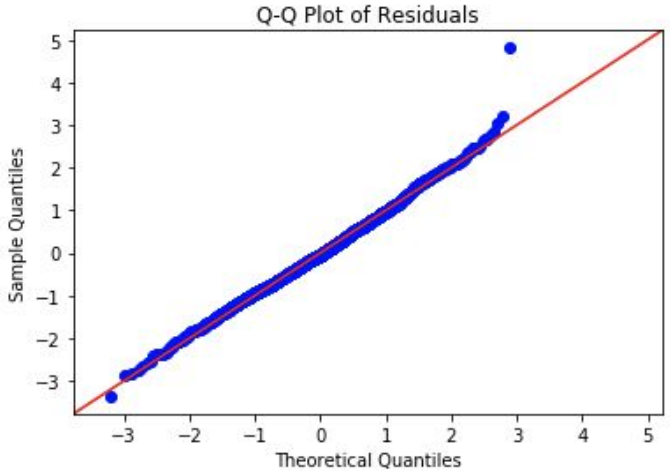
For selecting the order p for the AR model, the PACF plot is obtained as below.



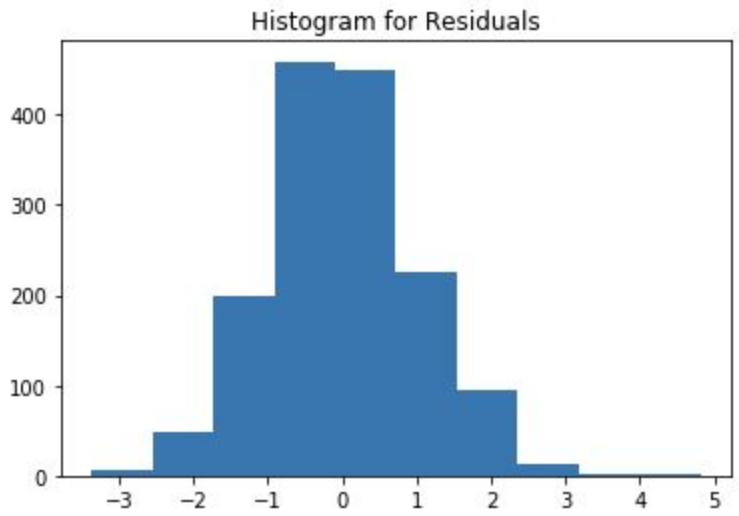

It is evident from the plot that the order is 2 since after that it drops down below 0. The RMSE for the AR model is 1.38. The predicted versus actual values graph is as below.



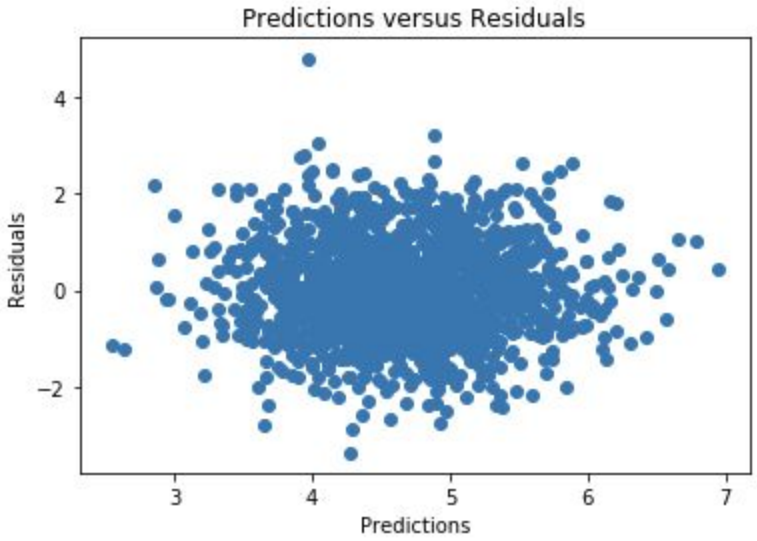

It is visible from the plots that the predicted values tries to follow the actual values however it is not as accurate as the above two models. Even the RMSE is the highest for this model.




The Q-Q Plot for the residuals of the AR model is as below. This plot indicates that the residuals follow a normal curve. The slight bend at the end could indicate that there is a lot of data towards the end. The normal test can be verified using histogram and chi-square test.




The histogram plot for the residuals does indicate that the residuals do indeed follow the normal curve. The chi-square test gave the p-value=1 and since it is greater than 0.1 we accept the null hypothesis that it follows the normal distribution.




The scatter plot indicates there is no trend between residuals and the predicted values.

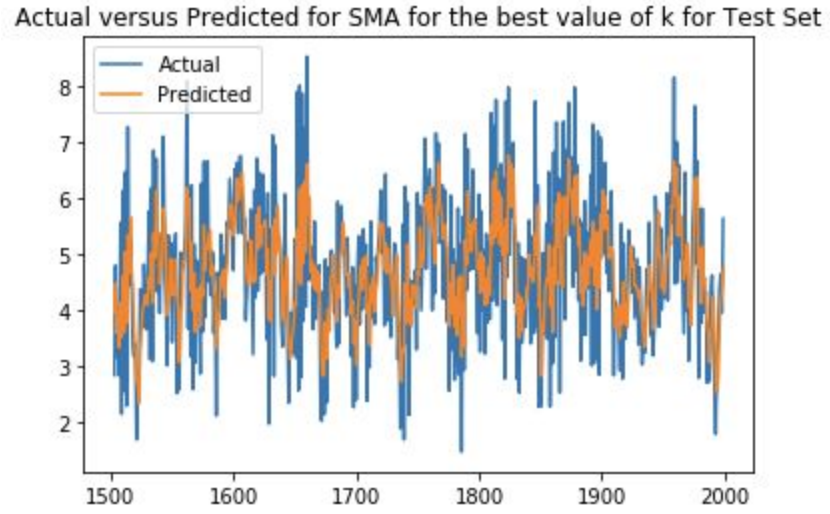
**Task 5. Comparison of all the models (use the testing set)**

The least RMSE in

- Simple Moving Average was for window size k=3 [RMSE: 0.775]
- Exponential Smoothing Model was for a=0.9 [RMSE: 0.158]
- Order 2 for AR Model [RMSE: 1.3877]

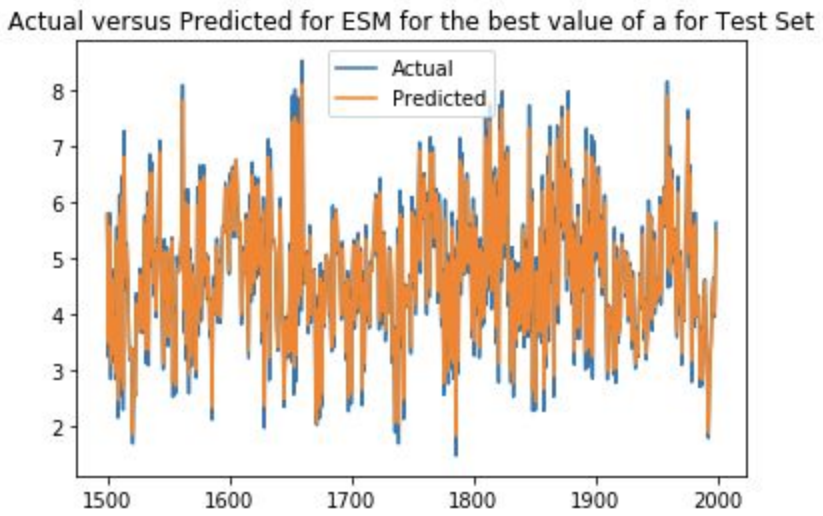
So we try these values for the test data.

1. Simple Moving Average for k=3
   RMSE: 0.887



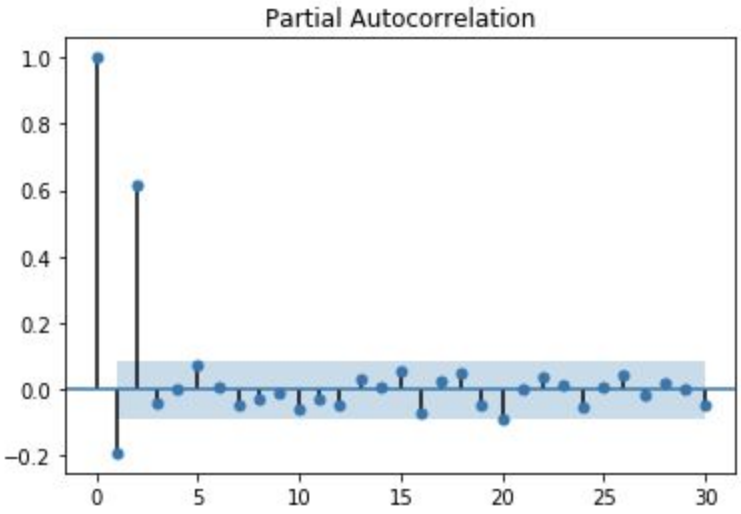

   Plot for actual versus predicted values looks good as it roughly follows the pattern but has errors.

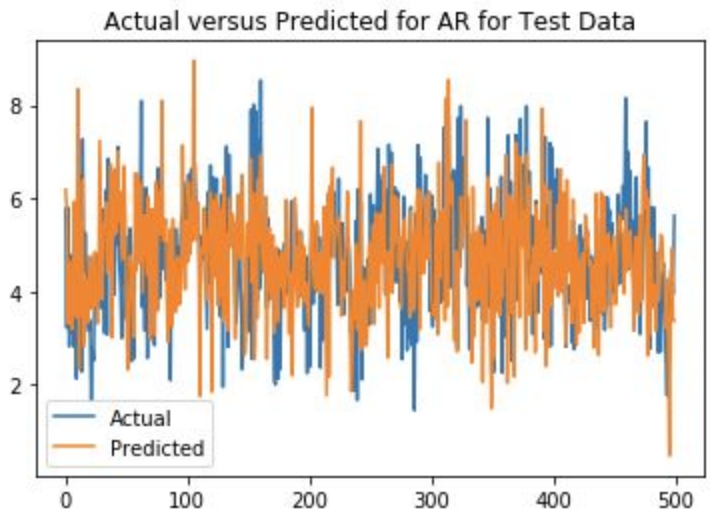
2. Exponential Smoothing Model for a=0.9
   RMSE: 0.1930




This plot looks much better as compared to the Simple Moving Average Model. Even the RMSE is lower.

3. AR Model for Order 2
   Rmse: 1.214






The predicted values tend to follow the actual values but it is not accurate enough. The RMSE is the highest for this model.

**Summary**

Below states the RMSE values obtains on both train and test data sets.

| | **Train Data** | **Test Data** |
|---|---|---|
| **Simple Moving Average for k=3** | 0.775 | 0.887 |
| **Exponential Smoothing Model for a=0.9** | 0.158 | 0.1930 |
| **AR(2)** | 1.3877 | 1.214 |

The above data indicates the best suitable model for our data is the exponential smoothing model at a=0.9. This is also visible from the plot of actual versus predicted values which gives the most accurate fit.