Auto Regressive Modelling of Stocks

* **Santhosh Dasari**

# Why is Time Series special?

Time Series is a collection of data points collected at **constant time intervals**. These are analyzed to determine the long-term trend so as to forecast the future or perform some other form of analysis. But what makes a Time Series different from say a regular regression problem?

There are 2 things:

1. It is time dependent. So, the basic assumption of a linear regression model that the observations are independent doesn’t hold in this case.
2. Along with an increasing or decreasing trend, most TS have some form of **seasonality trends**, i.e. variations specific to a particular time frame.

Because of the inherent properties of a Time Series, there are various steps involved in analyzing it.

First we should check for stationarity. Only then can we apply the mathematical models on a time series.

For practical purposes, we can assume the series to be stationary if it has constant statistical properties over time, i.e. the following:

1. Constant mean
2. Constant variance
3. Autocovariance that does not depend on time.

We can check stationarity using the following:

1. **Plotting Rolling Statistics:** We can plot the moving average or moving variance and see if it varies with time.
2. **Dickey-Fuller Test:** This is one of the statistical tests for checking stationarity. Here the null hypothesis is that the TS is non-stationary. The test results comprise of a **Test Statistic** and some **Critical Values** for difference confidence levels. If the ‘Test Statistic’ is less than the ‘Critical Value’, we can reject the null hypothesis and say that the series is stationary.

# How to make a Time Series Stationary?

Though stationarity assumption is taken in many TS models, almost none of practical time series are stationary. Actually, it’s almost impossible to make a series perfectly stationary, but we try to take it as close as possible.

There are 2 major reasons behind non-stationarity of a TS:

1. **Trend** – varying mean over time.
2. **Seasonality** – variations at specific time-frames.

The underlying principle is to model or estimate the trend and seasonality in the series and remove those from the series to get a stationary series. Then statistical forecasting techniques can be implemented on this series. The final step would be to convert the forecasted values into the original scale by applying trend and seasonality constraints back.

## Estimating & Eliminating Trend:

One of the first tricks to reduce trend can be **transformation**.  We can apply transformation which penalize higher values more than smaller values. These can be taking a log, square root, cube root, etc. But even then, there are trends that still have to be considered.

We can use some other techniques to estimate or model this trend and then remove it from the series. There can be many ways of doing it and some of most commonly used are:

1. **Aggregation** – taking average for a time period like monthly/weekly averages
2. **Smoothing** – taking rolling averages

Smoothing refers to taking rolling estimates, i.e. considering the past few instances.

### Moving average: We take average of ‘k’ consecutive values depending on the frequency of time series.

Now, we remove this moving average or rolling mean from the original time series to get a stationary time series. But we have to remove the first few NaN values as they are used to start the averaging.

Now we test this with stationarity tests like dickey fuller test and check the stationarity. If this doesn’t work, we can use exponential weighted mean average (EWMA) which may give a better result which in our modelling has been proved to be better and more stationary.

We can also use **Differencing** and **Decomposition** techniques also but they are not considered here as they are not giving better results in the end. Decomposition is not much preferred as it is difficult to get back the original values from the stationary time series.

# Using AR Model to fit the Time Series?

After getting a stationary time series all we need to do is to fit an AR Model and predict the future using that model. But before we do that we have to get the parameters that are needed to fit an AR Model

i.e. Partial Auto Correlation Coefficient. (p)

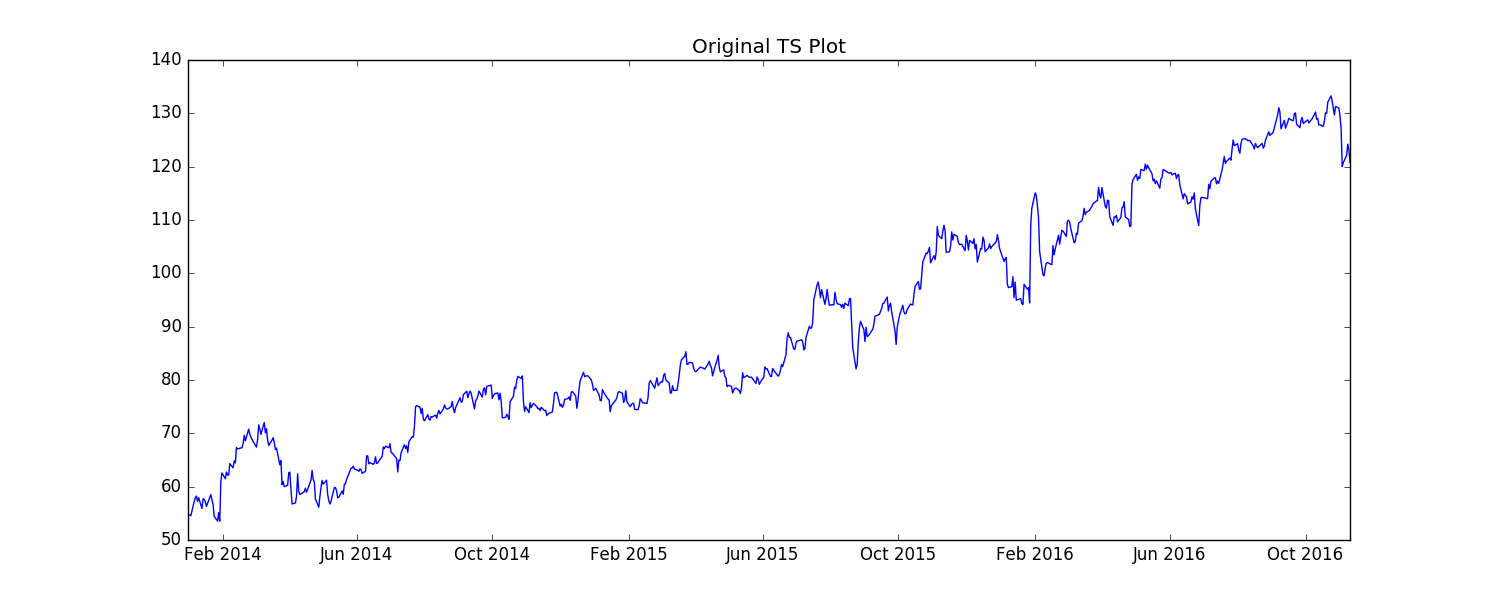
We do this by plotting PACF graphs across lags and get the perfect lag that can be used to fit the model.

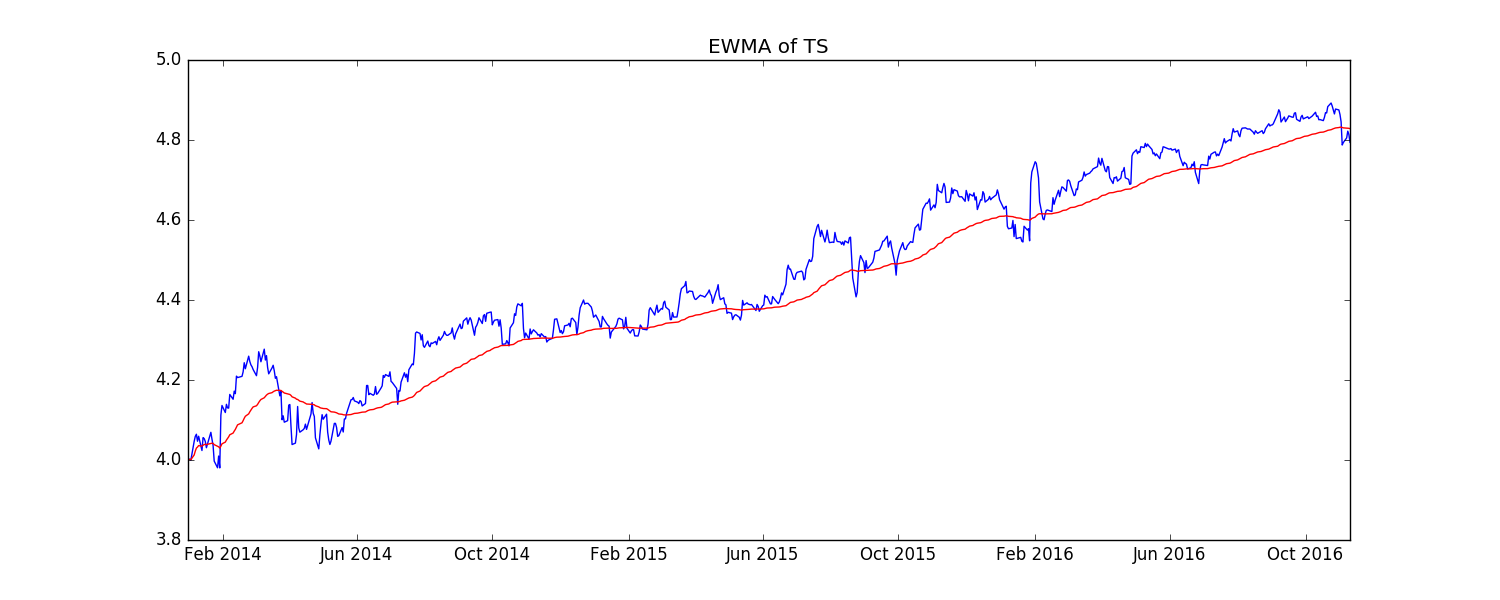
Now we can use this ‘p’ to build a model and use that model to predict future values. Then we have to add back the trend and seasonality and remove logarithm to get the values at original scale.

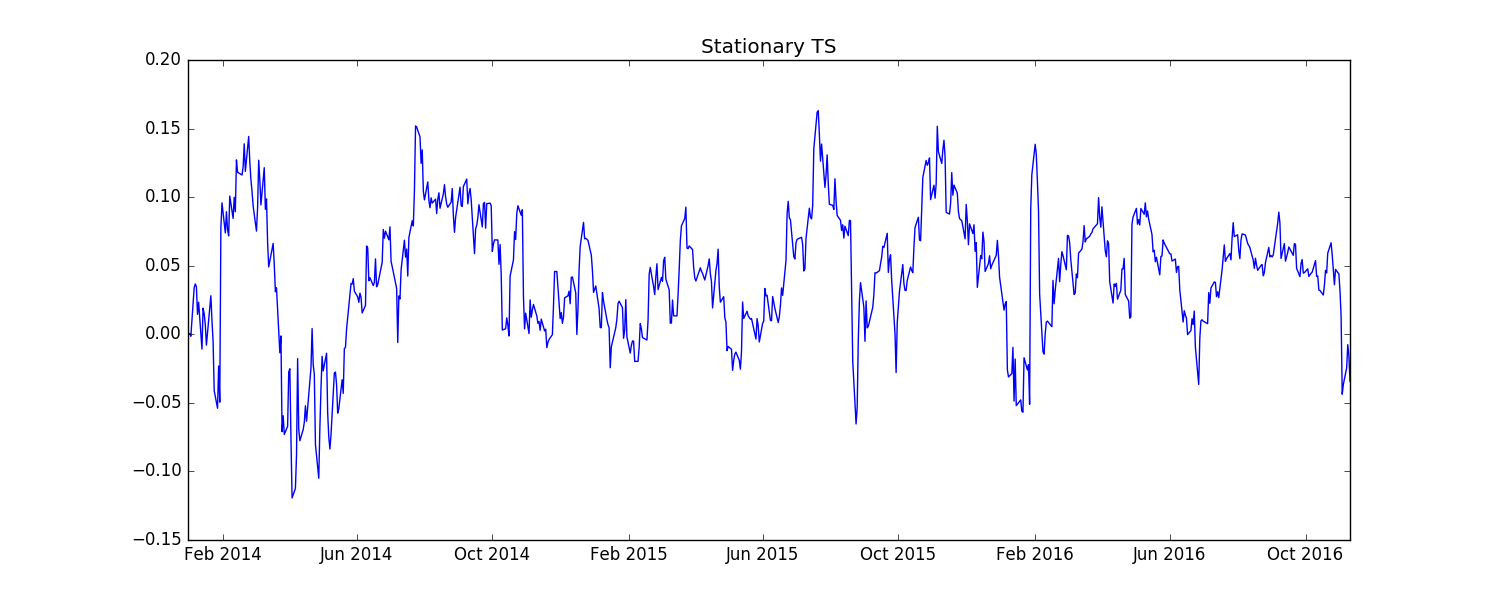
**WE USE PYTHON FOR OUR MODELLING WHICH CONTAINS ALL REQUIRED TOOLS FOR DOING THE MODELLING.**

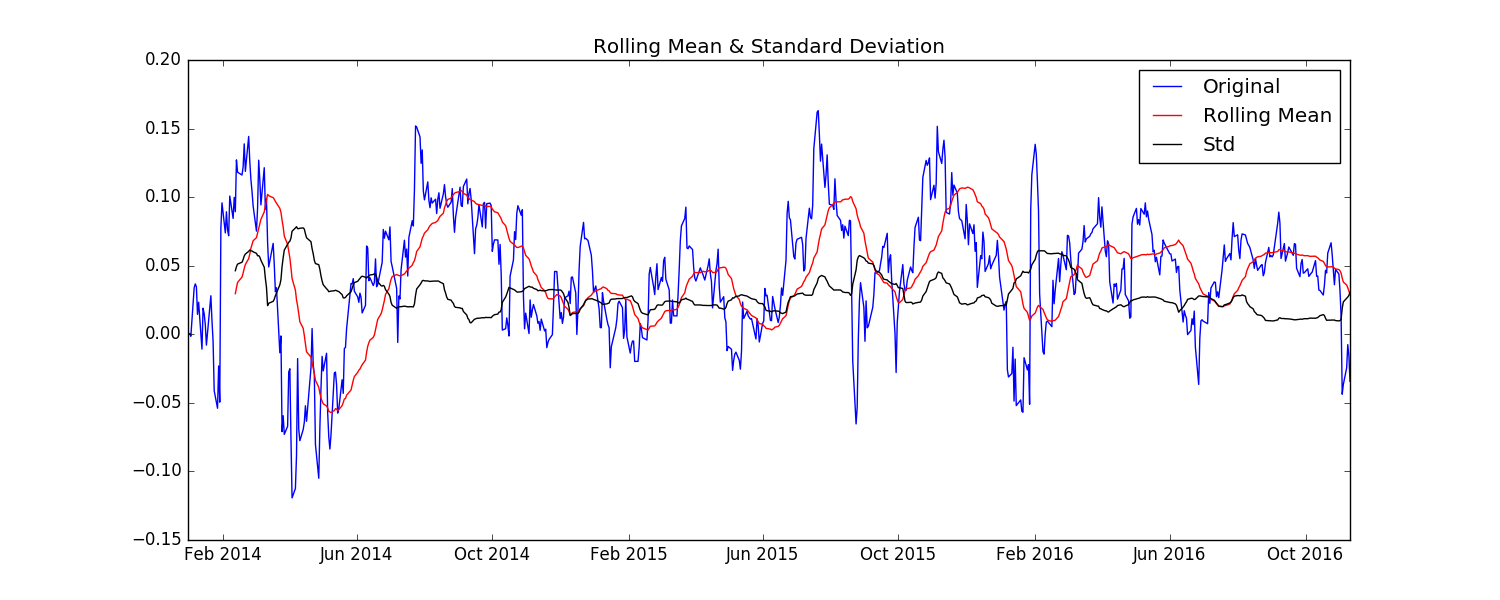
**Modelling for Facebook Stocks Data**:

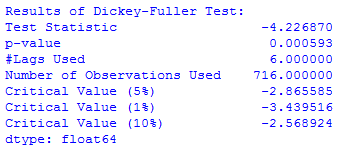
**FACEBOOK**

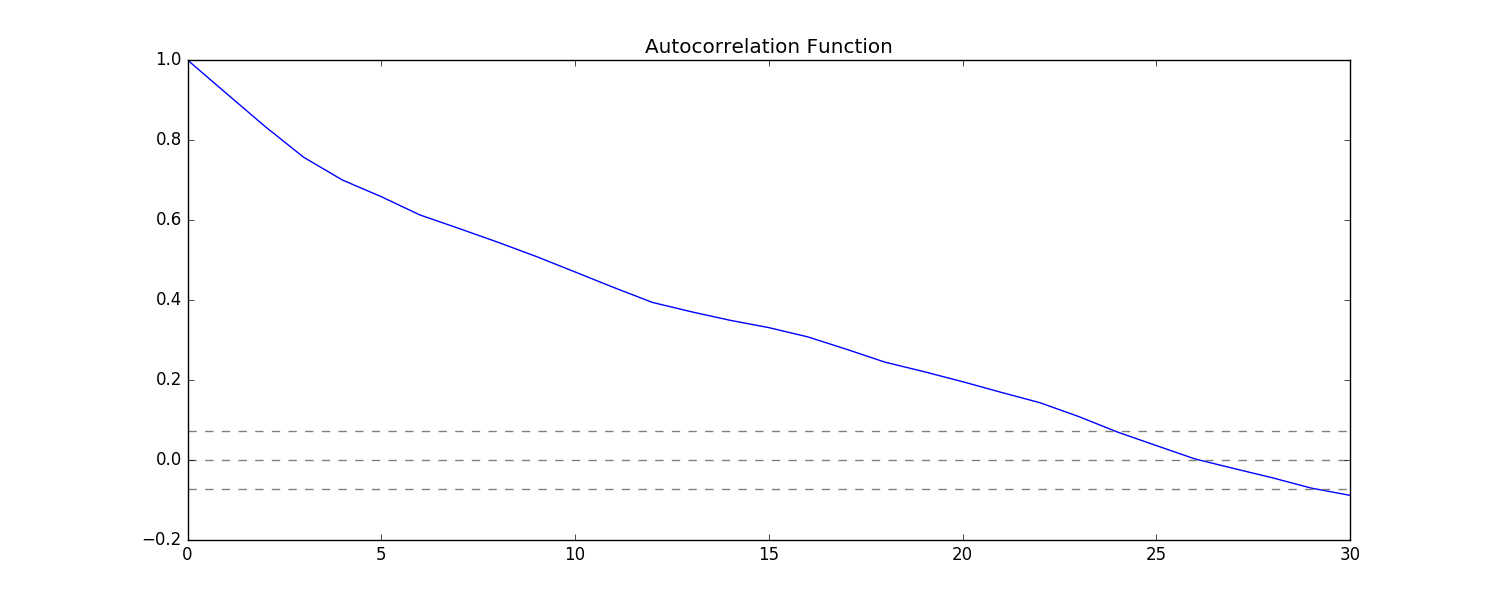
****

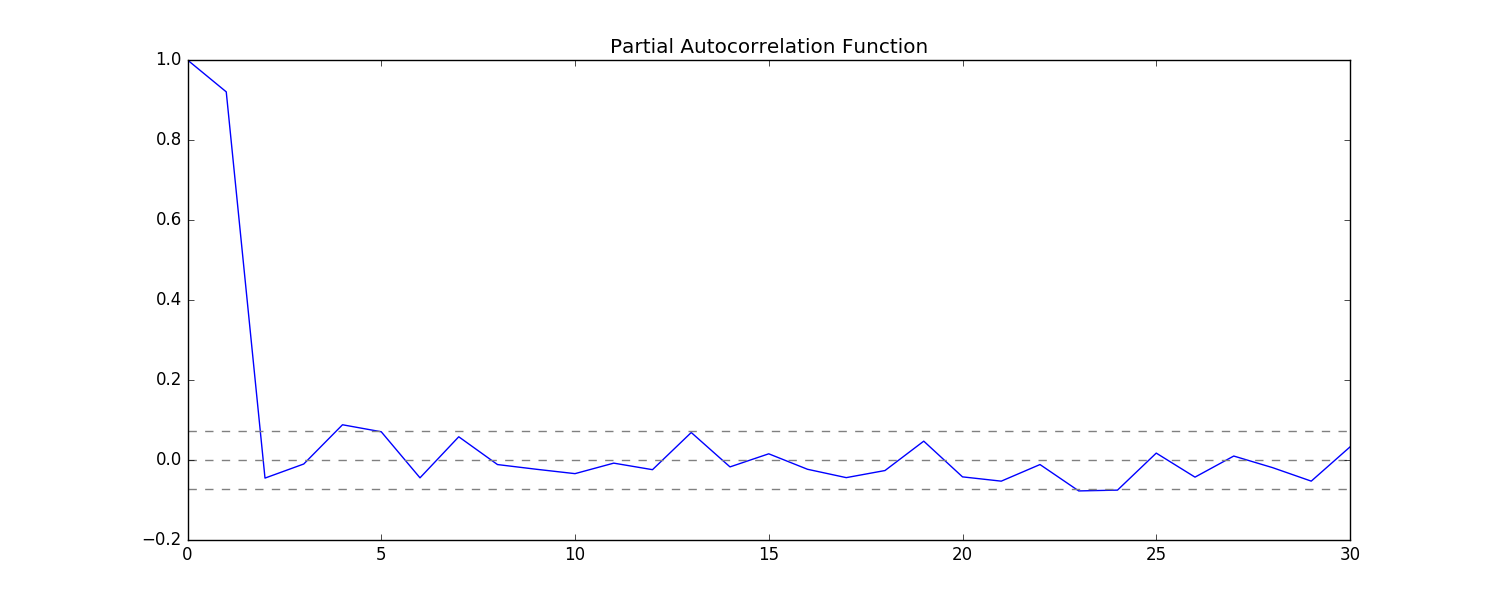
****

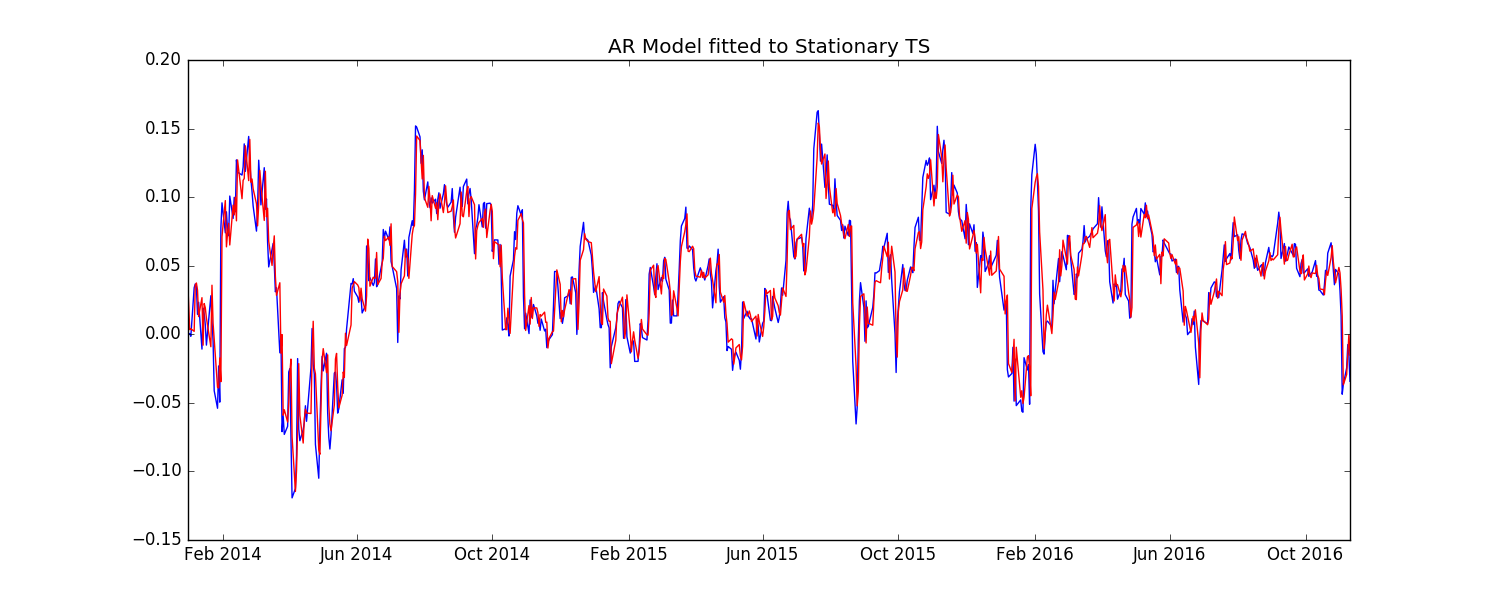
****

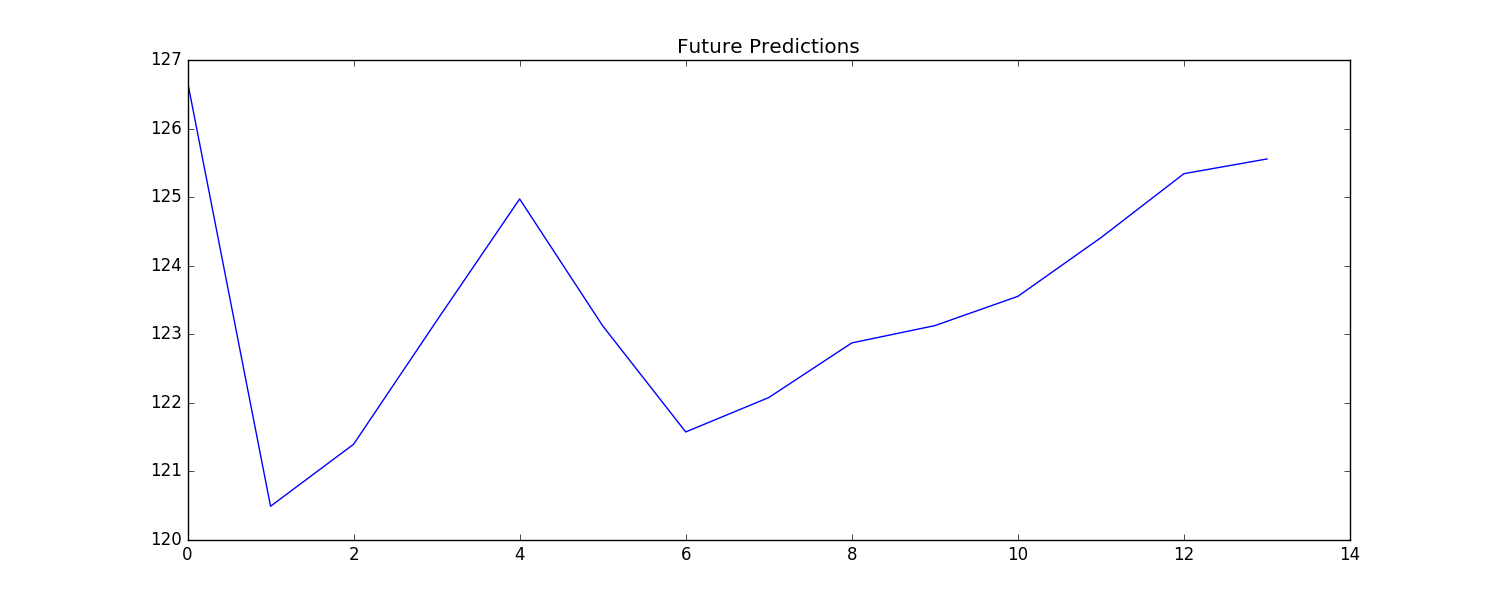
****

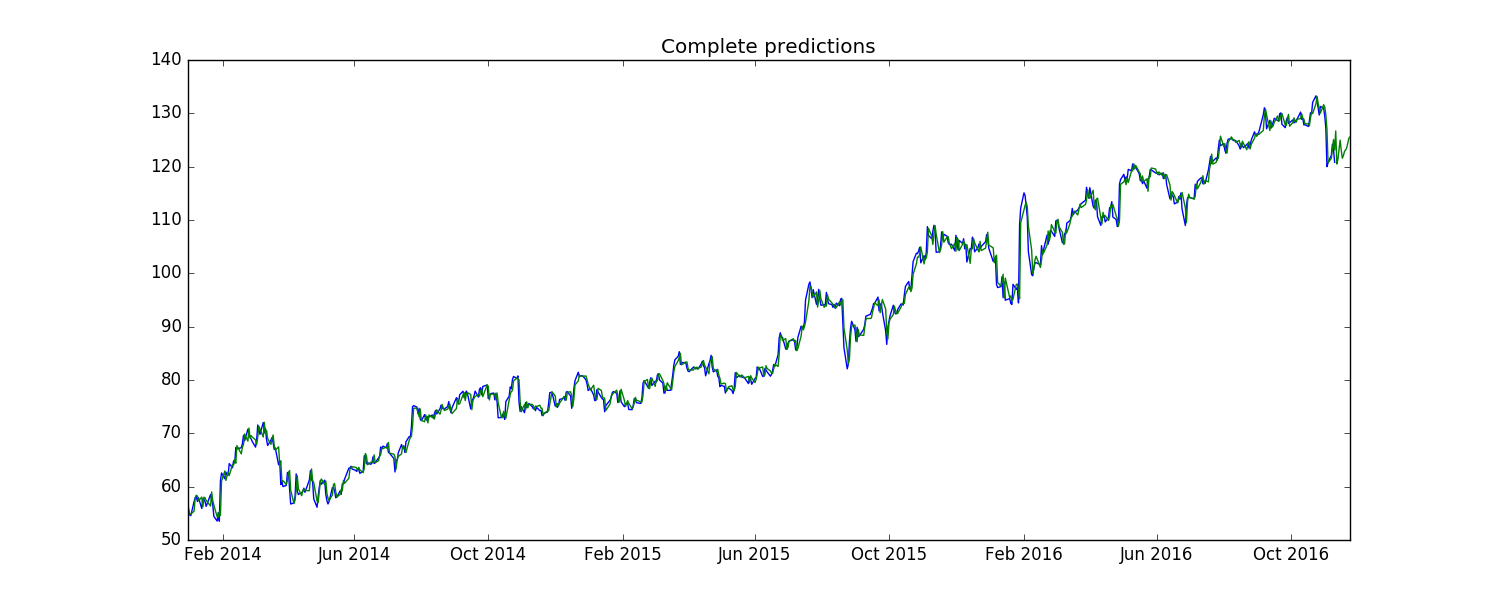
****

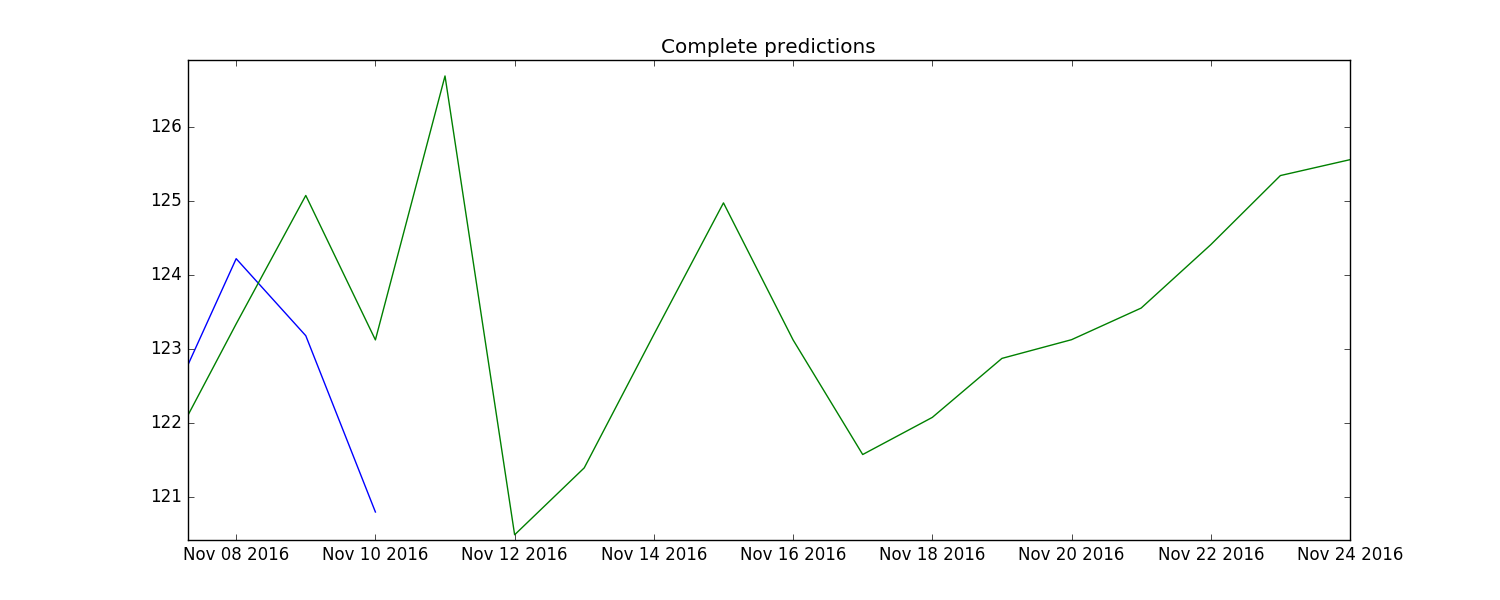
****

****

****

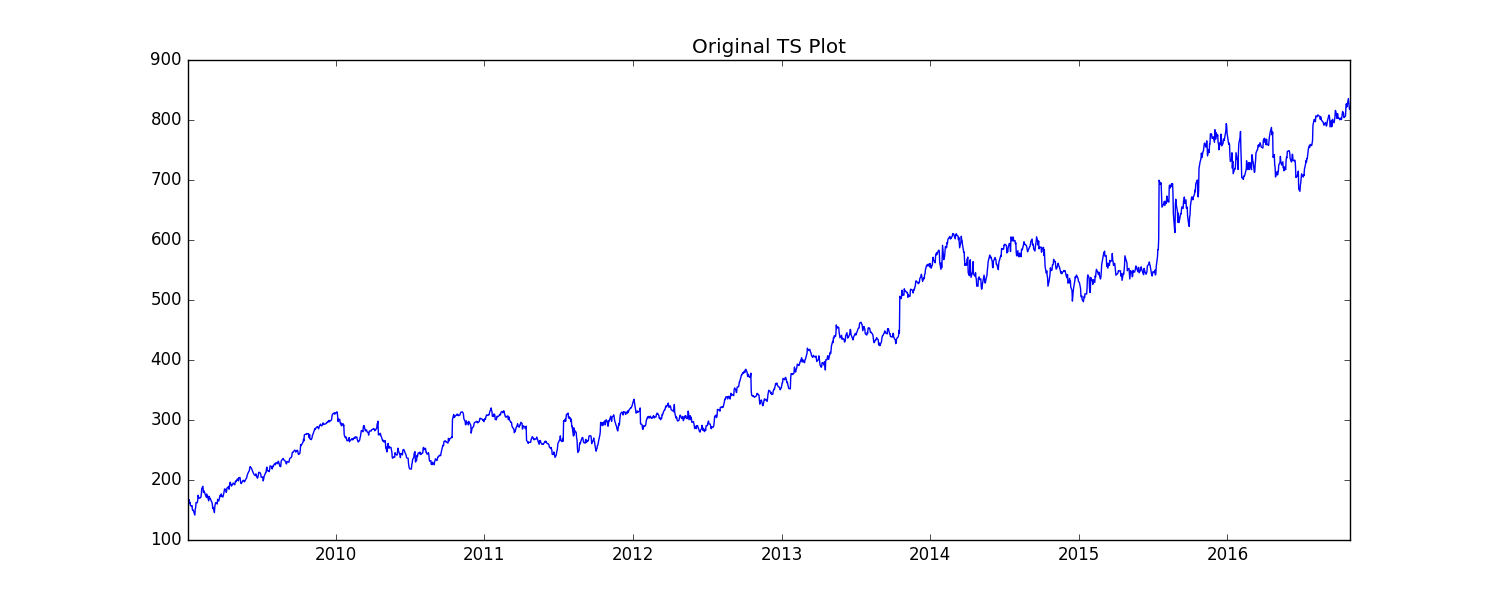
****

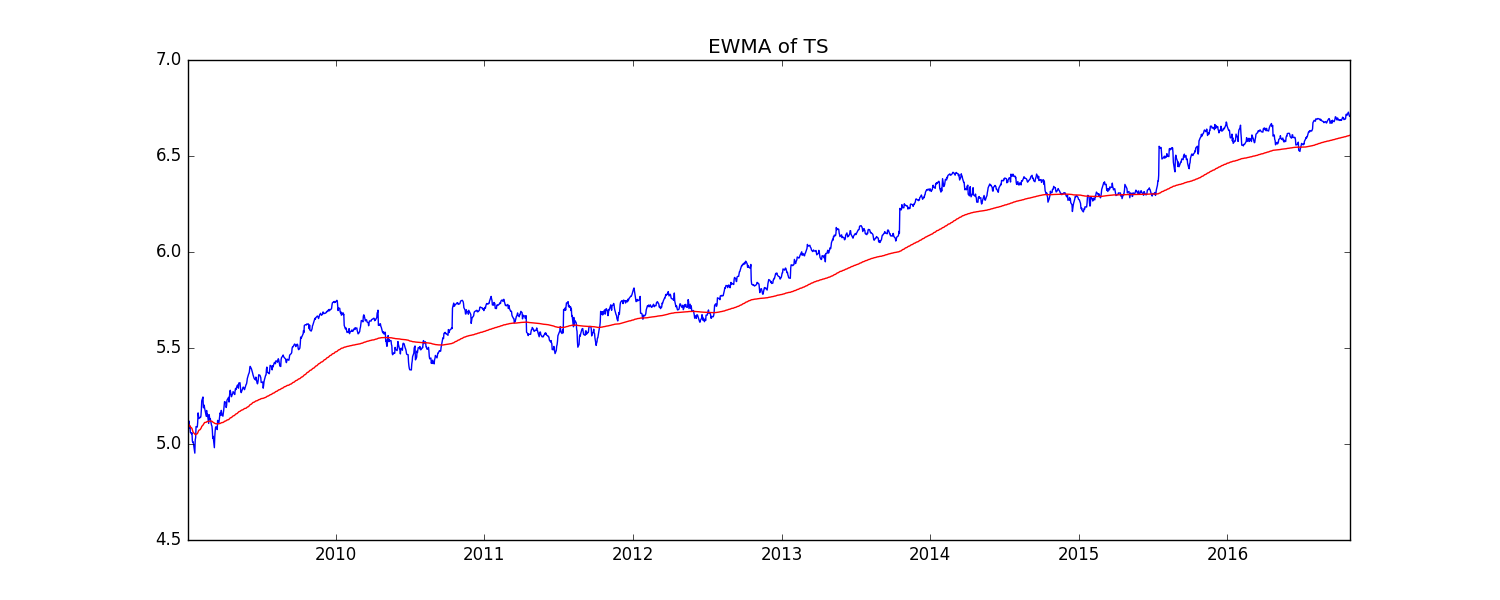
****

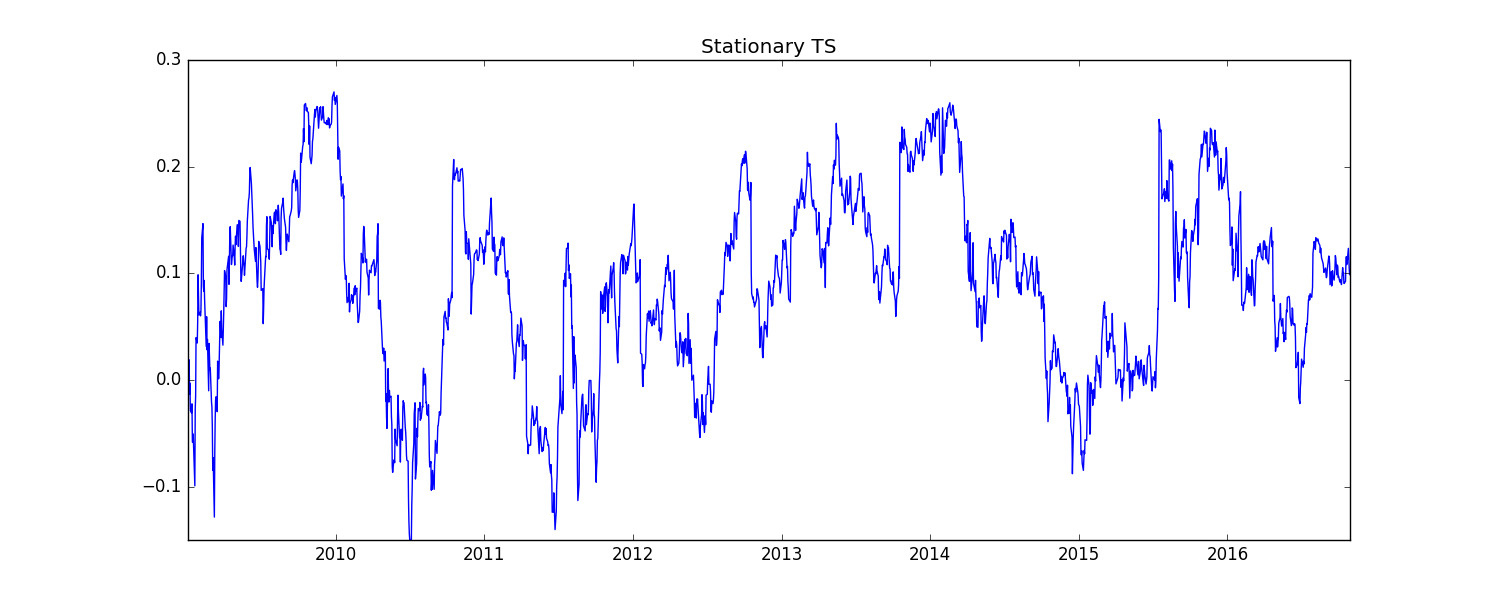
****

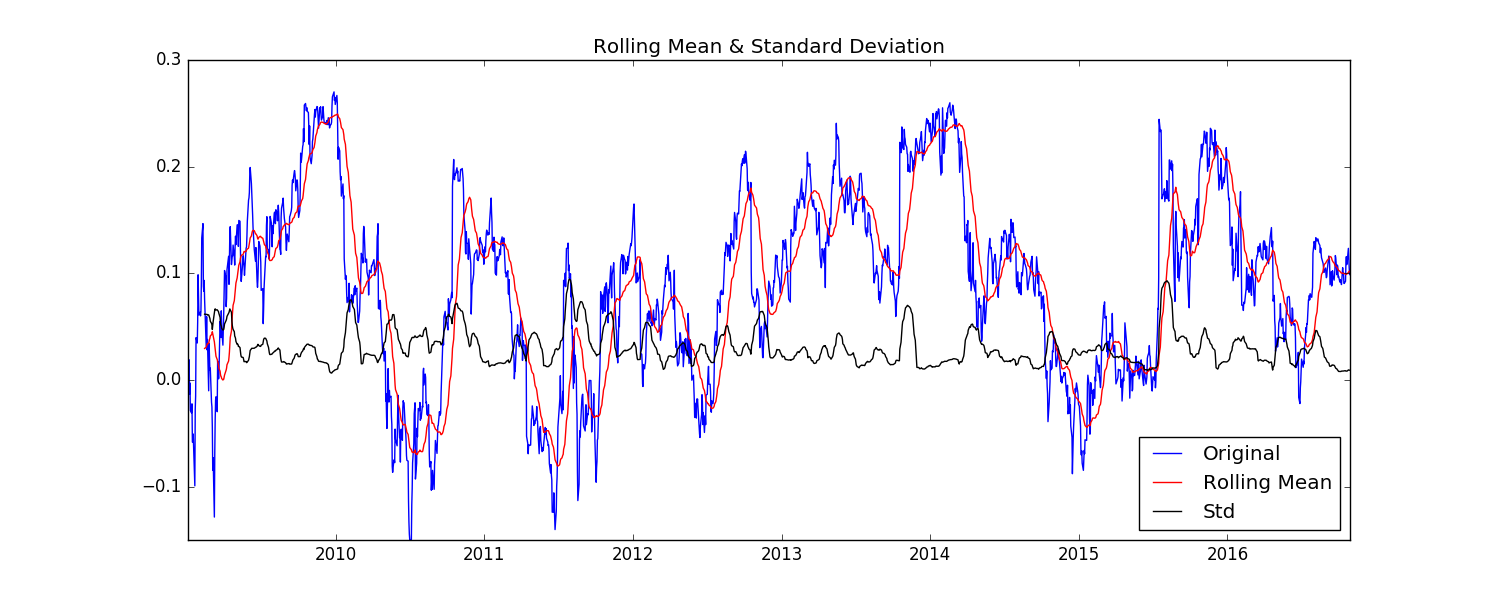
**Modelling for Google Stocks Data**:

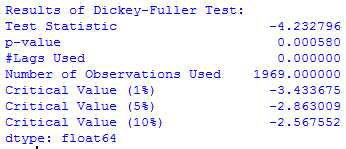
**GOOGLE**

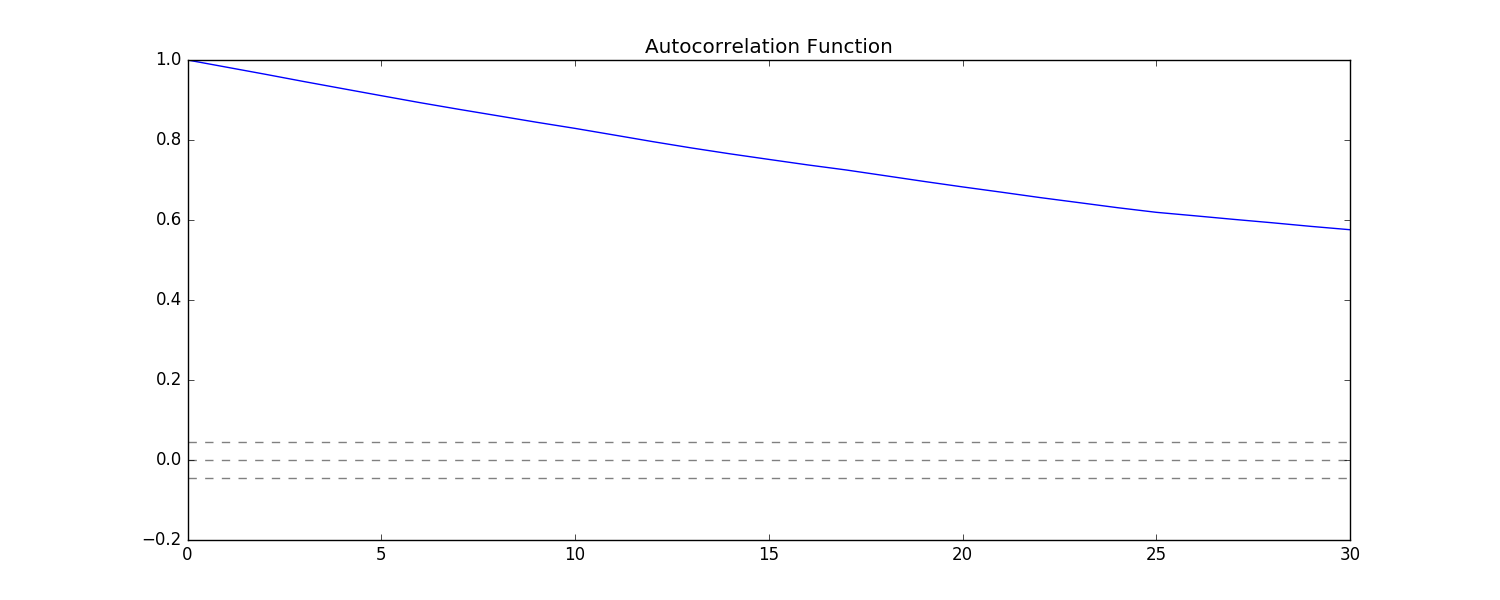
****

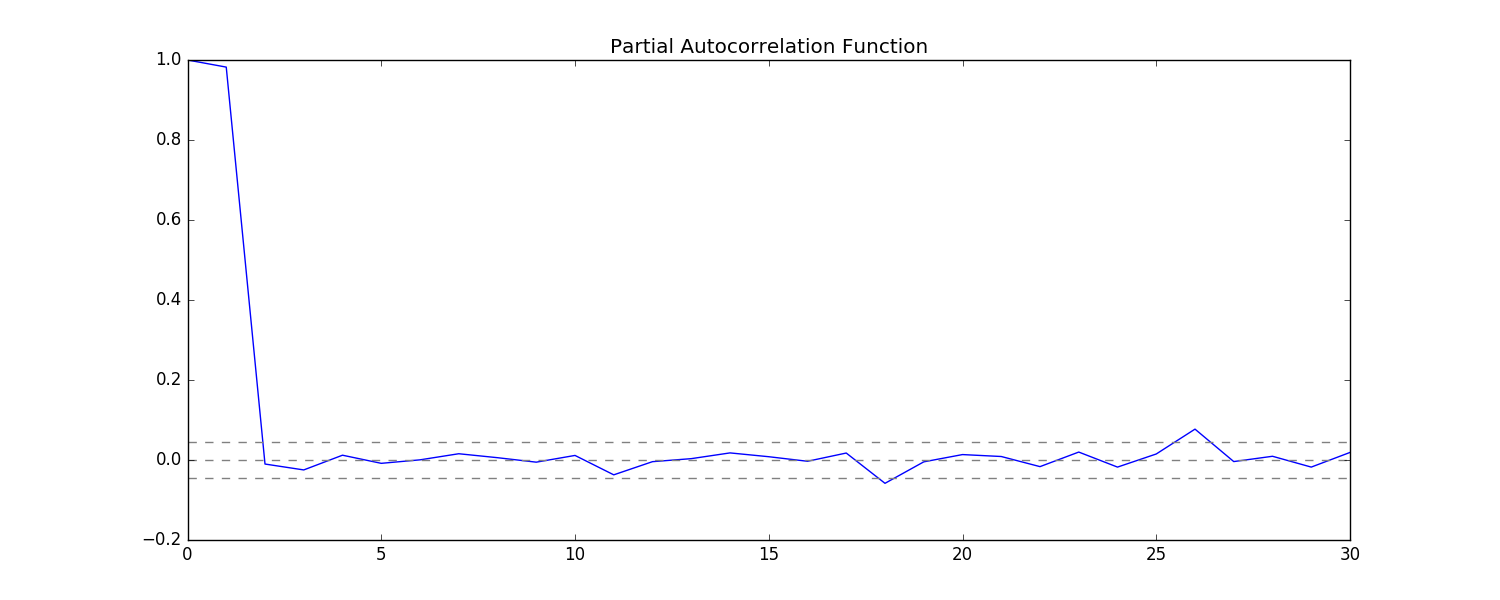
****

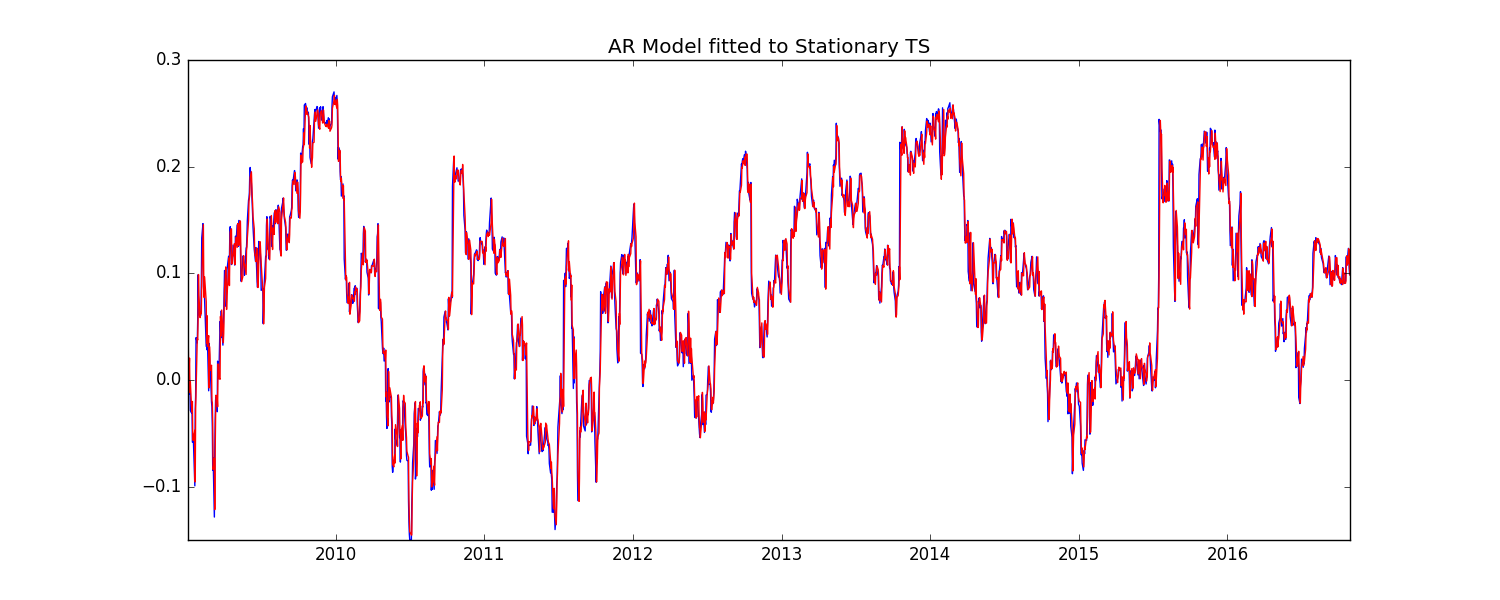
****

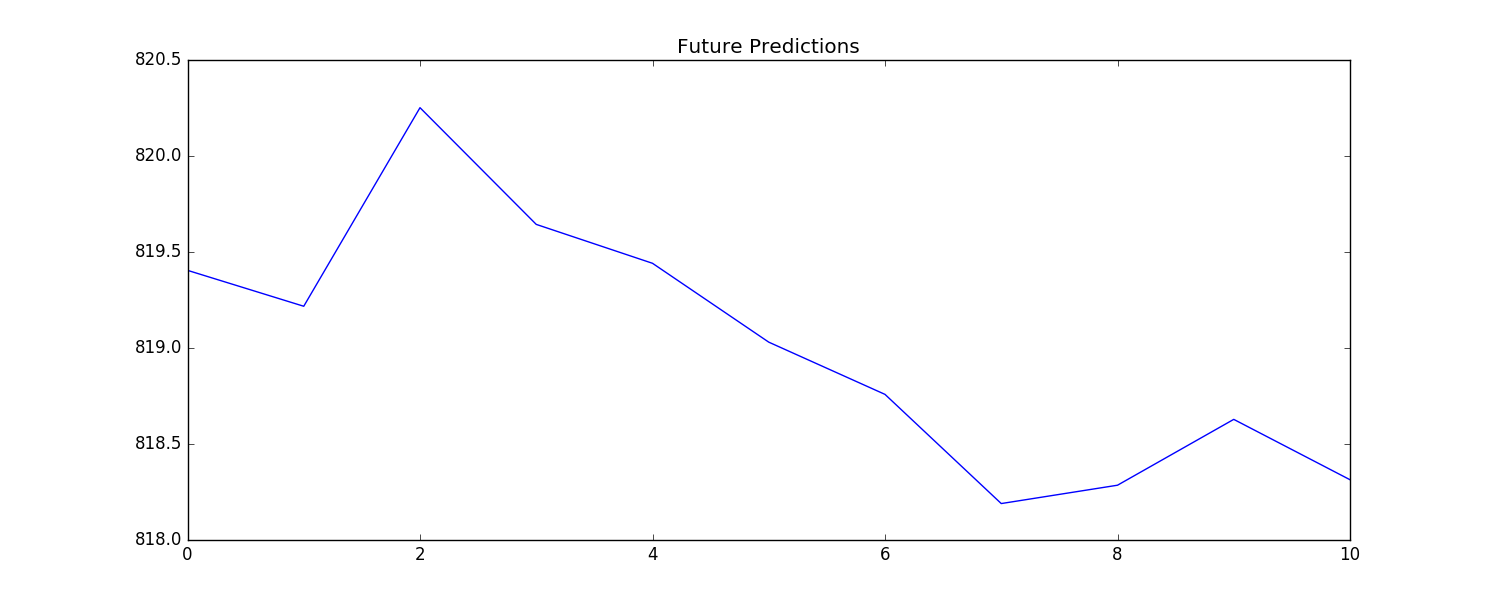
****

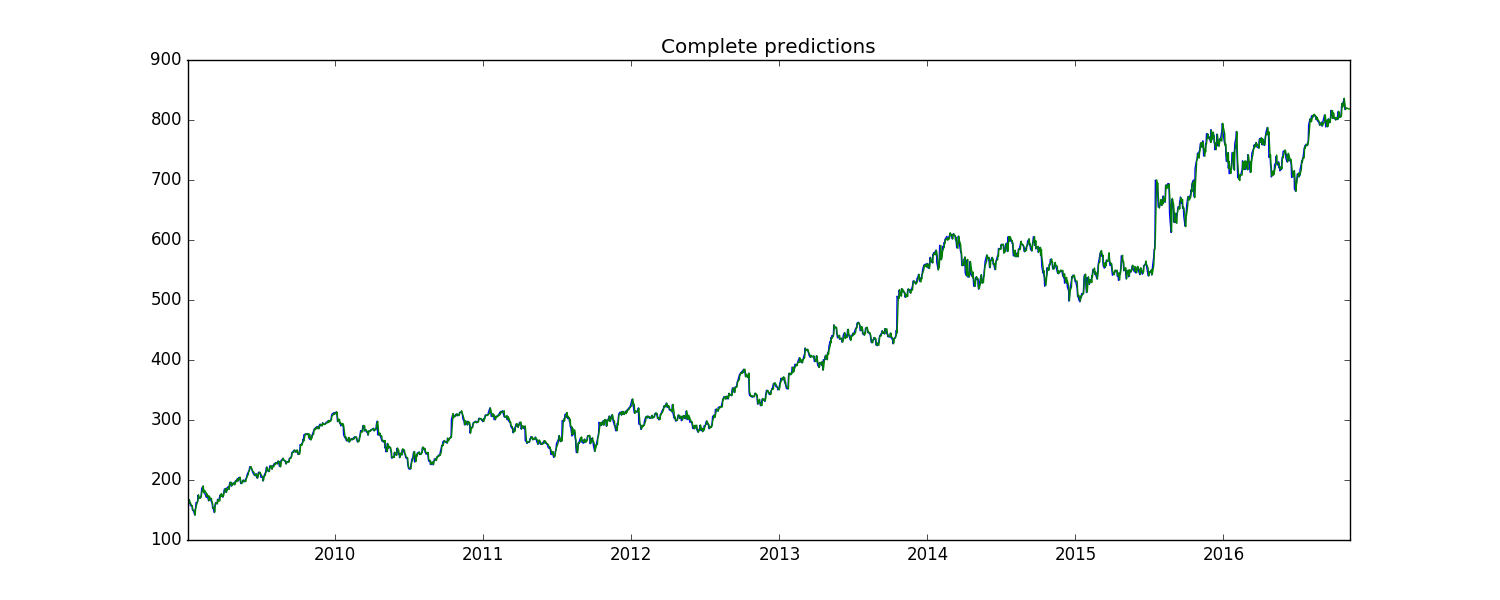
****

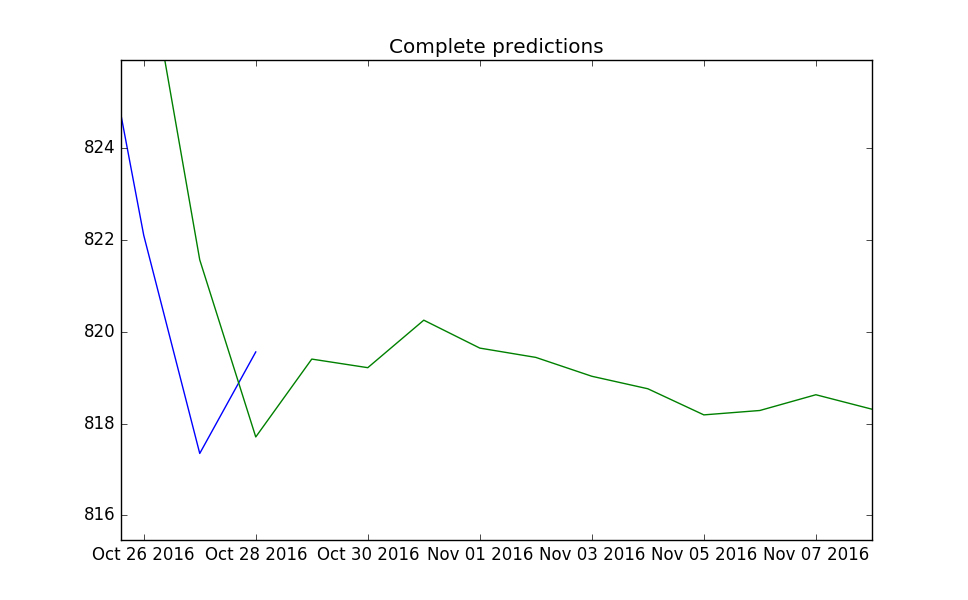
****

****









Results:

The prediction for Facebook future stocks is more varying and I think is reliable. But for Google, the predictions are not much varying but still can be reliable.