**Time Series Analysis of S&P 500**

A. Pawan DIXIT & Shubham CHITKARA

1. **Exploring Data**

The S&P 500 daily stock data was downloaded from Wall Street Journal dating from January 1995 to March 2018. The below figure shows the daily closing stock prices for the given data.

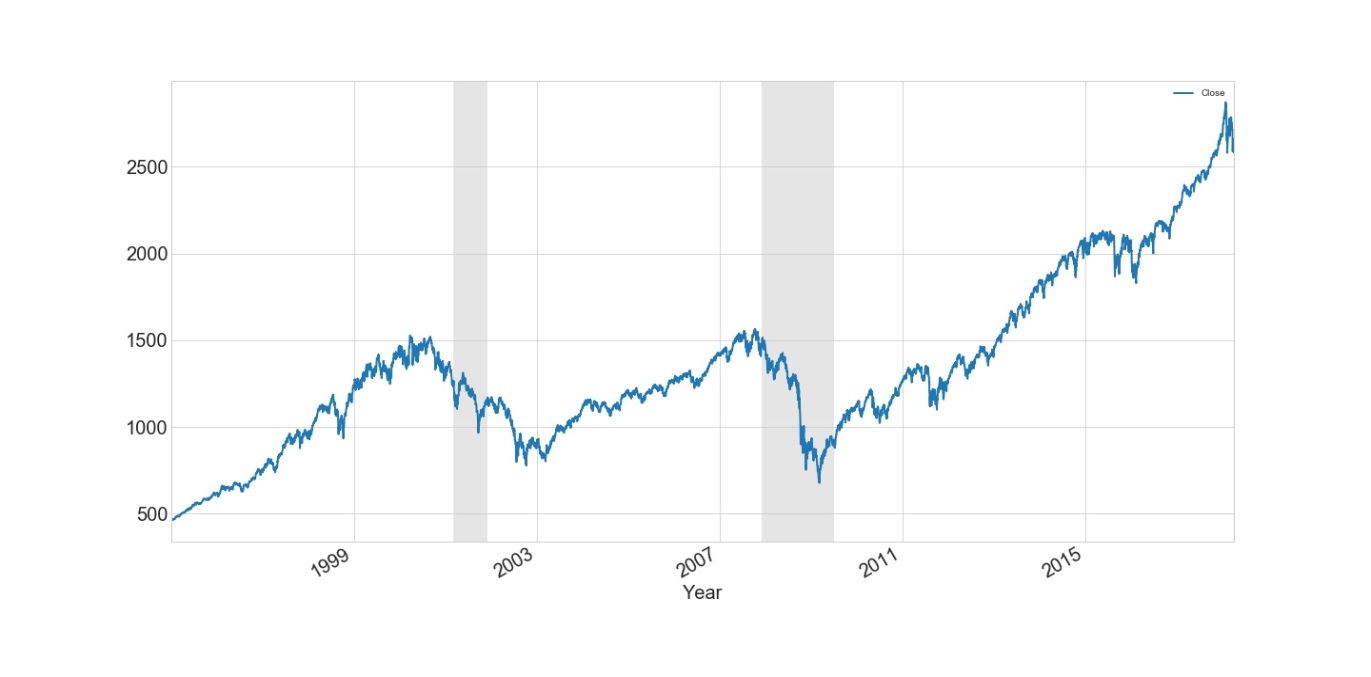


Figure -S&P 500 daily index

As we can see that stock prices went down significantly during the periods that are marked grey in the figure above. These represent periods of recession in United States during 2001 & 2008.

We took the moving average of the data over 1 year period to make the series smooth and to see the overall trend over last 23 years.

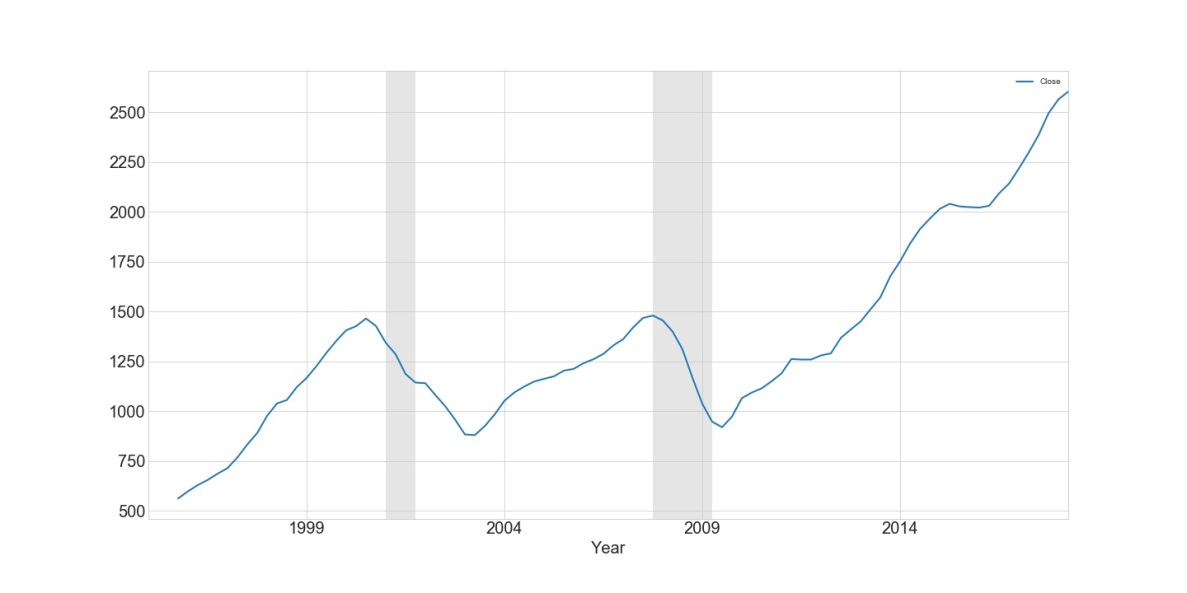


Figure - S&P 500 Moving average (1-year)

From the above figure, we can say that there is an overall upward trend in the stock prices except for during the recession period which kind of makes sense as the economy went down as well.

So, after seeing these plots we went ahead to see if we can find any relationship between US’s economy and S&P 500. We normalized quarterly GDP of United States and quarterly prices of S&P 500 to bring both of them to the same scale.

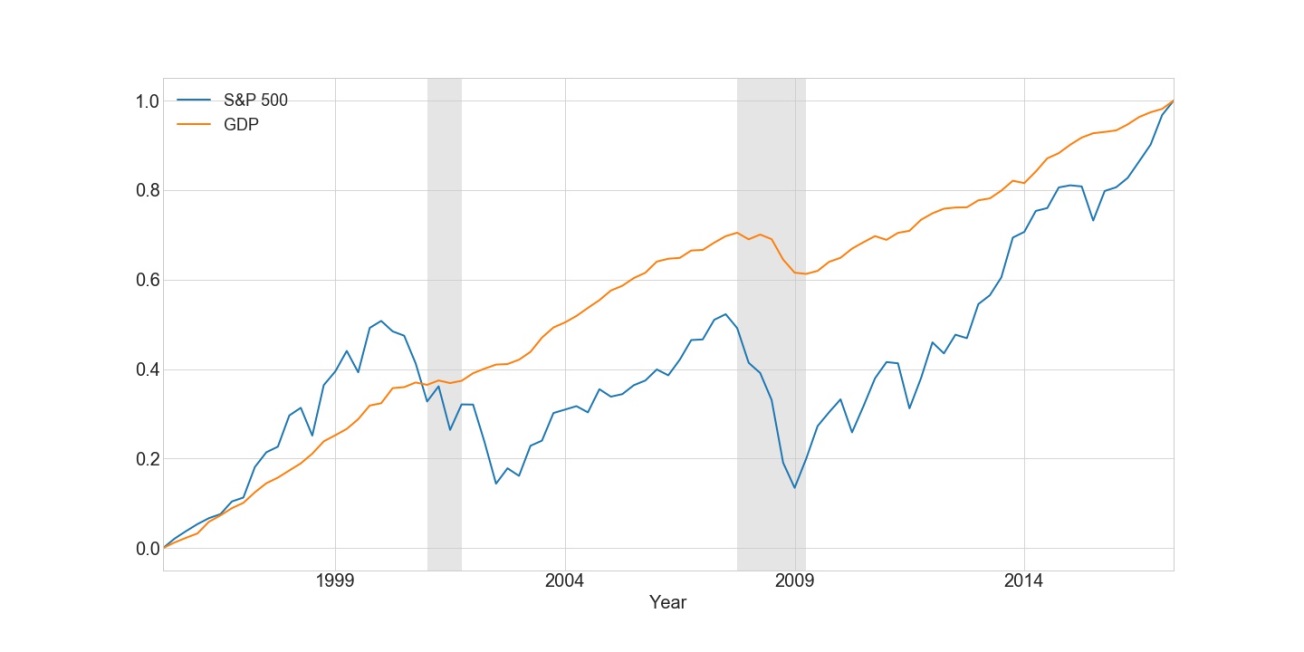
****

Figure - Normalized S&P 500 quarterly vs Normalized GDP quarterly

Although at first it might look as if there is no significant relationship between these two but on observing carefully, we can see that even small changes in GDP inflicts significant changes in S&P 500. During the 2001 recession the GDP goes flat from an increasing state and simultaneously S&P 500 stock index saw a significant decline. Similarly, during the 2008 recession, the economy went down and S&P Index also declined to a great extent. Since then GDP has always been increasing and so are S&P 500 Indices. This is probably because the stock prices are more volatile than a country’s GDP.

Exploring the data further, we looked for autocorrelation in the series using autocorrelation plot.

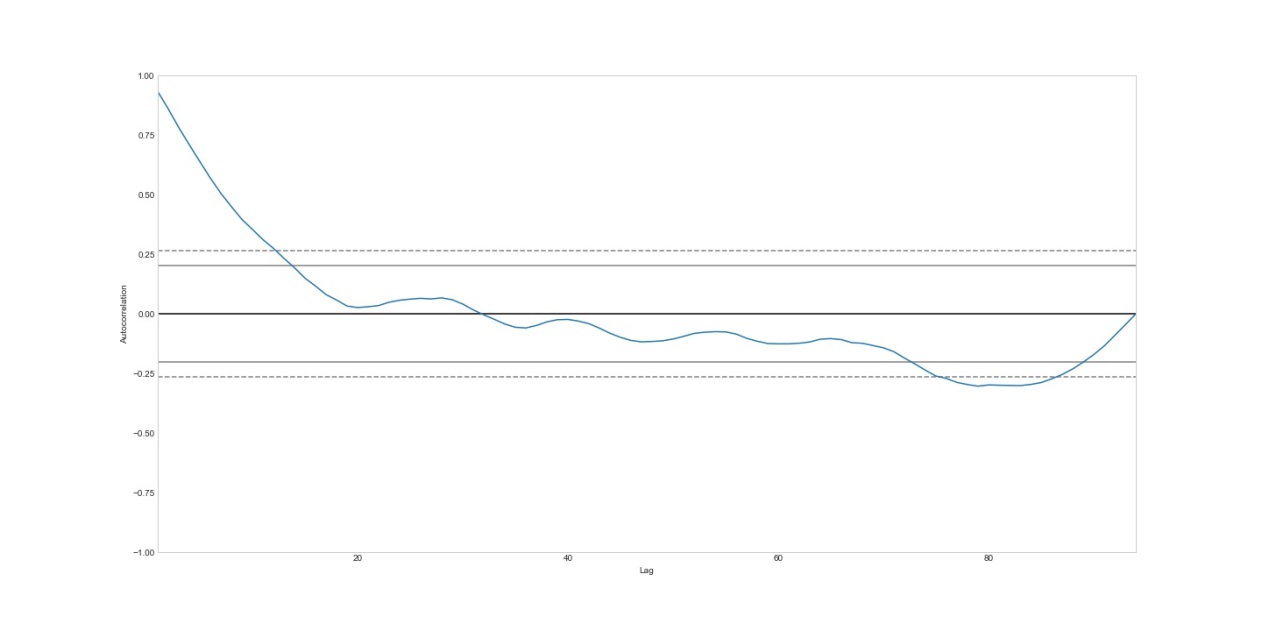


Figure - Autocorrelation of quarterly S&P 500 index

The first few lags in the series are highly correlated which gives the indication of a stochastic trend. Further we used the Dickey-Fuller test to check for stationarity of the series which indicated that the series is non-stationary and confirmed the stochastic trend. To remove this stochastic trend, we took the first order difference of the series which turned out to be stationary. The below graph represents the first order differenced series. It is clear that none of the lags are highly correlated.

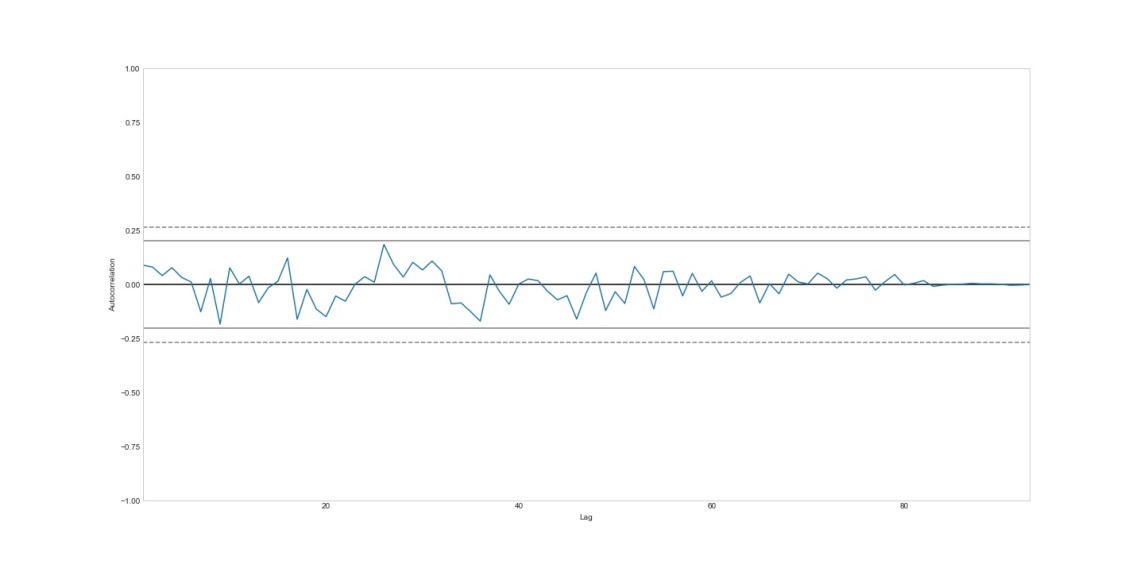


Figure - Auto-correlation plot of first order difference of S&P 500 index

Now that we know that the first order differenced series is stationary, we can go ahead and model the series using ARIMA. We used the *auto.arima* function in the *forecast* package of R to find the best ARIMA model for this time-series on the basis of BIC (Bayesian Information Criterion). After creating the model, we used one-step ahead recursive forecast method for prediction. The figure below shows the original data vs the forecasted values for the series.

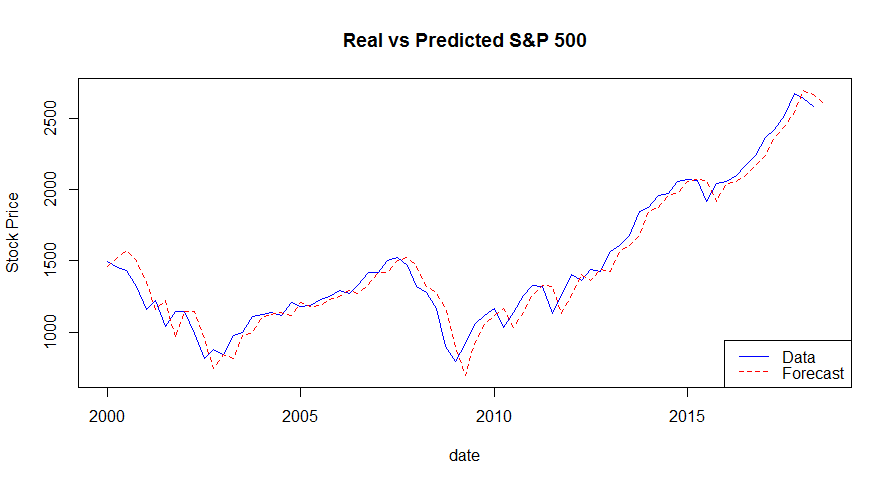


Figure - ARIMA Model and its Forecast

The Mean Squared Forecast Error value for this ARIMA model was 10835.62

We then created a Recursive Neural Network (RNN) model to forecast the time-series using *tensorflow* package in Python. The RNN had 1 hidden layer with 500 neurons and a learning rate of 0.001. This model used a basic RNN cell and it ran for 200 epochs. The figure below shows the original data vs the forecasted values for the series.

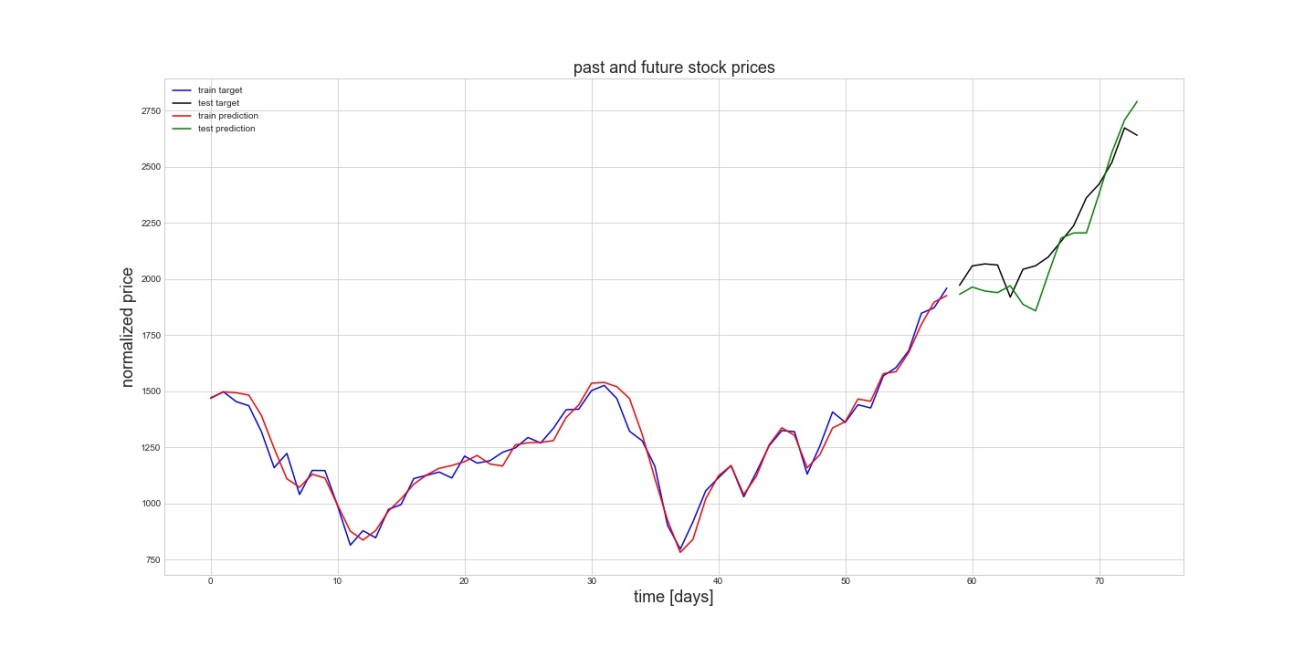


Figure - RNN Model and its forecast

The Mean Squared Forecast Error value for this RNN model was 1860.64

1. **Conclusion**

To sum up things we found that the S&P 500 index increased from 459.11 to 2581.88 which clearly shows there is an increasing trend. On further investigation, we found that it was a stochastic trend so we used first order differencing to make it stationary.

We created ARIMA and RNN models to forecast S&P 500 series and found that RNN outperforms ARIMA on the basis of Mean Squared Forecast Error (MSFE). So if we have to forecast the stock prices just on the basis of previous data we would recommend Investors to use our RNN model over ARIMA.

On an end note we would like to mention that it is never easy to forecast stock indices in real world because it is not just the trend but also various other factors that affect the stock prices which are even more difficult to predict. Had it been easy to predict stock prices, everyone would have been a billionaire. So, it is always better to have insights of an economist along with the results of statistical model to get a good forecast.

**References**

1. Time Series Analysis and Its Applications, Robert H. Shumway, David S Stuffer, 2006
2. Introduction to Econometrics, Stock, J.H. & M.W. Watson, 2014
3. <https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/>
4. [https://en.wikipedia.org/](https://en.wikipedia.org/https://www.pyimagesearch.com/2016/09/26/a-simple-neural-network-with-python-and-keras// )
5. [https://www.pyimagesearch.com/2016/09/26/a-simple-neural-network-with-python-and-keras//](https://en.wikipedia.org/https://www.pyimagesearch.com/2016/09/26/a-simple-neural-network-with-python-and-keras// )
6. <https://iamtrask.github.io/2015/11/15/anyone-can-code-lstm//>
7. <https://blog.statsbot.co/time-series-prediction-using-recurrent-neural-networks-lstms-807fa6ca7f>
8. <https://www.investopedia.com/articles/trading/07/stationary.asp>