STOCK MARKET PREDICTION USING TIME SERIES ANALYSIS

Pooja Radhakrishnan, Gokul Krishnaswamy, Gowtham Krishnaswamy

UNIVERSITY of DENVER

Daniel Felix Ritchie School of Engineering and Computer Science

ABSTRACT

Stock markets are uniquely complicated. A successful prediction of stock's future price could yield significant profits. We have used the closing points of the stock market index of a month as a factor to predict the future stock prices. Time series analysis is done to achieve this task.

RESEARCH QUESTION

To predict the closing value of stock in the months to come in the near future, can we implement the same using the **ARIMA** model of Time Series Analysis?

DATA SET

The 'Stock Market' dataset was extracted from the service provider '*Yahoo.com*' using the "**^GSPC**" function of '**quantmod**' package in R for the time span 03-01-1999 to 03-01-2019 (20 years). It has daily data with the following attributes: Date, Opening value, Highest value, Lowest value, Closing Value, Volume and Adjusted value.

date	open	high	low	close	volume	adjusted
1999-03-01	1238.33	1238.70	1221.88	1236.16	699500000	1236.16
1999-03-02	1236.16	1248.31	1221.87	1225.50	753600000	1225.50

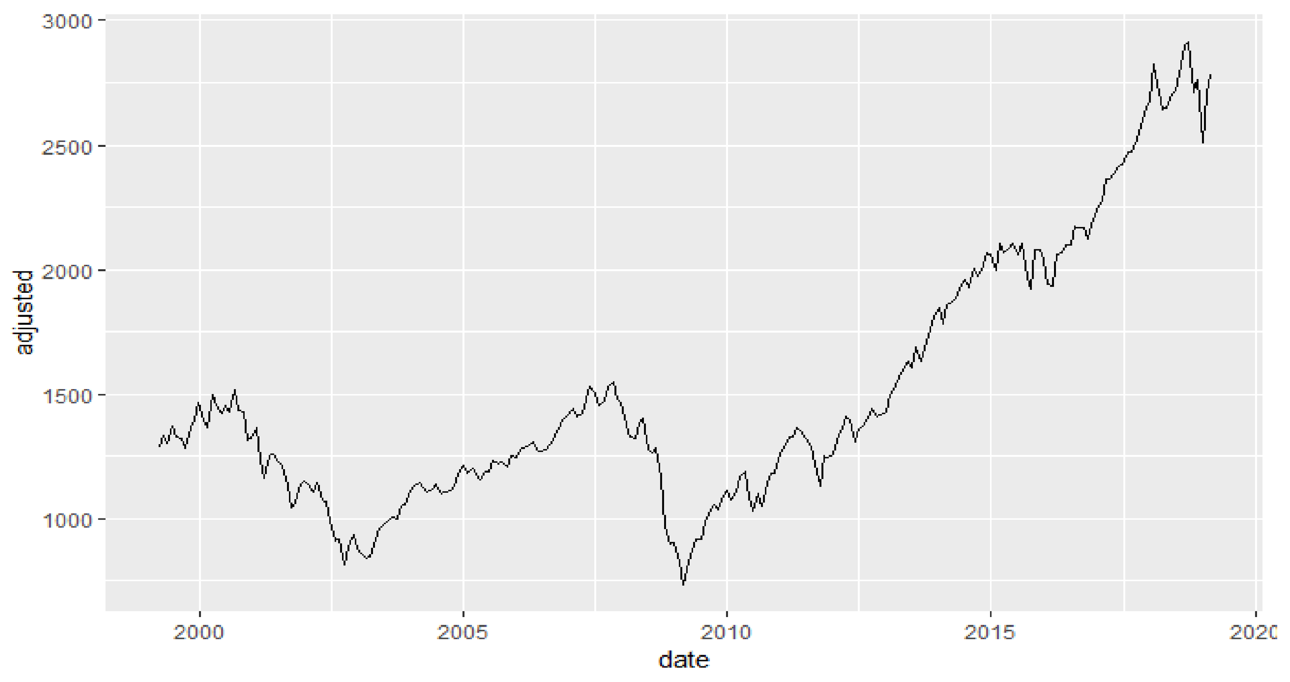


Figure 1a): Raw Stock Data, 1b) Monthly returns plot

ARIMA MODEL

ARIMA stands for '**AutoRegressive Integrated Moving Average'.**

AR – the values of a given time series data are regressed on their own lagged values. ***(p)***

I – differencing the time series data to remove the trend and convert a non-stationary time series to a stationary one. ***(d)***

MA – the number of lagged values of the error term. ***(q)***

METHODOLOGY

- Exploratory Data Analysis
- Decomposing the data to check for the presence of the following non stationary components: Trend, Seasonality, Cyclicity, Randomness.
- Check for Stationarity using Augmented Dicky Fuller's test.
- If not, making the data stationary by differencing.
- Extract the values of *p* and *q* using the ACF and PACF plots.
- Fit an **ARIMA** model both manually, and using the kernel.
- Forecast for the required period desired.

DATA SATISFACTION OF METHOD

- The dataset was explored and analyzed, and contained no missing values.
- Upon Decomposition, the presence of the non stationary components namely Trend, Seasonality, Cyclicity and Randomness were observed.
- This was confirmed by the Augmented Dickey – Fuller's test.
- The data was made stationary using differencing.
- The ACF and PACF plots gave the p and q values.
- Hence, now, the dataset can be fitted into ARIMA model and meets all requirements.

APPLICATION OF METHOD

- Test for stationarity using decomposition:

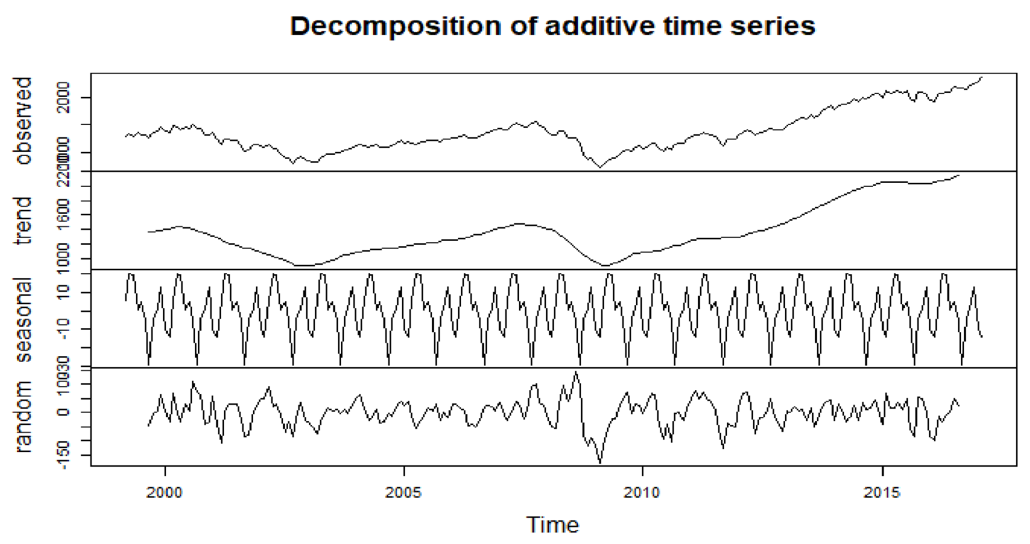


Figure 2 : Checking for stationarity using decomposition

- Retrieving the p—value from ADF test to accept or reject the null hypothesis that the data is non stationary.

```
        Box-Ljung test

data:  stock_ts
X-squared = 3183, df = 20, p-value < 2.2e-16


        Augmented Dickey-Fuller Test

data:  stock_ts
Dickey-Fuller = -1.1724, Lag order = 6, p-value = 0.9091
alternative hypothesis: stationary

-value is greater than 0.05 => The Null hypothesis that the data is not stationary is accepted.
```

Figure 3: Accepting null hypothesis using ADF test

- Since the data is non-stationary, we make it stationary using differencing technique, which returns the value of 'd' of the **ARIMA** model.

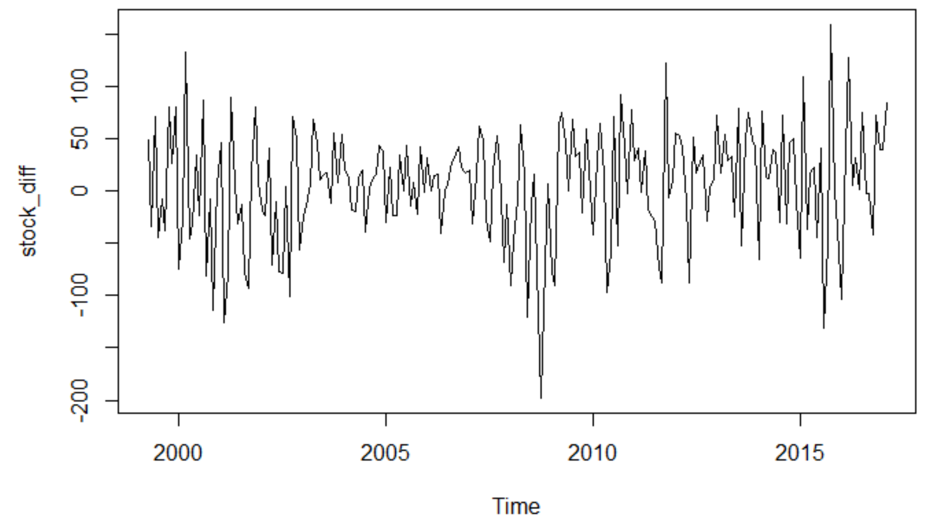


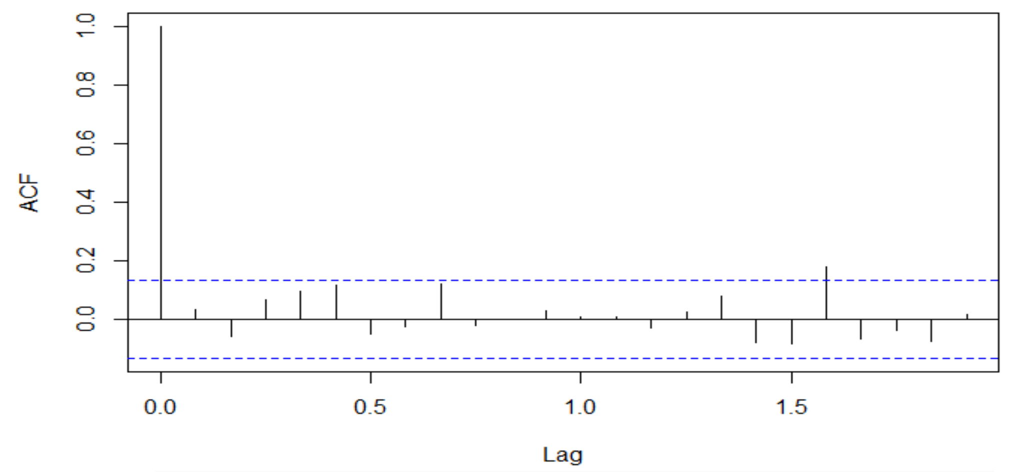
Figure 4: Stationary data with constant mean and variance

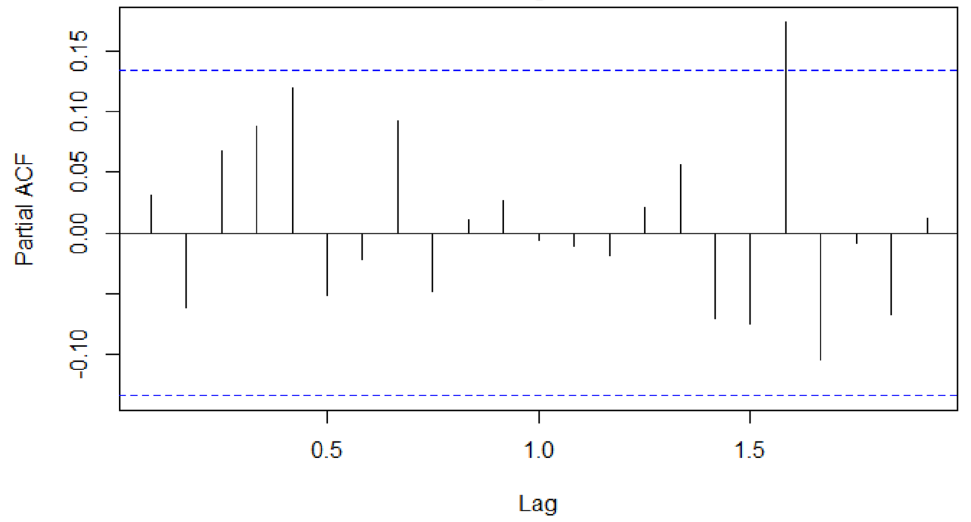
- Rechecking the stationarity of data using ADF test. p-value is less than 0.05, and hence we reject the null hypothesis.

```
        Box-Ljung test

data:  stock_diff
X-squared = 24.548, df = 20, p-value = 0.2193

p-value smaller than printed p-value
        Augmented Dickey-Fuller Test

data:  stock_diff
Dickey-Fuller = -5.2098, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

The p-value is less than 0.05, which means the Null hypothesis that the data is not stationary is rejected.

Therefore, the data is stationary.
```

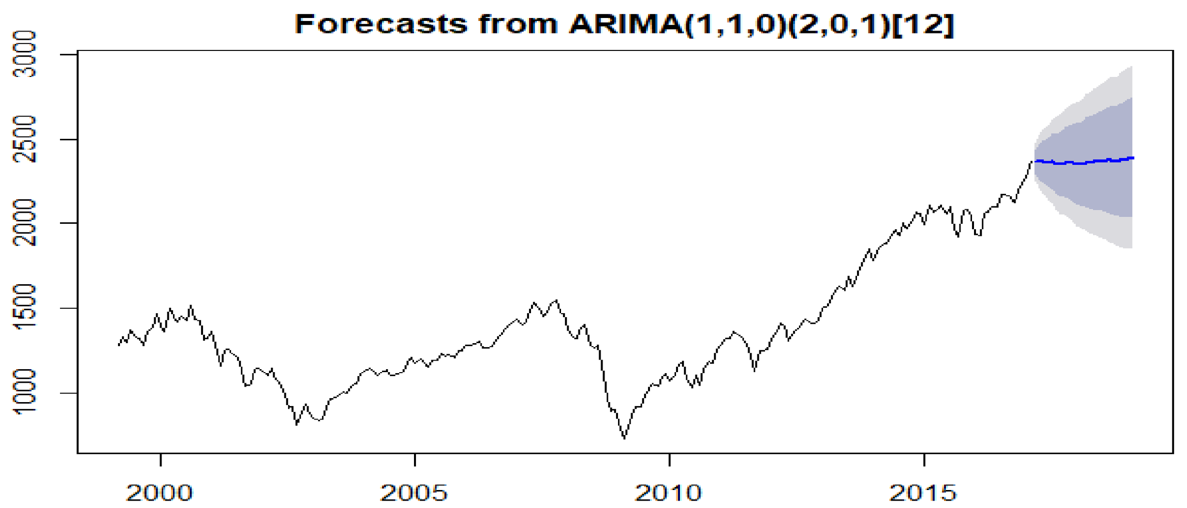
Figure 5: Rejecting null hypothesis using ADF test

- The *AutoCorrelation Function (ACF)* plot and *Partial AutoCorrelation (PACF)* plots give us the 'p and 'q' values of the **ARIMA** model.



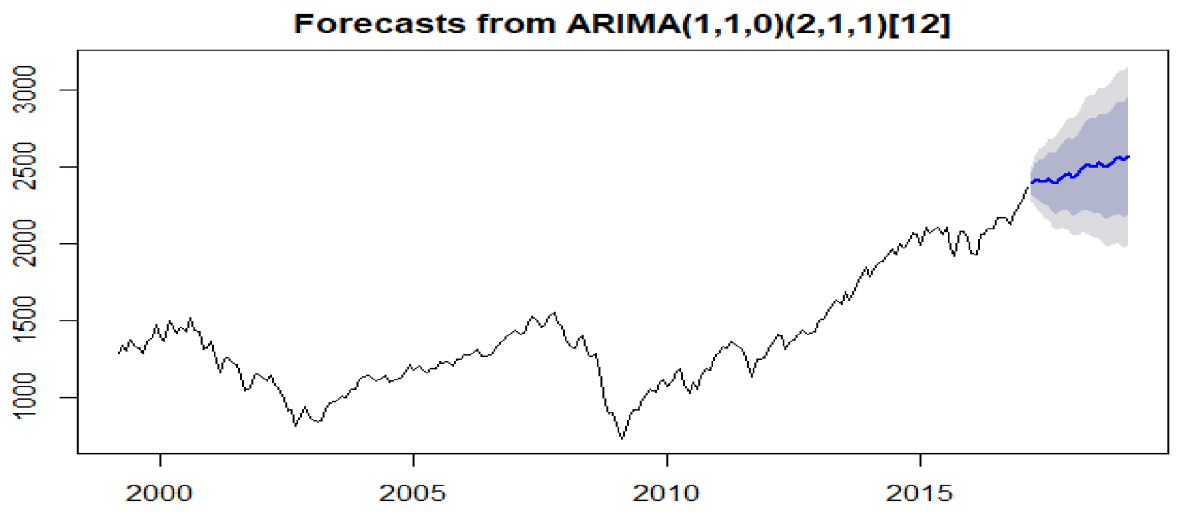


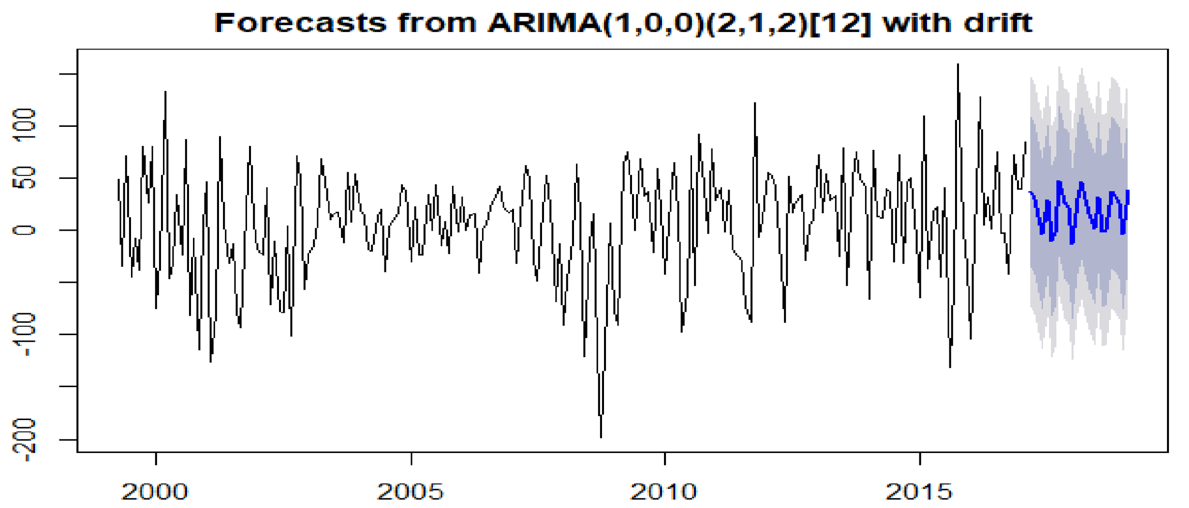
- The following result shows the result of the manual **ARIMA** model without SEASONALITY DIFFERENCE ORDER PARAMETER:



- The following result shows the result of the kernel **ARIMA** model with the SEASONALITY DIFFERENCE ORDER PARAMETER:

```
Series: stock_data_train
ARIMA(1,1,0)(2,1,1)[12]

Coefficients:
         ar1     sar1    sar2     sma1
      0.0432  -0.0517  0.0279  -0.8707
s.e.  0.0659   0.1041  0.1003   0.0929
```





RESULT

The forecast estimates above are provided with 80% confidence limits shaded in darker blue, and 95% in lighter blue regions in comparison with the actual values.