Stock Prediction for FaceBook

Ravi Teja | Siddharth | Karishma | Sruthi

The Objective

Company

- Founded: 2004
- Market leader in Social Networking
- Multiple child companies

FB ~ NYSE

- Investors
- Going public to raise capital : 2012
- Strong Investment

Objective

- Predict the stock value using time series model
- Understanding Time series Models

Data Description

	Date	Open	High	Low	Close	Volume	Adjusted Close
1	2013-01-02	27.44	28.18	27.42	28.00	69846400	28.00
2	2013-01-03	27.88	28.47	27.59	27.77	63140600	27.77
3	2013-01-04	28.01	28.93	27.83	28.76	72715400	28.76
4	2013-01-07	28.69	29.79	28.65	29.42	83781800	29.42
5	2013-01-08	29.51	29.60	28.86	29.06	45871300	29.06

- 1008 observations of 7 variables
- Dataset contains:

Date : Date of the observation

Open : Opening price of the stock for that particular day

High : Highest price of the stock on that day

Low : lowest price of the stock on that day

Close: closing price of the stock for that day

Volume : Number of Stocks traded

Adjusted Close: Adjusted closed price ~~of the~~ stock

Challenges deep-dive

Challenge 1

Missing Data

No Data for Saturday and Sunday

- As Stock market runs only for 5 days a week.

Challenge 2

Managing Missing Data

Created a continuous Time Series object

Selecting the Variable

- Close price / Adj Close Price

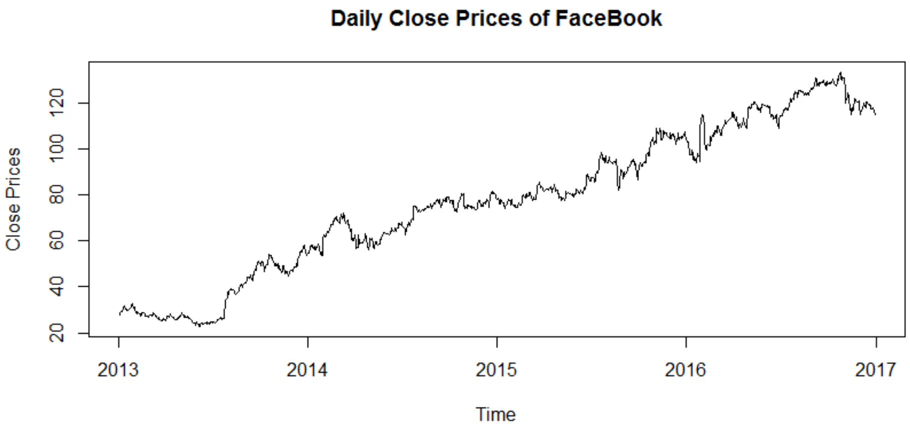
Challenge 3

Tradeoff b/w Close & Adj. Close

Selected Close Price so that we will have clear idea of the trading day without adjustments.

Data Analysis

Trend of Close price across Time



Not Stationary

Increasing Trend across Time

No Seasonality

Data Analysis cntd..

ADF Test:

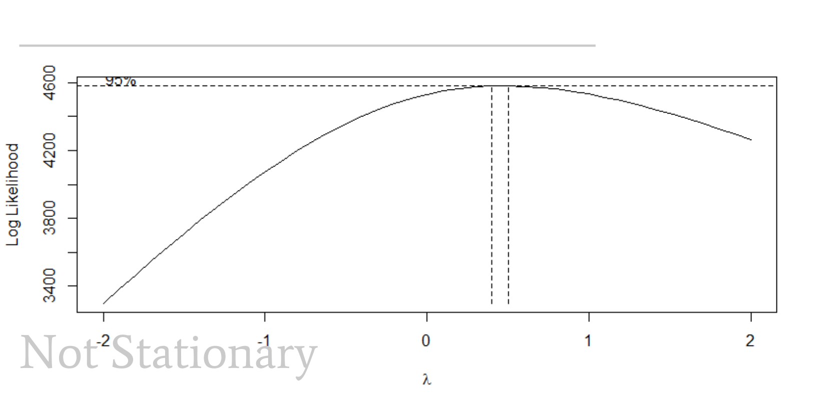
```
> adf.test(close_stock, alternative = "stationary")

        Augmented Dickey-Fuller Test

data:  close_stock
Dickey-Fuller = -2.9316, Lag order = 10, p-value = 0.1839
alternative hypothesis: stationary
```

Not Stationary

P-value > 0.05

Failed to reject Null Hypothesis(H_o)

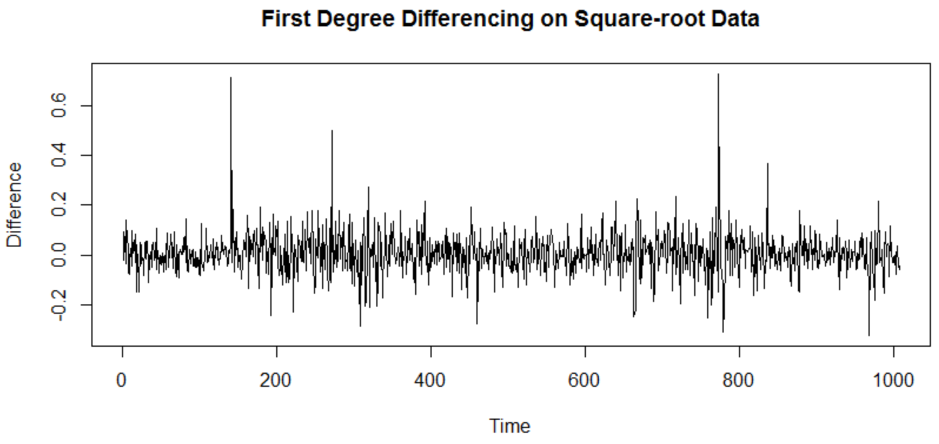
Box Cox Curve:



Not Stationary

Lambda = 0.5

Data Analysis cntd..



Transformation and Differencing:

Transformed Data to Square root

Differencing at lag 1 to achieve stationarity

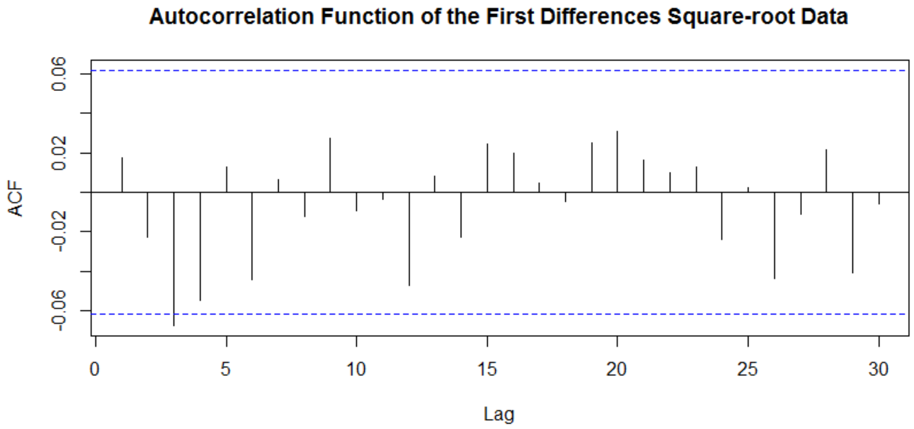
ADF Test:

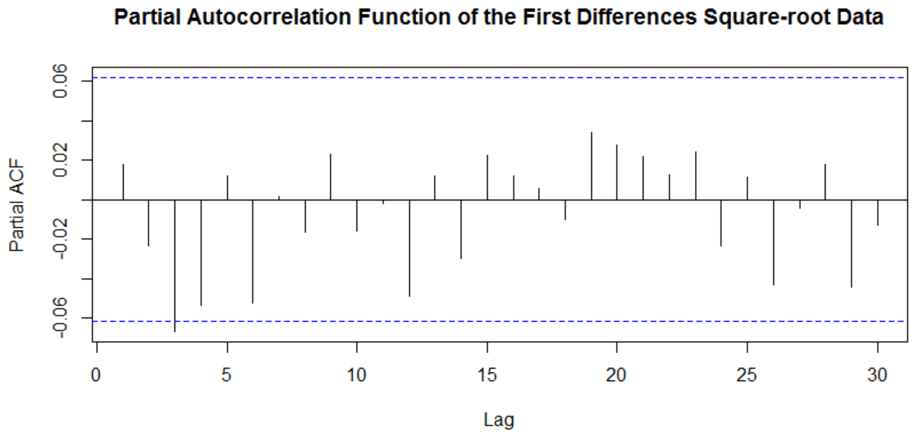
Stationary

P- value < 0.05

```
> adf.test(sd1, alternative = "stationary")

	Augmented Dickey-Fuller Test

data:  sd1
Dickey-Fuller = -10.104, Lag order = 10, p-value = 0.01
alternative hypothesis: stationary
```

ACF & PACF Plots





Random Noise

Data is independent

Model Selection

Model 1

Model from auto.arima function:

Suggested for ARIMA(1,1,4)

AIC = 3657.15

```
> auto.arima(close_stock)
Series: close_stock
ARIMA(1,1,4) with drift

Coefficients:
          ar1     ma1      ma2      ma3      ma4   drift
      -0.9408  0.9643  -0.0089  -0.1316  -0.1321  0.0867
s.e.   0.0378  0.0487   0.0447   0.0460   0.0342  0.0406

sigma^2 estimated as 2.194:  log likelihood=-1821.52
AIC=3657.04   AICc=3657.15   BIC=3691.44
>
```

Model 2

Model from ACF & PACF plots

Random walk : ARIMA(0,1,0)

AIC = 3669.69

$Y_t = y_{t-1} + e(t)$

Model Forecast

Forecast from ARIMA(1,1,4)

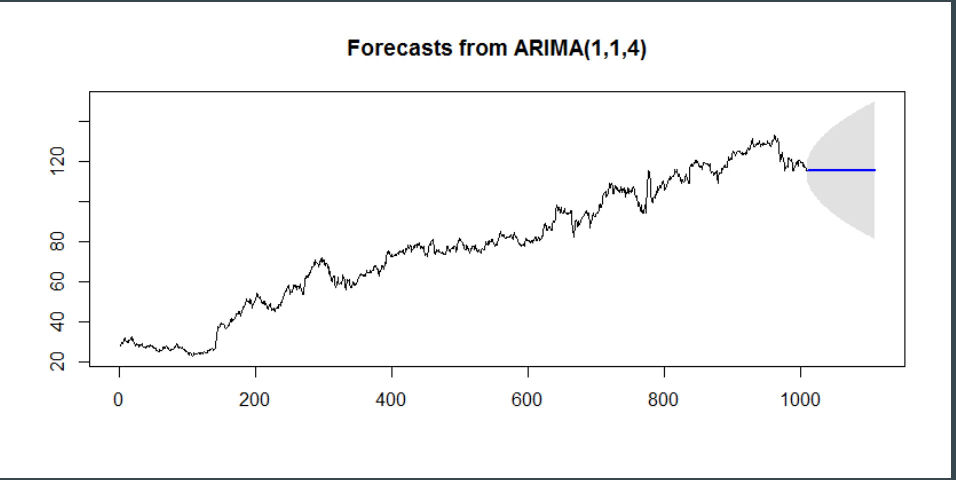
Forecast is not following Trend

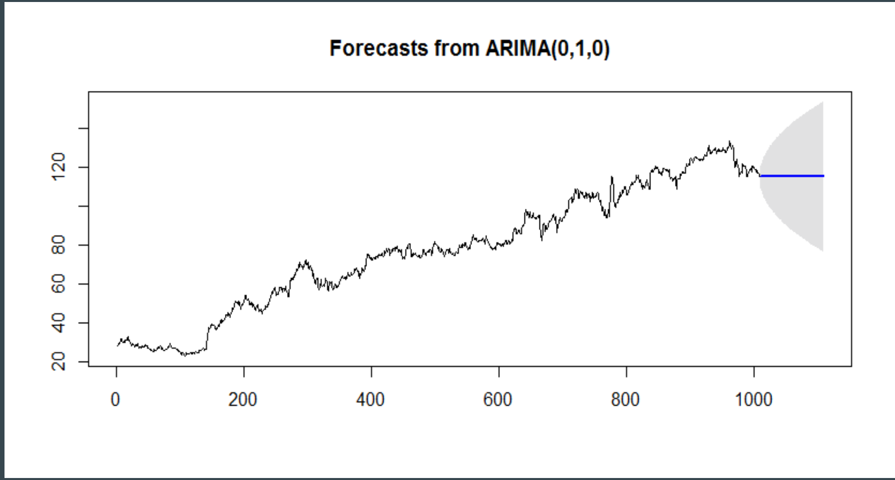
Confidence Interval = 99%

Forecast from ARIMA(0,1,0)

Forecast is not following Trend

Confidence Interval = 99%





Comparing Actual Values VS Forecasted Values

Model 1 (Arima (1,1,4)

```
> forecast(object = model1, h = 100, level = 99)
```

	Point Forecast	Lo 99	Hi 99
1009	115.1329	111.31135	118.9545
1010	115.2830	109.79950	120.7666
1011	115.3582	108.66582	122.0506
1012	115.4434	107.90584	122.9809
1013	115.3631	107.12041	123.6057
1014	115.4388	106.49922	124.3783
1015	115.3674	105.82328	124.9115
1016	115.4347	105.28490	125.5845
1017	115.3713	104.68318	126.0594
1018	115.4310	104.20066	126.6614
1019	115.3747	103.65392	127.0955
1020	115.4278	103.21199	127.6436
1021	115.3777	102.70829	128.0472
1022	115.4249	102.29740	128.5525
1023	115.3805	101.82867	128.9322

Actual Stock Price

```
> FB
```

	FB.Close
2017-01-03	116.86
2017-01-04	118.69
2017-01-05	120.67
2017-01-06	123.41
2017-01-09	124.90
2017-01-10	124.35
2017-01-11	126.09
2017-01-12	126.62
2017-01-13	128.34
2017-01-17	127.87
2017-01-18	127.92
2017-01-19	127.55
2017-01-20	127.04
2017-01-23	128.93
2017-01-24	129.37
2017-01-25	131.48
2017-01-26	132.78

Model 2 (Arima(0,1,0)

```
> forecast(object = model2, h = 100, level = 99)
```

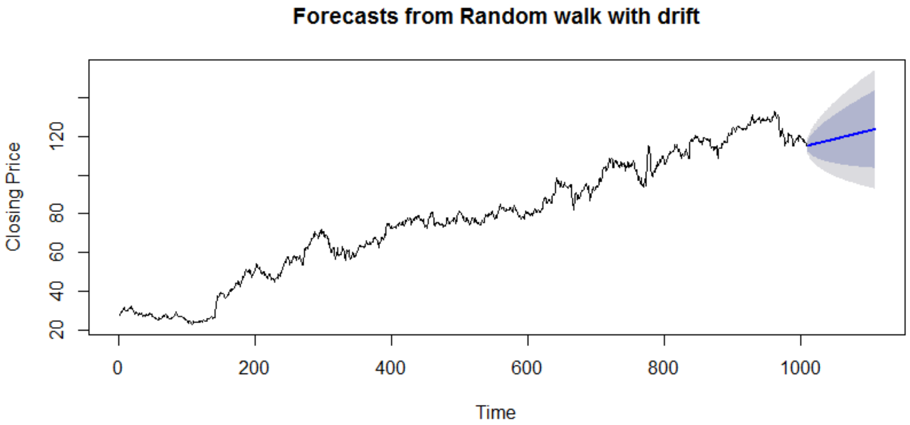
	Point Forecast	Lo 99	Hi 99
1009	115.05	111.19903	118.9010
1010	115.05	109.60390	120.4961
1011	115.05	108.37992	121.7201
1012	115.05	107.34805	122.7520
1013	115.05	106.43896	123.6610
1014	115.05	105.61708	124.4829
1015	115.05	104.86128	125.2387
1016	115.05	104.15780	125.9422
1017	115.05	103.49707	126.6029
1018	115.05	102.87215	127.2279
1019	115.05	102.27776	127.8222
1020	115.05	101.70983	128.3902
1021	115.05	101.16511	128.9349
1022	115.05	100.64097	129.4590
1023	115.05	100.13524	129.9648

Random Walk

- Change in a variable that follows no particular pattern or trend
- Variance and Covariance as a function of time
- Real world scenarios that follow random walk
  - Stock Prices
  - Movements of gaseous atoms
- Stock Market applications of random walk
  - Comparison with traditional methods
- Random walk with drift
  - Quantifying the trend as a function of time

Random Walk with Drift

```
> summary(rwalkfore)

Forecast method: Random walk with drift

Model Information:
Call: rwf(y = close_stock, h = 100, drift = T, level = 99)

Drift: 0.0864  (se 0.0471)
Residual sd: 1.4933

Error measures:
                       ME     RMSE      MAE         MPE     MAPE
Training set 3.520967e-15 1.492542 1.023192 -0.02334237 1.443462
```



$$Y_t = y_{t-1} + e(t) + \theta_o$$

Comparing Actual Values VS Forecasted Values

Model 1 (Arima (0,1,0)) with Drift

Forecasts:

	Point Forecast	Lo 99	Hi 99
1009	115.1364	111.29000	118.9829
1010	115.2229	109.78050	120.6653
1011	115.3093	108.64049	121.9782
1012	115.3958	107.69145	123.1001
1013	115.4822	106.86425	124.1002
1014	115.5687	106.12349	125.0139
1015	115.6551	105.44811	125.8621
1016	115.7416	104.82443	126.6587
1017	115.8280	104.24293	127.4131
1018	115.9145	103.69670	128.1322
1019	116.0009	103.18050	128.8213
1020	116.0873	102.69029	129.4844
1021	116.1738	102.22285	130.1247
1022	116.2602	101.77556	130.7449
1023	116.3467	101.34627	131.3471

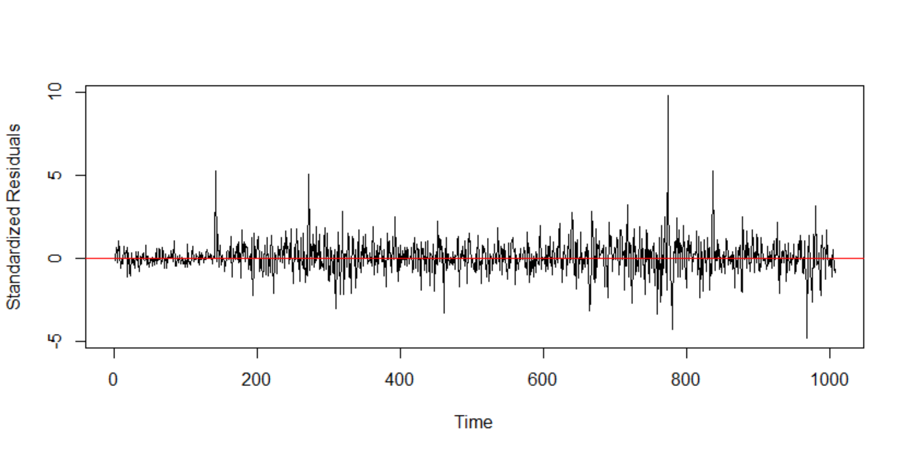
Actual Stock Price

> FB

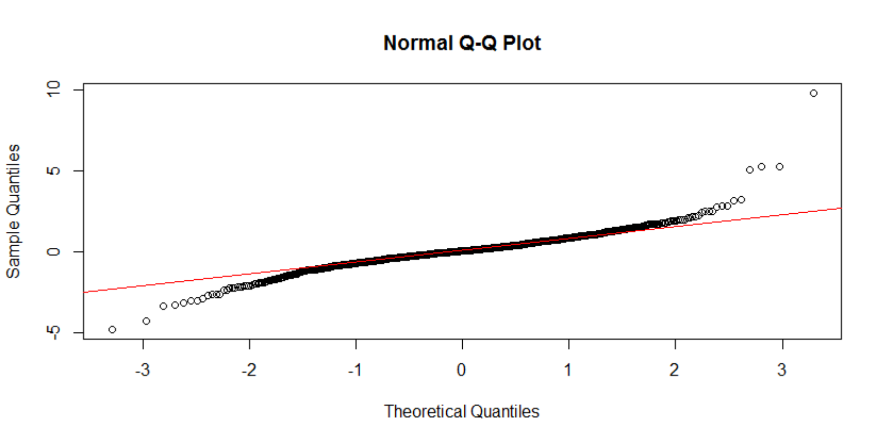
	FB.Close
2017-01-03	116.86
2017-01-04	118.69
2017-01-05	120.67
2017-01-06	123.41
2017-01-09	124.90
2017-01-10	124.35
2017-01-11	126.09
2017-01-12	126.62
2017-01-13	128.34
2017-01-17	127.87
2017-01-18	127.92
2017-01-19	127.55
2017-01-20	127.04
2017-01-23	128.93
2017-01-24	129.37
2017-01-25	131.48
2017-01-26	132.78

Residual Analysis

Plot for Standardized residuals



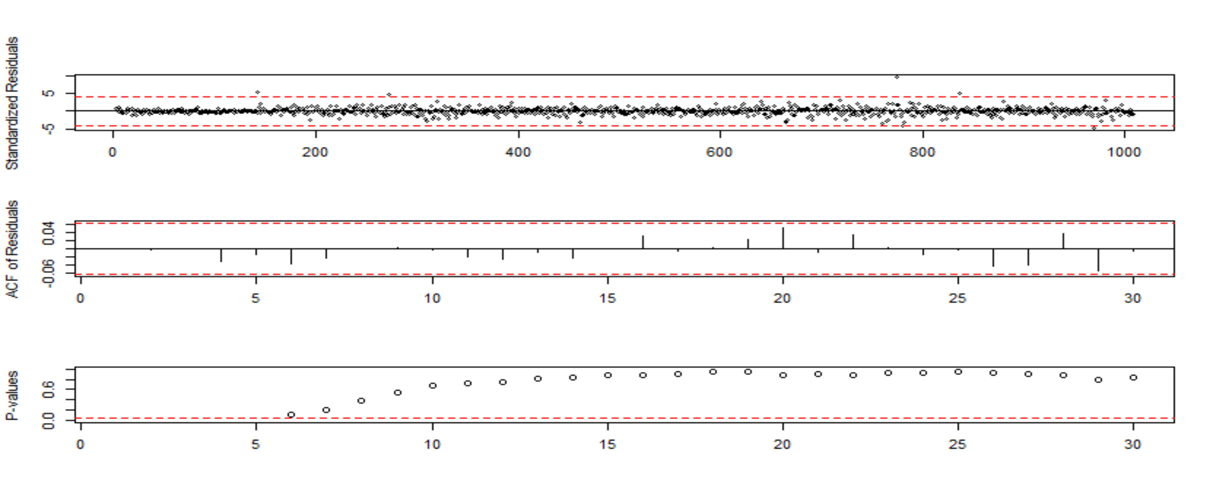
QQ plot for Residuals




```
> ad.test(rwalkfore$residuals)

        Anderson-Darling normality test

data:  rwalkfore$residuals
A = 11.89, p-value < 2.2e-16
```

Residual Analysis

Ljung-Box test



H0: The data are random.
Ha: The data are not random

Conclusion

Best model is Random walk with Drift

Always Stock price is Volatile

Used for proper Timely reaction for Traders for a rough estimate