USING GENERALIZED LINEAR MODEL TO PREDICT STOCK PERFORMANCE

Minchan Shi
12/10/2020

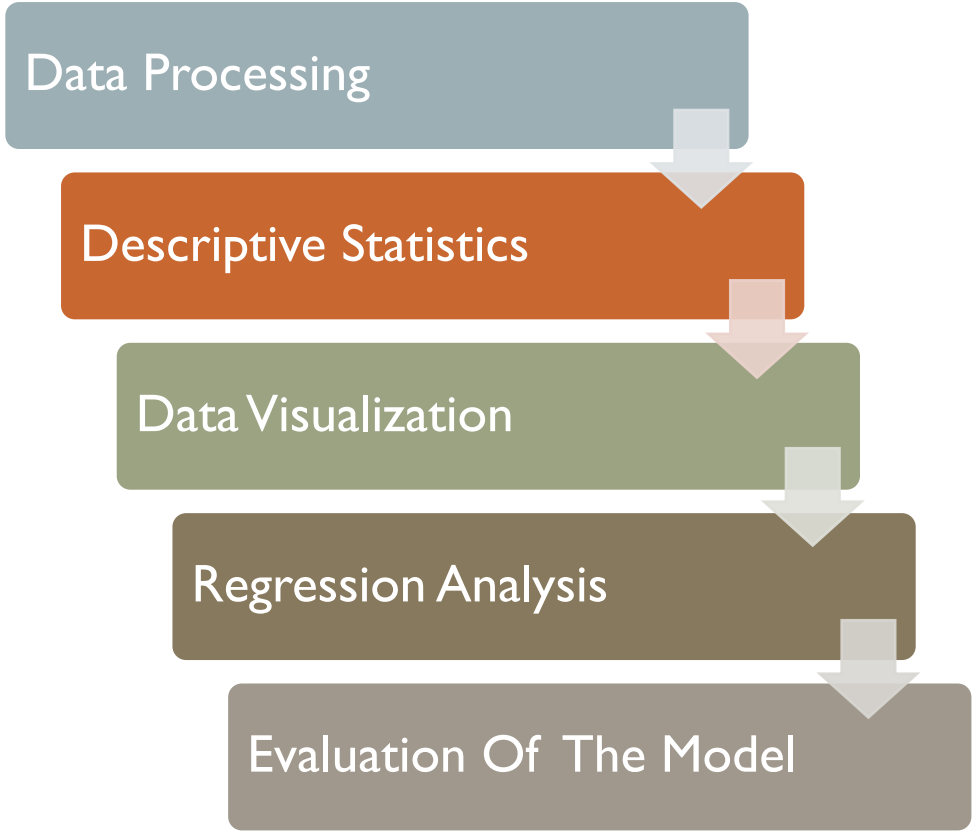
WHAT I WILL BE TALKING ABOUT

- Introduction
- Data Description and Diagnostics
- Data Analysis and Main Results
- Conclusion

PROBLEM STATEMENT & MOTIVATION

PREDICTING &FORECASTING STOCK PRICE(GDX)

METHODS AND DATA DIAGNOSTICS



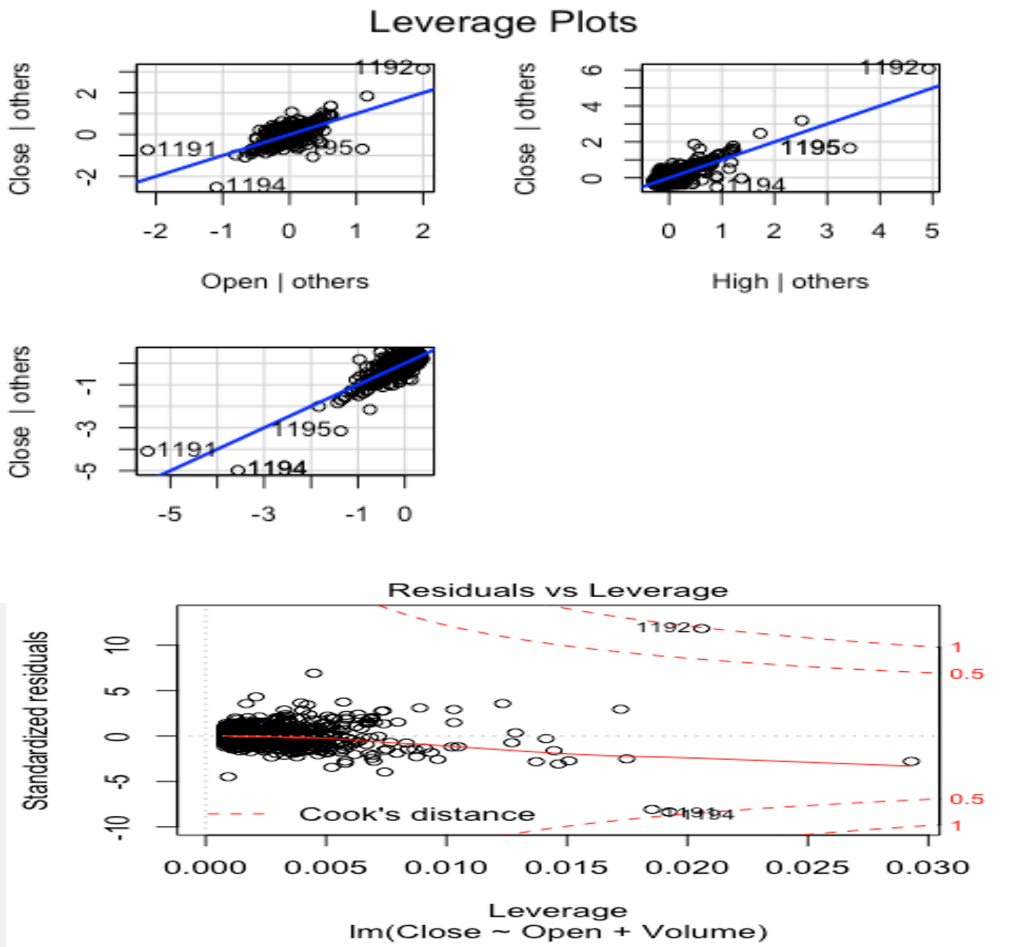
DATA DESCRIPTION

The dataset is from yahoo finance historical database.<"https://finance.yahoo.com/quote/GDX/performance?p=GDX

VanEck Vectors Gold Miners ETF (GDX)

1259 sample size

"Date" ,"Open" ,"High" ,"Low" ,"Close" , "Adj.Close","Volume" ."Nextday", "Profit_In_Percentage","Month" ,"Year" .




```
##        rstudent unadjusted p-value Bonferroni p
## 1192 12.553054          3.9216e-34   4.9334e-31
## 1194 -8.591092          2.5206e-17   3.1709e-14
## 1191 -8.284355          3.0175e-16   3.7960e-13
## 1193  7.047061          3.0077e-12   3.7837e-09
## 1196 -4.514901          6.9303e-06   8.7183e-03
## 1211  4.349618          1.4744e-05   1.8548e-02
```

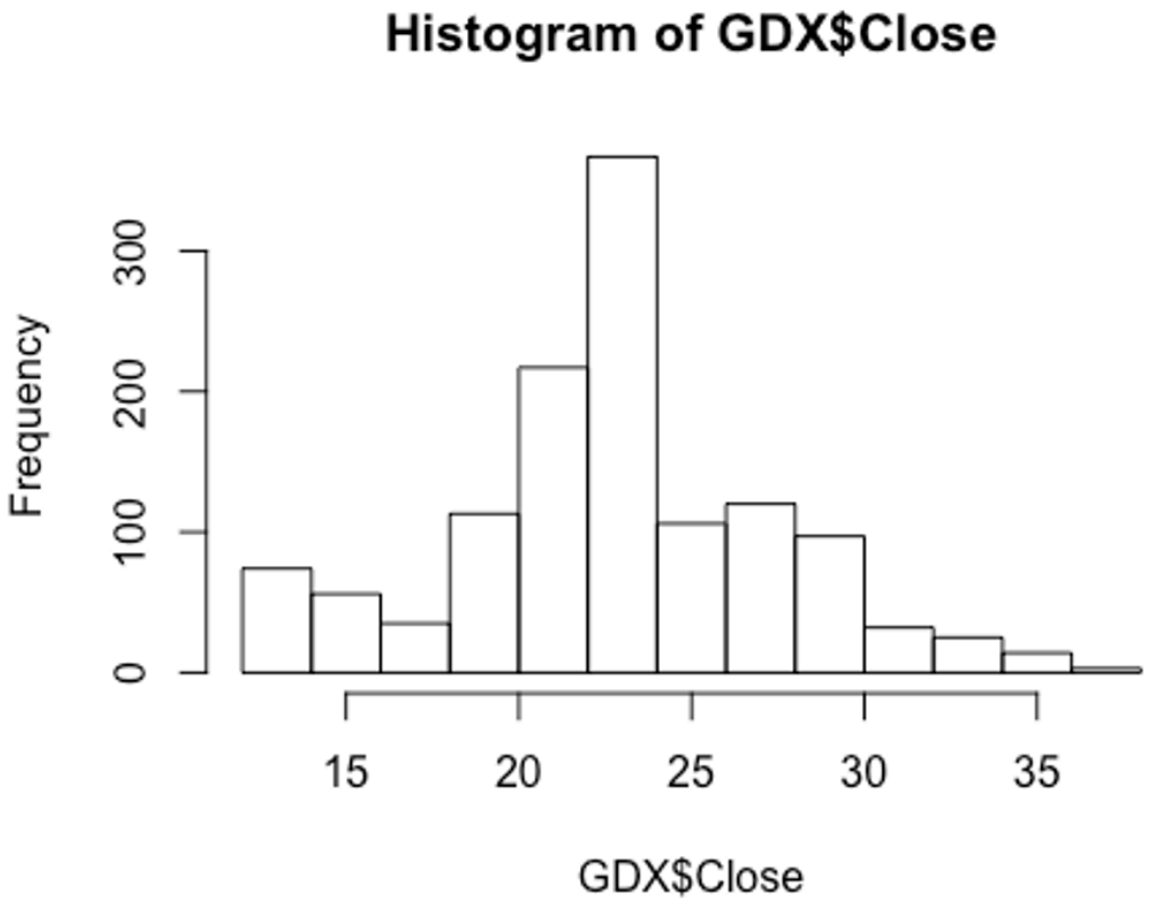
DATA DIAGNOSTICS

OUTLIER APPEARS ONE NA VALUE

KOLMOGOROV-SMIRNOV TEST
SHAPIRO-WILKS TEST

SKEWNESS : 0.07943889
KURTOSIS : 3.213588

CLOSE TO NORMAL DISTRIBUTION



DESCRIPTIVE STATISTICS

$Open

skewness	kurtosis	sd	min	mean	median	max
0.08	3.00	4.66	12.70	22.77	22.59	37.36

$High

skewness	kurtosis	sd	min	mean	median	max
0.11	3.00	4.71	12.92	23.07	22.80	37.49

$Low

skewness	kurtosis	sd	min	mean	median	max
0.03	3.00	4.59	12.40	22.46	22.36	36.45

$Close

skewness	kurtosis	sd	min	mean	median	max
0.08	3.00	4.66	12.47	22.76	22.59	37.21

$Adj.Close

skewness	kurtosis	sd	min	mean	median	max
0.11	3.00	4.71	12.18	22.43	22.24	37.21

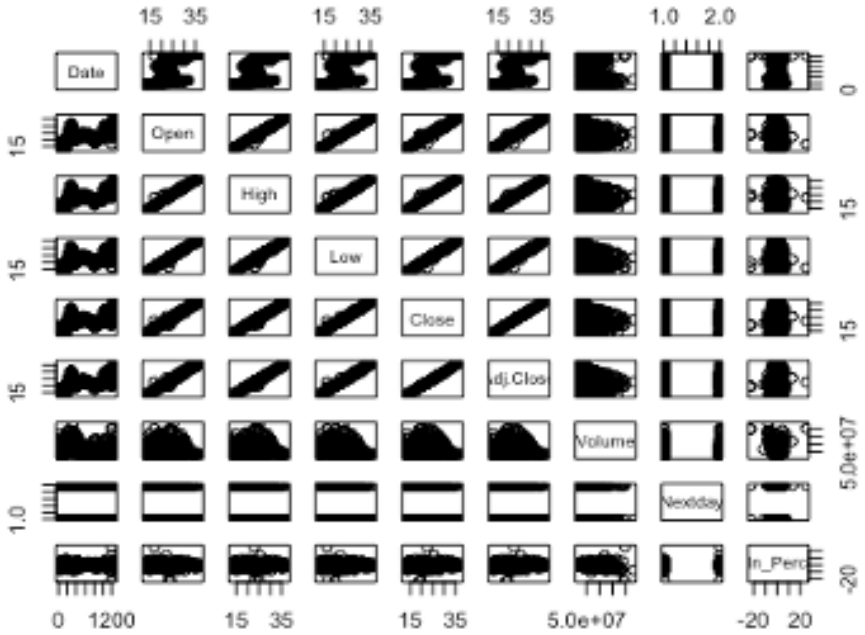
$Volume

skewness	kurtosis	sd	min	mean	median	max
1.68	7.00	29647328.88	13437500.00	55842388.24	48710300.00	232153600.00

$Profit_In_Percentage

skewness	kurtosis	sd	min	mean	median	max
-0.48	31.00	2.29	-22.58	-0.06	0.00	25.52

HIGH CORRELATION



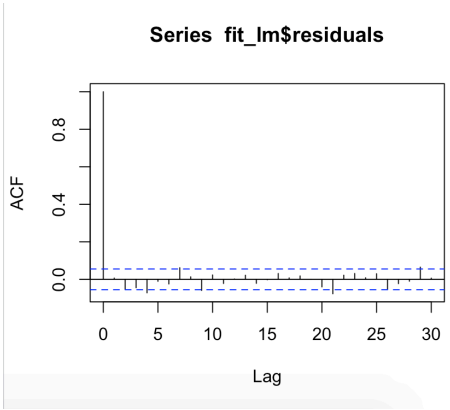
The **Durbin-Watson test** statistic **tests** the **null hypothesis**
that the residuals from an ordinary least-squares regression are not autocorrelated

```
        Durbin-Watson test

data:  fit_lm
DW = 1.9851, p-value = 0.3743
alternative hypothesis: true autocorrelation is greater than 0
```

Null hypothesis: the variance is unchanging in the residuals
Alternative hypothesis: the variance is changing in the residuals

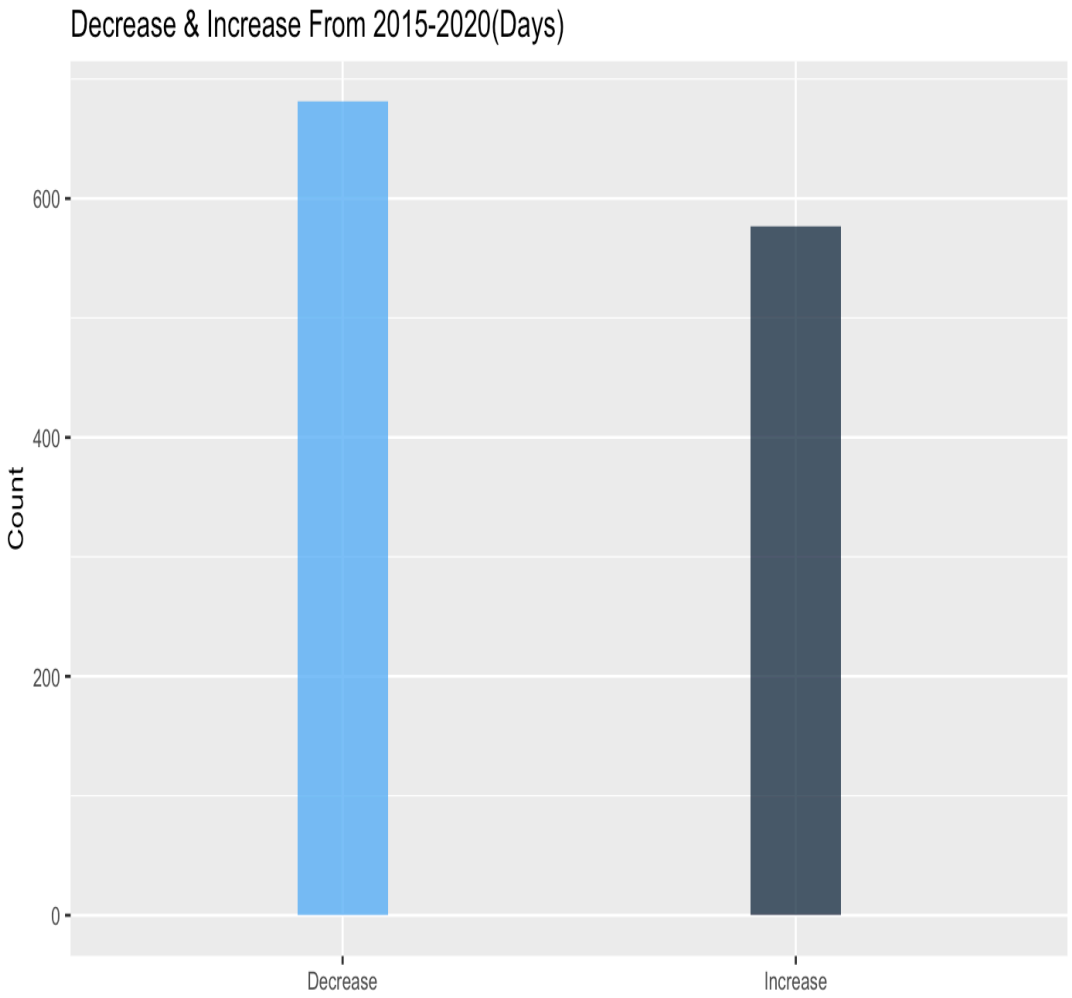
```
        studentized Breusch-Pagan test

data:  fit_lm
BP = 683.19, df = 4, p-value < 2.2e-16
```

Auto-correlation test

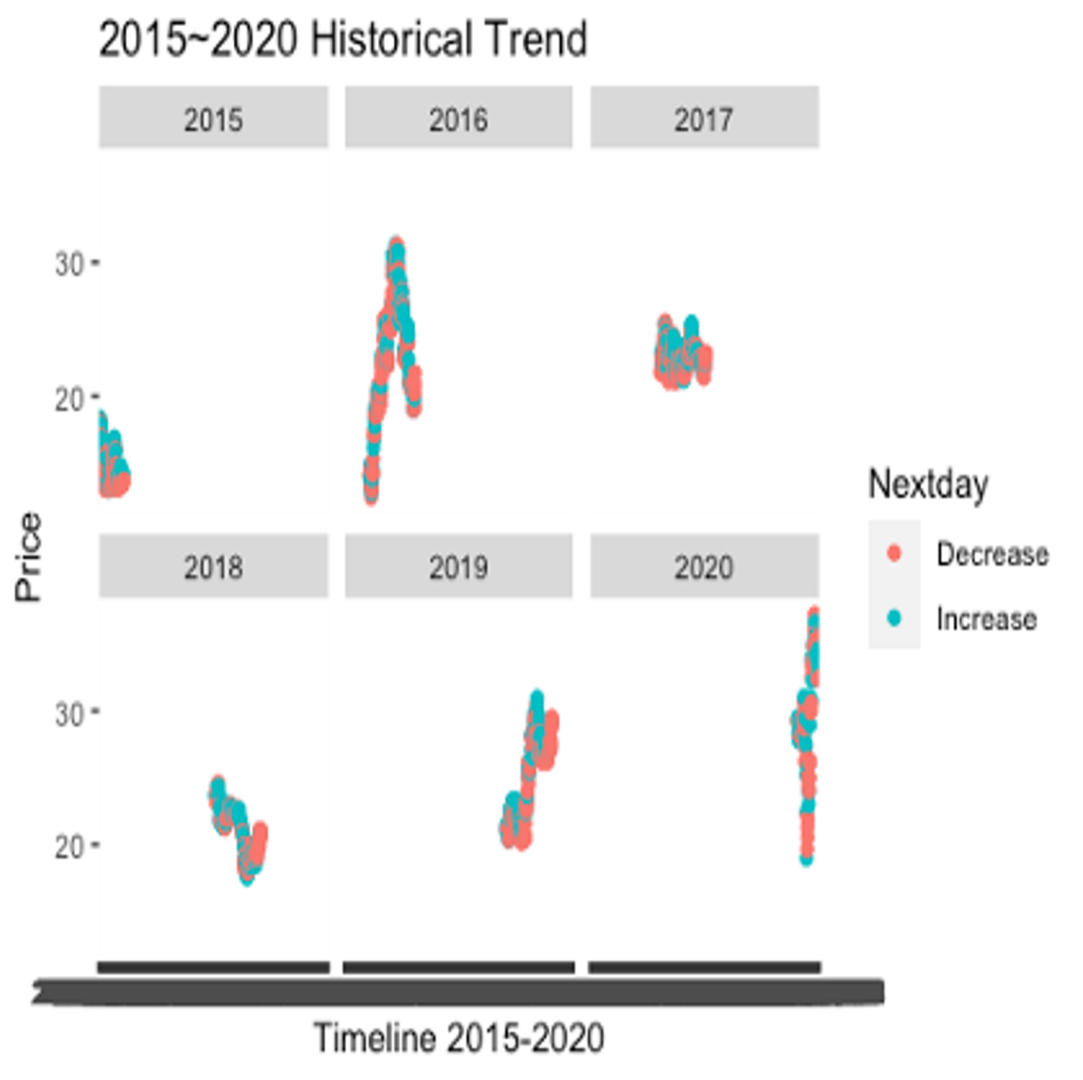


VISUALIZATION

	Nextday	n
1	Decrease	681
2	Increase	577



VISUALIZATION



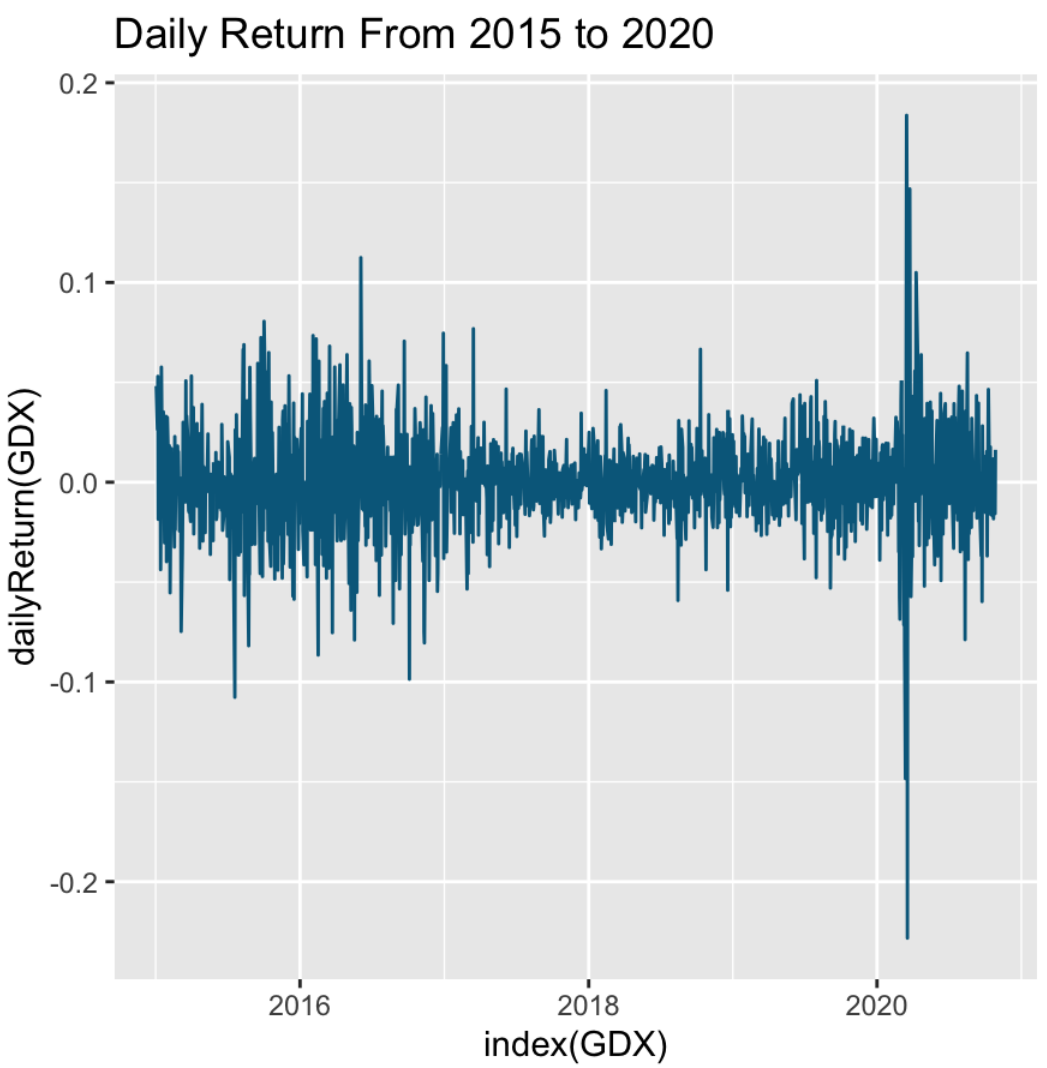
MAY & JUNE & AUGUST PRICE IS ON THE SIDE

NOV & DEC & FEB PRICE IS ON THE LOW SIDE

HOW HAS THE RETURN ON INVESTMENT PERFORMED OVER THE LAST FIVE YEARS?

Earning and returns_Unit %

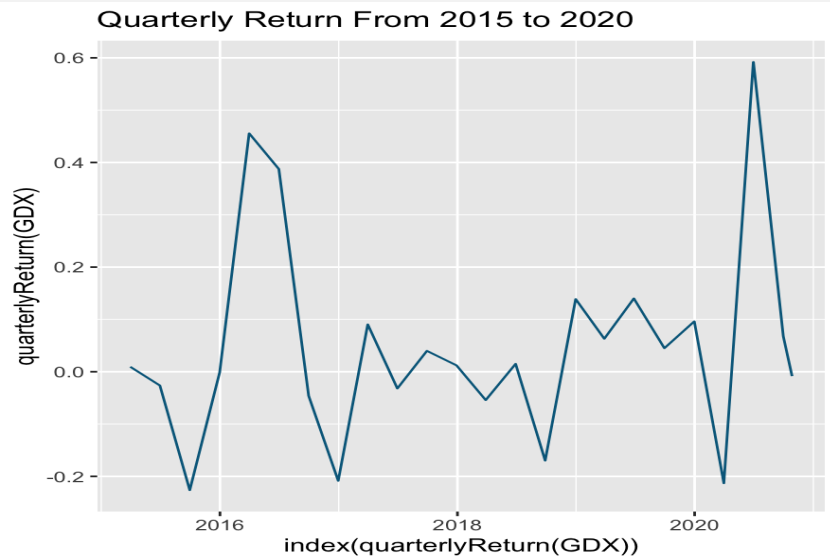
Index	daily.returns
Min. :2015-01-02	Min. :-0.2282353
1st Qu.:2016-06-16	1st Qu.:-0.0118852
Median :2017-11-28	Median : 0.0004482
Mean :2017-11-29	Mean : 0.0008605
3rd Qu.:2019-05-15	3rd Qu.: 0.0137548
Max. :2020-10-27	Max. : 0.1836842

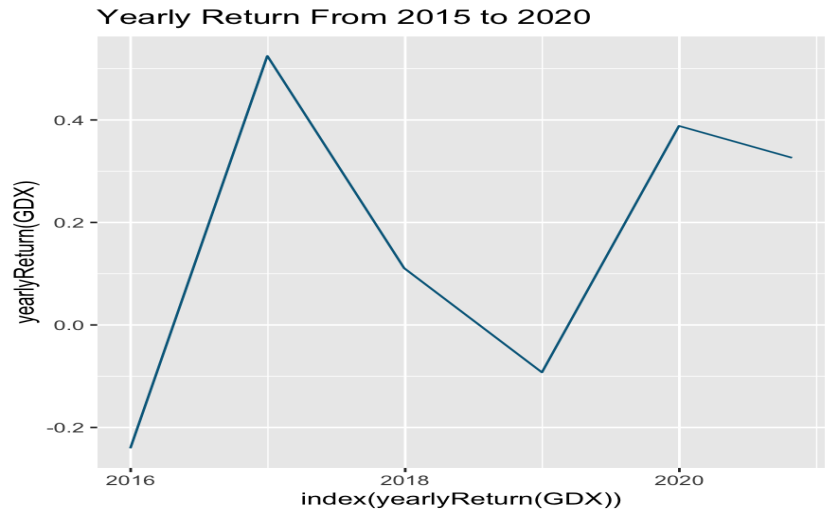


Earning and returns_Unit %

Index	quarterly.returns
Min. :2015-03-31	Min. :-0.22635
1st Qu.:2016-09-07	1st Qu.:-0.03555
Median :2018-02-12	Median : 0.01360
Mean :2018-02-10	Mean : 0.04859
3rd Qu.:2019-07-21	3rd Qu.: 0.09181
Max. :2020-10-27	Max. : 0.59201

	yearly.returns
2015-12-31	-0.24073049
2016-12-30	0.52478134
2017-12-29	0.11089866
2018-12-31	-0.09251291
2019-12-31	0.38833575
2020-10-27	0.32616123





GENERALIZED LINEAR MODEL

Model Result

P-value ***

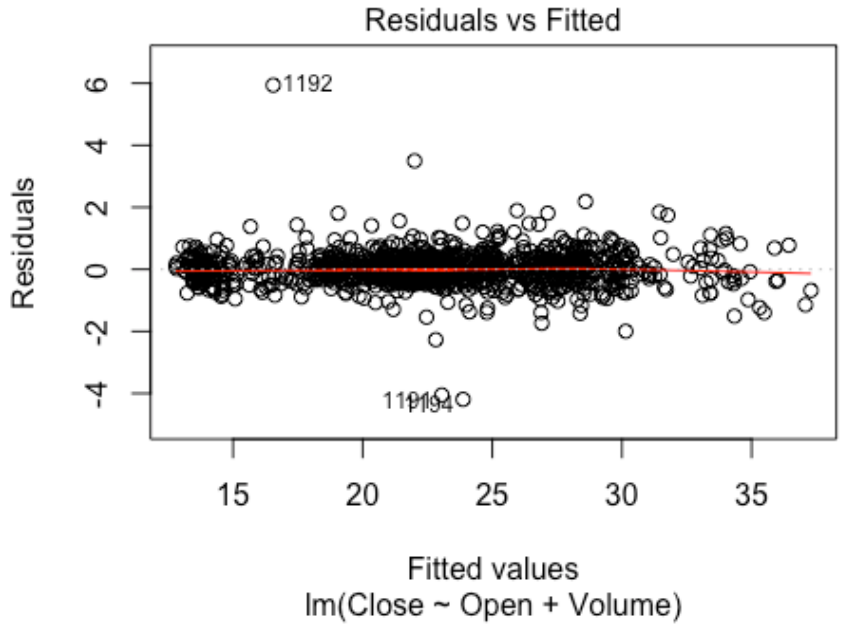
R-squared ~ 0.99

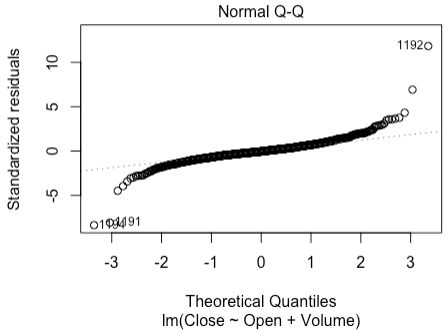
ANOVA

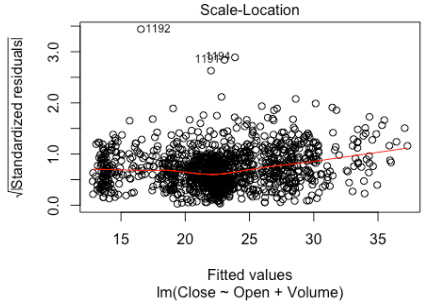
Test result indicates F=3329 ,
Pr(>F) 2.2e-16 ***

BIC/VIF preformed

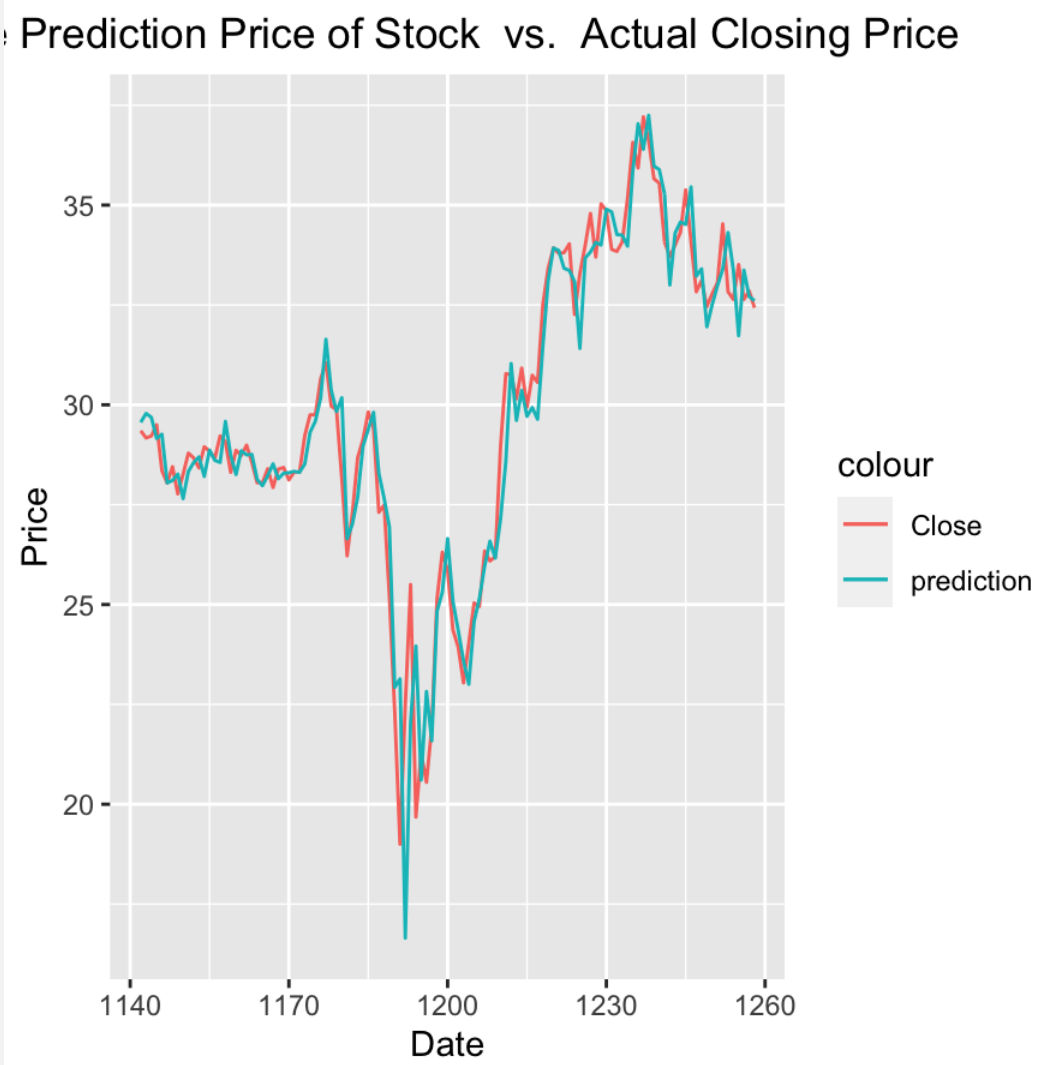
```
Residuals:
    Min      1Q  Median      3Q     Max
-4.1925 -0.2166 -0.0250  0.2080  5.9407

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.358e-01  7.711e-02   3.057 0.002281 **
Open         9.932e-01  3.070e-03 323.481  < 2e-16 ***
Volume      -1.616e-09  4.826e-10  -3.348 0.000837 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5071 on 1256 degrees of freedom
Multiple R-squared:  0.9882,    Adjusted R-squared:  0.9881
F-statistic: 5.242e+04 on 2 and 1256 DF,  p-value: < 2.2e-16
```


Residuals vs Fitted
Residuals
Fitted values
lm(Close ~ Open + Volume)
1192
1194


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(Close ~ Open + Volume)
1192


Scale-Location
√|Standardized residuals|
Fitted values
lm(Close ~ Open + Volume)
1192

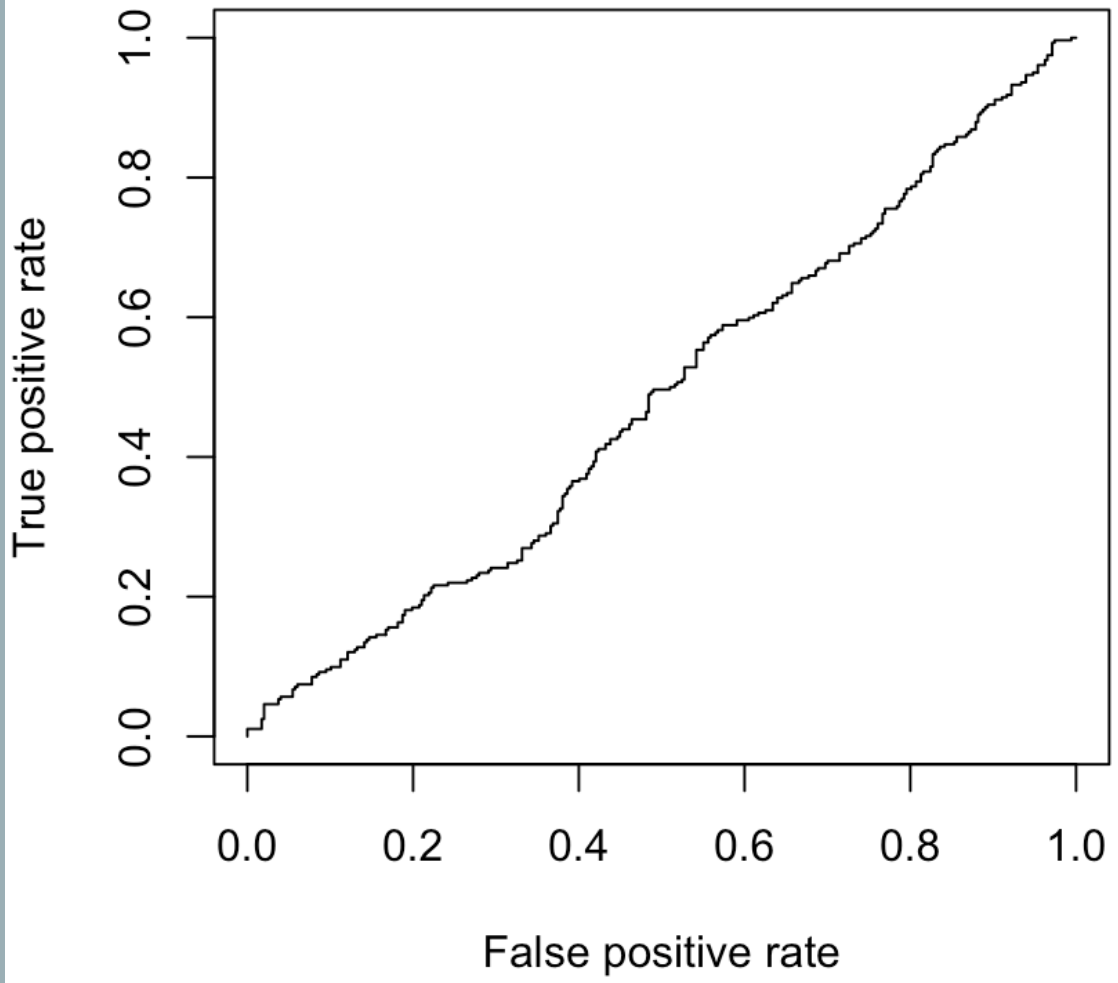
AROUND 25% ERROR



LOGISTIC REGRESSION WAS USED TO PREDICT THE PROBABILITY OF A DAILY RISE OR FALL, AND THE ACCURACY WAS 55%

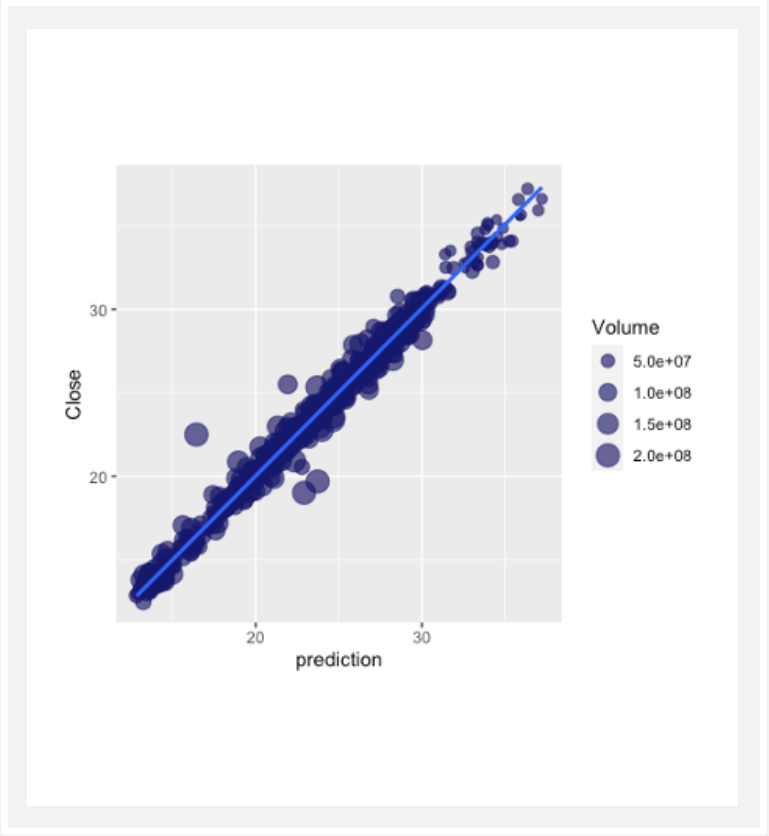
Predict	Decrease	Increase
Decrease	656	547
Increase	25	30

LOGISTIC REGRESSION WAS USED TO PREDICT THE PROBABILITY OF A DAILY RISE OR FALL, AND THE ACCURACY WAS 55%



GENERALIZED LINEAR MODEL

- From Generalized Linear Model we can see that the closing price of a stock can be predicted by establishing a regression model, however, the real value is far away from the model result. Around 25% error rate in the regression model compare with the real closing price. I used the model generated on the training set to predict outcomes for validation set and calculated the test error rate, the percentage of the model misclassified an outcome as compared to the observed results. In general, linear regression analysis does not seem to be suitable for predicting stock prices.
- I use logistic regression to estimate that the model correctly predicted that the market opening price order to go up on 13 days and that it order to go down on 669 days, for a total of 669 + 13 = 682 correct predictions. Among them, the error rate reaches 1-0.5421304=0.4578696, which means the error rate of logistic regression is about 46%. According to the observation, we can know that the Generalized Linear Model is not very accurate in predicting stocks.



CONCLUSION AND FUTURE RESEARCH

Insight gain

- Short-term investment in this stock is not an ideal strategy
- November and December are low buying times

Future Research

- Other creative and effective methods
- Model can be modified
- Study the short selling mechanism

Thank you!

Questions?