Hannah Lee

"I have neither given nor received help (apart from the instructor) to complete this assignment."

Financial Analysis, Planning, and Forecasting Section 1

Professor Naumova

Abstract

This report is an exploration of the development of a multiple linear regression model and its possible business applications for the company, Netflix. The model looks at 20 factors and how those 20 factors affect the decision variable, the stock price, specifically at closing. Theoretically, Netflix will be able to use this model in order to forecast future stock prices with historical data and with forecasted factors. This will enable Netflix to make better informed business decisions including changes to the workforce, taking on a loan, and investments.

Background information

The statistical model developed in this project is meant to predict the company Netflix's future stock prices. The stock price prediction solves the business problem of trying to predict and plan for future performance. The stock price can indicate future business performance because it indicates investors' demand for Netflix, the financial performance of Netflix, and Netflix's potential for future growth. The more the stock price increases, the better it appears to be for the company.

Positive stock performance tends to drive large investors to invest in the company.

To give an overview of the company, Netflix, it is an internet television network company whose business model consists of streaming memberships. Members can watch Netflix's library of content, including movies and TV shows, all over the world from any device with internet access. Netflix's revenue comes from monthly membership fees.

```
> predict(model, newdata=prediction)
        1         2         3         4
364.1156 13498.9134 16120.6540 26492.2624
```

For example, I inputted new data into the variables and ran the linear regression model again. These are four example stock prices (at close) based off of the x's inputted. Of course, these predicted stock prices are not accurate at all, but Netflix can use this model to try and predict future stock prices by inputting forecasted(predicted) factors (ie. revenue, expenses, employees, subscribers). By using this linear regression model, Netflix can also see how theoretical changes in each factor impact predicted stock prices. This information can be useful in business decisions. For example, if Netflix is deciding whether they should conduct a layoff of their employees next month or next year, this model can inform Netflix how that layoff would affect their stock price depending on what quarter they decide to do it in. The same applies to the factor: long-term debt. If Netflix is deciding whether to take out a loan this year, or wait until

next year, or whether to take it out at all, the model will display the affect on Netflix's stock price by increasing or decreasing long-term debt, total liabilities, and a decrease in assets.

Description of Factors

***Categorical Variables: Low = 0, High = 1**

Quarters from Start

This variable indicates the progression of time.

Open

Stock price at market opening.

High

Stock price highest number during market day.

Low

Stock price lowest number during market day.

Close

Stock price at market close.

Volume

Number of shares being traded that day.

Market Strength

Looking at the graph of the S&P 500 performance, if quarter was lower than the annual average, I recorded it as low. If the quarter was higher than the annual average, I recorded it as high. This is a categorical variable because the S&P 500 is a rough estimate of how the entire stock market is doing.

Company Performance

This is a categorical variable. I chose to measure this as low or high depending on if the quarterly stock price was lower or higher than the annual average stock price.

Revenue

Quarterly revenue of Netflix.

Subscribers Worldwide

Quarterly count of worldwide subscribers of Netflix.

App Downloads

Number of app downloads.

Operating Expenses

Operating expenses including rent, utilities, salaries, etc.

Long Term Debt

Debt such as loans with payment period greater than 12 months.

Total Liabilities

Expenses, anything not an asset. Includes debt, wages, payments for items and services, etc.

Total Assets

Property, equipment, revenue, etc.

Cash on hand

The amount of cash that is liquid and can be withdrawn from a bank and used presently.

Shareholder Equity

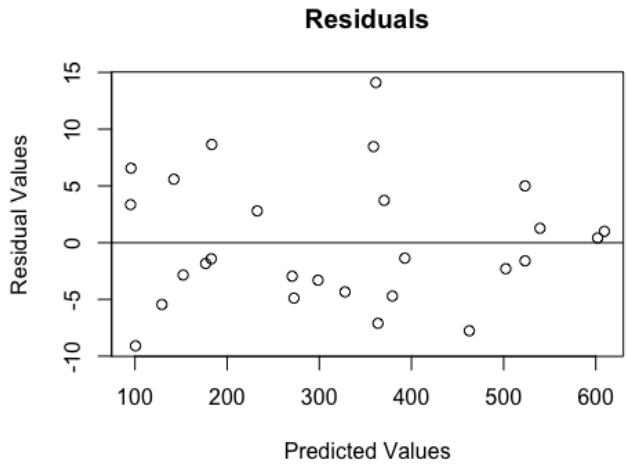
The difference between assets and liabilities. Sum of preferred and common equity.

PE Ratio
Price to Earnings Ratio. Good at valuing stock compared to other company stocks.
EPS
Earnings per share. It is a measure of a company's profit.
Employees
The number of employees annually. A large layoff or increase of employees can have an effect on stock price because it means the company is rapidly expanding or downsizing.

These factors are commonly used when deciding whether to invest in a stock or not. It gives an overview of the company's financial health and performance, as well as the stock performance over time. The ratios help when comparing multiple companies who have different financial metrics when deciding which company is more profitable, or which company has alarming metrics.

Discussion of the assumptions of the model backed up by graphs and analysis

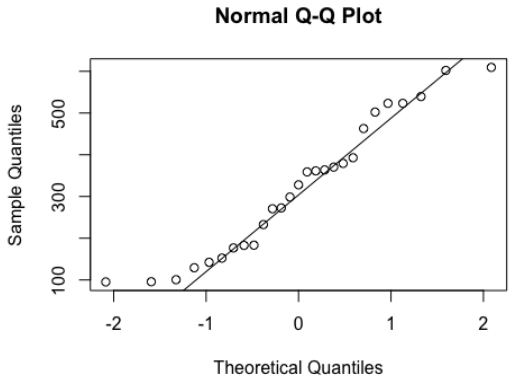
1. Linearity
    a. In the residual plot, about the same number of points lie above and below the x-axis. There are no visible patterns showing dependence of the residuals on the x-values. Thus, the assumptions of linearity, independence, and randomness are not seriously violated. (Lecture 4 Part 2)



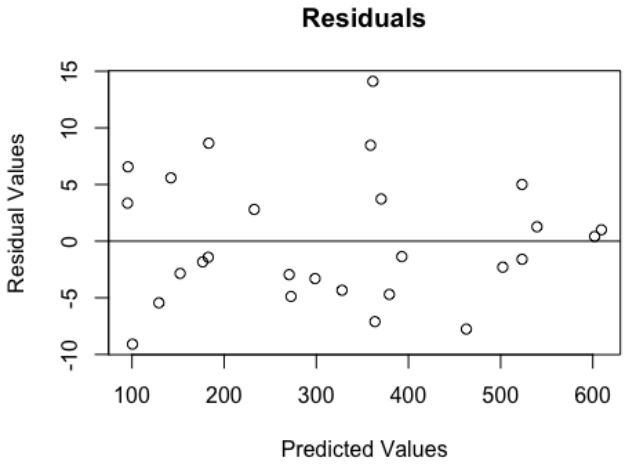
2. Independence and Randomness
    a. For providing independence, when running the model() function in R, none of the variables returned NA values, thus the R function decided that the chosen variables were sufficiently independent to be used in the model.
    b. When choosing factors, I had two variables return NA values and the statistics tests would not run. Upon researching why, the NA returning variables were not sufficiently independent to be used in the model, so I had to try other factors for variables that would not return NA values.
    c. https://zditect.com/blog/10934312.html
    d. “You are getting NA for the last variable because it is linearly dependent on the other 11 variables. R's lm function (and all properly constructed R regression functions as well) will automatically exclude linearly dependent variables for you.”(Blog Post)
3. Zero conditional mean
    a. The residuals are normally distributed around 0. So we can assume that this assumption is true.
4. Homoscedasticity
    a. The residual plot shows the residuals lie in corridors about the same width, so we can say its homoscedastic.

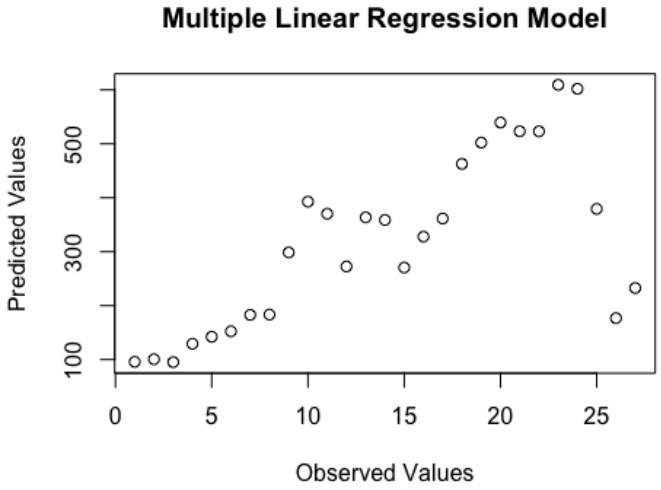
b. Based off of the results of the Bresuch-pagan test and the White's test, we can assume that the model is homoscedastic.

5. Normality



This normal distribution plot shows that the linear model follows a normal distribution because the data points stick close to the fitted line.






```
Coefficients:
          (Intercept)  Quarters.from.Present                  Open                  High
            7.470e+00             3.635e+00            -1.896e-01             9.941e-01
                  Low                Volume         Market.Strength  Company.Performance
            2.840e-01            -6.269e-08             9.584e+00            -6.000e-01
              Revenue  Subscribers.Worldwide          App.Downloads   Operating.Expenses
           -2.976e-08             1.819e-06            -1.183e+00            -2.407e-02
       Long.Term.Debt   Total.Liabilities           Total.Assets          Cash.on.Hand
            3.427e-02             5.155e+00            -5.176e+00            -1.696e-02
  Shareholder.Equity              PE.Ratio                   EPS             Employees
            5.187e+00            -9.203e-03            -5.889e+00            -2.595e-04
```

```
> fitted(model)
       1        2        3        4        5        6        7        8        9
 95.66750 100.57174  95.19612 129.24391 142.22138 152.26007 182.77054 183.31147 298.65485
      10       11       12       13       14       15       16       17       18
392.79325 370.40489 272.54196 363.66578 358.85113 270.57608 327.90446 361.38962 462.80143
      19       20       21       22       23       24       25       26       27
502.32513 539.45993 523.26247 523.20544 609.33801 602.02380 379.29328 176.70555 232.64026
> residuals(model)
         1          2          3          4          5          6          7          8
 6.5625049 -9.0917352  3.3538857 -5.4439087  5.5886147 -2.8500709 -1.4205292  8.6485383
         9         10         11         12         13         14         15         16
-3.3048474 -1.3632548  3.7251130 -4.8819600 -7.1057782  8.4688778 -2.9560847 -4.3344535
        17         18         19         20         21         22         23         24
14.1103814 -7.7614166 -2.2951260  1.2700534 -1.6024995  5.0045854  1.0020151  0.4161985
        25         26         27
-4.7032872 -1.8355553  2.7997394

Residuals:
   Min     1Q Median     3Q    Max
-9.092 -3.820 -1.421  3.539 14.110

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             7.470e+00  4.184e+02   0.018   0.9863
Quarters.from.Present   3.635e+00  1.427e+01   0.255   0.8062
Open                   -1.896e-01  1.956e-01  -0.969   0.3646
High                    9.941e-01  3.922e-01   2.535   0.0390 *
Low                     2.840e-01  3.075e-01   0.924   0.3865
Volume                 -6.269e-08  9.541e-08  -0.657   0.5321
Market.Strength         9.584e+00  9.871e+00   0.971   0.3639
Company.Performance    -6.000e-01  8.655e+00  -0.069   0.9467
Revenue                -2.976e-08  4.793e-08  -0.621   0.5543
Subscribers.Worldwide   1.819e-06  1.266e-06   1.437   0.1940
App.Downloads          -1.183e+00  4.846e+00  -0.244   0.8142
Operating.Expenses     -2.407e-02  2.037e-02  -1.182   0.2759
Long.Term.Debt          3.427e-02  1.650e-02   2.077   0.0764 .
Total.Liabilities       5.155e+00  5.476e+00   0.941   0.3778
Total.Assets           -5.176e+00  5.476e+00  -0.945   0.3760
Cash.on.Hand           -1.696e-02  8.794e-03  -1.929   0.0951 .
Shareholder.Equity      5.187e+00  5.480e+00   0.947   0.3754
PE.Ratio               -9.203e-03  2.626e-01  -0.035   0.9730
EPS                    -5.889e+00  1.193e+01  -0.494   0.6367
Employees              -2.595e-04  1.357e-02  -0.019   0.9853
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.77 on 7 degrees of freedom
Multiple R-squared:  0.9988,    Adjusted R-squared:  0.9956
F-statistic: 308.9 on 19 and 7 DF,  p-value: 1.999e-08
```

```
> durbinWatsonTest(model)
 lag Autocorrelation D-W Statistic p-value
   1      -0.5015319      2.940323    0.54
 Alternative hypothesis: rho != 0
```

H0 = a= 0, no autocorrelation.
H1 = a > 0, autocorrelation exists.
The results of the Durbin-Watson test shows that because the p-value is not less than the significance level 0.05, the residuals are not autocorrelated.

```
> bptest(model)

        studentized Breusch-Pagan test

data:  model
BP = 20.658, df = 19, p-value = 0.356
```

H0: $a_1 = a_2 = \ldots = 0$
H1: At least one value a_i is not 0.
Since the p-value is not less than 0.05, we fail to reject the null hypothesis. We do not have sufficient evidence to say that heteroscedasticity is present in the regression model. We cannot claim conditional heteroscedasticity.

```
White's test results

Null hypothesis: Homoskedasticity of the residuals
Alternative hypothesis: Heteroskedasticity of the residuals
Test Statistic: 1.81
P-value: 0.405257
```

Because Breush-Pagan test was inconclusive, we perform White's test.
H0: $a_1 = a_2 = \ldots = 0$
H1: At least one value a_i is not 0.
Because the p-value is not less than than alpha = 0.05, we fail to reject H0. We do not have sufficient evidence to say that heteroscedasticity is present in the regression model.

Based off these two tests, we can assume that the model is homoscedastic because there is not enough evidence that heteroscedasticity is present.

Recommendations in case of violation of the assumptions

As stated previously, several variables previously chosen to be in the model were not sufficiently independent to provide accurate model results. The recommendation in this situation that I followed was to exclude these linearly dependent variables, and to find new variables that are independent enough to be used in the linear regression model.

If any of the other assumptions are violated, either modify the model until the assumptions are valid, or do a different model approach on another topic.

Statistical analysis (hypothesis testing)

The multiple linear regression model created is:

$$\mathbf{y = 109.2615 - 3.6354x_1 - 0.1896x_2 + 0.9941x_3 + 0.2840x_4 + 0x_5 + 9.5841x_6 - 0.5999x_7 + 0x_8 + 0x_9 - 1.1829x_{10} - 0.0241x_{11} + 0.03427x_{12} + 5.1552x_{13} - 5.1764x_{14} - 0.0169x_{15} + 5.1872x_{16} - 0.0092x_{17} - 5.8889x_{18} - 0.0002x_{19}}$$

"Quarters.from.Start" = x_1
"Open" = x_2
"High" = x_3
"Low" = x_4
"Close" = y
"Volume" = x_5
"Market.Strength" = x_6
"Company.Performance" = x_7
"Revenue" = x_8
"Subscribers.Worldwide" = x_9
"App.Downloads" = x_{10}
"Operating.Expenses" = x_{11}
"Long.Term.Debt" = x_{12}
"Total.Liabilities" = x_{13}
"Total.Assets" = x_{14}
"Cash.on.Hand" = x_{15}
"Shareholder.Equity" = x_{16}
"PE.Ratio" = x_{17}
"EPS" = x_{18}
"Employees" = x_{19}

Adjusted Coefficient of Determination (R-squared) = 0.9956 .
Thus, because the coefficient of determination is close to 1, the fitted model explains 99.56% of the variability in y. (Lecture 4 Part 2). And, the model is a good fit, so the predicted y value is likely to be accurate. The adjusted R squared is used here instead of the normal R squared, because with multiple linear regression and the large number of variables used in the model, the adjusted R squared is more accurate because it adjusts the calculation to the number of variables.

H0: B1 = B2 =...= Bk = 0
H1: Not all Bi are the same. (Lecture 4-5 Notes)

Since the p-value corresponding to the F-statistic = 308.9 is -1.999x10^-8 or 0.00000001999, we reject H0 at significance level alpha = 0.05. There is enough evidence that at least one independent variable x_i is significant in the model. (Lecture 4-5 Notes)

Proper citation and reference list

Yahoo Finance Netflix Stock Performance Data
https://finance.yahoo.com/quote/NFLX/history?period1=1607817600&period2=1670889600&interval=1mo&filter=history&frequency=1mo&includeAdjustedClose=true

Netflix Quarterly Revenue
https://www.statista.com/statistics/273883/netflixs-quarterly-revenue/

Netflix Quarterly Number of Subscribers Worldwide
https://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers-worldwide/

Netflix Quarterly Operating Expenses
https://www.macrotrends.net/stocks/charts/NFLX/netflix/operating-expenses

S&P 500 Annual Performance
https://www.macrotrends.net/2526/sp-500-historical-annual-returns

S&P 500 Monthly Performance
https://www.spglobal.com/spdji/en/indices/equity/sp-500/#overview

Netflix Annual Average Stock Price
https://www.macrotrends.net/stocks/charts/NFLX/netflix/stock-price-history

How much Netflix spent on Content (movies/shows) annually
https://www.statista.com/statistics/964789/netflix-content-spend-worldwide/

Stock Market Metrics
https://www.investopedia.com/financial-edge/0312/5-stock-market-metrics-explained.aspx
https://www.investopedia.com/terms/p/price-earningsratio.asp
https://www.investopedia.com/articles/basics/04/100804.asp
https://www.investopedia.com/investing/why-do-companies-care-about-their-stock-prices/#:~:text=A%20company's%20stock%20price%20reflects,and%20receive%20increases%20in%20compensation.

Netflix SEC Filing - Company Background
https://www.sec.gov/ix?doc=/Archives/edgar/data/1065280/000106528022000036/nflx-20211231.htm

Lecture Notes (Lecture 4-5 mainly)

Appendix