S&P 500 Price Change Predictors

Andy Barnes
Ryan Leveille
Tong Sun
Jared Turner

12 November, 2018

Project Scope

S&P 500

The Standard & Poor's 500, often abbreviated as the S&P 500, or just the S&P, is an American stock market index based on the market capitalizations of 500 large companies having common stock listed on the NYSE or NASDAQ. The NYSE and NASDAQ are the American stock exchanges where the stocks that make up the S&P 500 are traded. Market capitalization (market cap) is the market value of a publicly traded company's outstanding shares. Market capitalization is equal to the share price multiplied by the number of shares outstanding. As outstanding stock is bought and sold in public markets, capitalization could be used as an indicator of public opinion of a company's net worth and is a determining factor in some forms of stock valuation. The reason that we chose to use the S&P 500 for our companies is that it is a very broad and diversified index that includes companies from many different industries. The S&P 500 is one of the most commonly followed equity indices, and many consider it one of the best representations of the U.S. stock market, and a bellwether for the U.S. economy. The share prices for each company within the S&P 500 change daily (during trading days) based on trading interest. A share price is the price of a single share of a company, derivative or other financial asset. The share price is the highest amount someone is willing to pay for the stock, or the lowest amount that it can be bought for.

Time Frame

Our project is going to analyze the companies within the S&P 500 over a 7 year period. This period is going to be from 1/01/2011-12/31/2017.

Approach:

Due to the volatility of the stock market and how the data can sometimes be very misleading depending upon the time frame which you choose to analyze the data from, we are going to approach this project with the intention of minimizing any major outliers for our response variable (share price).

For example, if we were to randomly select some date in 2011 and observe the share price for Microsoft it could be $50 per share. Likewise if we took another random date in 2011 to observe the price of Microsoft it could be $25 per share, perhaps even just a week before or after the price was at $50 per share. Furthermore, when we analyze the response variable in 2017 it could be $50 or $25, depending on the time frame you select for the observation.

Therefore, the way that we will make sure that we do not misrepresent our response variable and by keeping the project within the scope of our class we will be taking the average of the share price for each company within the S&P 500 for the years 2011 and 2017. The 5 years in between (2012, 2013, 2014, 2015, 2016) will allow plenty of time for our response variable to change based on our predictor variables.

Another reason why we want to calculate the average share price for each company by year is because our predictor variables will be yearly as well.

The S&P 500 is based on market cap and some companies will grow in size or fall in size so we will be using companies that were in the S&P 500 starting on January 1, 2011, and are still in the S&P 500 on December 31, 2017. As of now, we have 476 companies out of 500 that are still within the S&P 500 to analyze.

Response Variable

Share Price (Y): Our response variable is going to be the share price of each company within the S&P 500. We are going to take the average share price for 2011 and the average share price for 2017 and then calculate the percent difference for each company.

Predictor Variables

We have multiple variables that we are going to use as our predictor variables. These variables have been chosen based on what what we think investors use to analyze a company before making an investment. Also, some other potential variables' data is incomplete or missing so we choose variables that were best for the scope of the project.

Our goal with this project is to find the predictor variables that have the biggest impact on our response variable (share price).

All of our variables are going to be using their percent change in yearly unit during our time period of 2011-2017. We will start with the predictor variables' 2011 observations and end with the predictor variables' 2017 observations. We chose this time frame because it gave us our yearly data of 2011, a 5 year time frame, then our yearly data of 2017.

1. Earnings per Share: Earnings per share (EPS) is the portion of a company's profit allocated to each share of common stock. Earnings per share serves as an indicator of a company's profitability.
2. Dividend: A dividend is a payment made by a corporation to its shareholders, usually as a distribution of profits. When a corporation earns a profit or surplus, the corporation is able to re-invest the profit in the business and pay a proportion of the profit as a dividend to shareholders.
3. Book Value per Share: Book value per share indicates the book value (or accounting value) of each share of stock. Book value is a company's net asset value, which is calculated by total assets minus intangible assets and liabilities. An easy way to think of book value per share is what is the expected value of the company.
4. Return on Assets: The return on assets shows the percentage of how profitable a company's assets are in generating revenue.
5. Return on Equity: Return on equity is a measure of the profitability of a business in relation to the book value of shareholder equity, also known as net assets or assets minus liabilities. ROE is a measure of how well a company uses investments to generate earnings growth.
6. Return on Invested Capital: Return on capital (ROC), or return on invested capital(ROIC), is a ratio used in finance, valuation and accounting, as a measure of the profitability and value-creating potential of companies after taking into account the amount of initial capital invested.
7. Debt / Equity: Debt/Equity (D/E) Ratio, calculated by dividing a company's total liabilities by its stockholders' equity, is a debt ratio used to measure a company's financial leverage. The D/E ratio indicates how much debt a company is using to finance its assets relative to the value of shareholders' equity.
8. Revenue: Revenue is the income that a business has from its normal business activities, usually from the sale of goods and services to customers.

The Data

Our data looks at the 476 stocks in in the S&P 500 today with data dating back to 2011. Each variable is given as the percent change in yearly value or average value from 2011 to 2017. The first 6 rows of data are as follows

Company	Delta.price	Earnings.Per.Share.USD	Dividends.USD	Book.Value.Per.Share…USD
A	1.0040759	-0.2631579	NA	0.1531023
AAL	5.3661833	-1.2463677	NA	-1.2150259
AAP	0.9557057	0.2563601	0.000000	3.0917603
AAPL	1.8951786	1.3316456	NA	1.1926995
ABC	1.2259627	-0.3543307	2.395349	0.0664845
ABMD	8.9755385	-4.6562500	NA	2.4388489

Return.on.Assets..	Return.on.Equity..	Return.on.Invested.Capital..	Debt.Equity	Revenue.USD.Mil
-0.2187210	-0.4385704	-0.3128872	-0.1777778	-0.3239607
-1.4622991	NA	NA	NA	0.7601651
-0.4973357	-0.6409276	-0.6044682	-0.3673469	0.5192869
-0.4876247	-0.1151908	-0.5160819	NA	1.1176547
-0.7791667	-0.2841845	-0.5969474	3.8823529	0.9090977
-2.1888889	-2.1493213	-2.1049774	NA	3.4059406

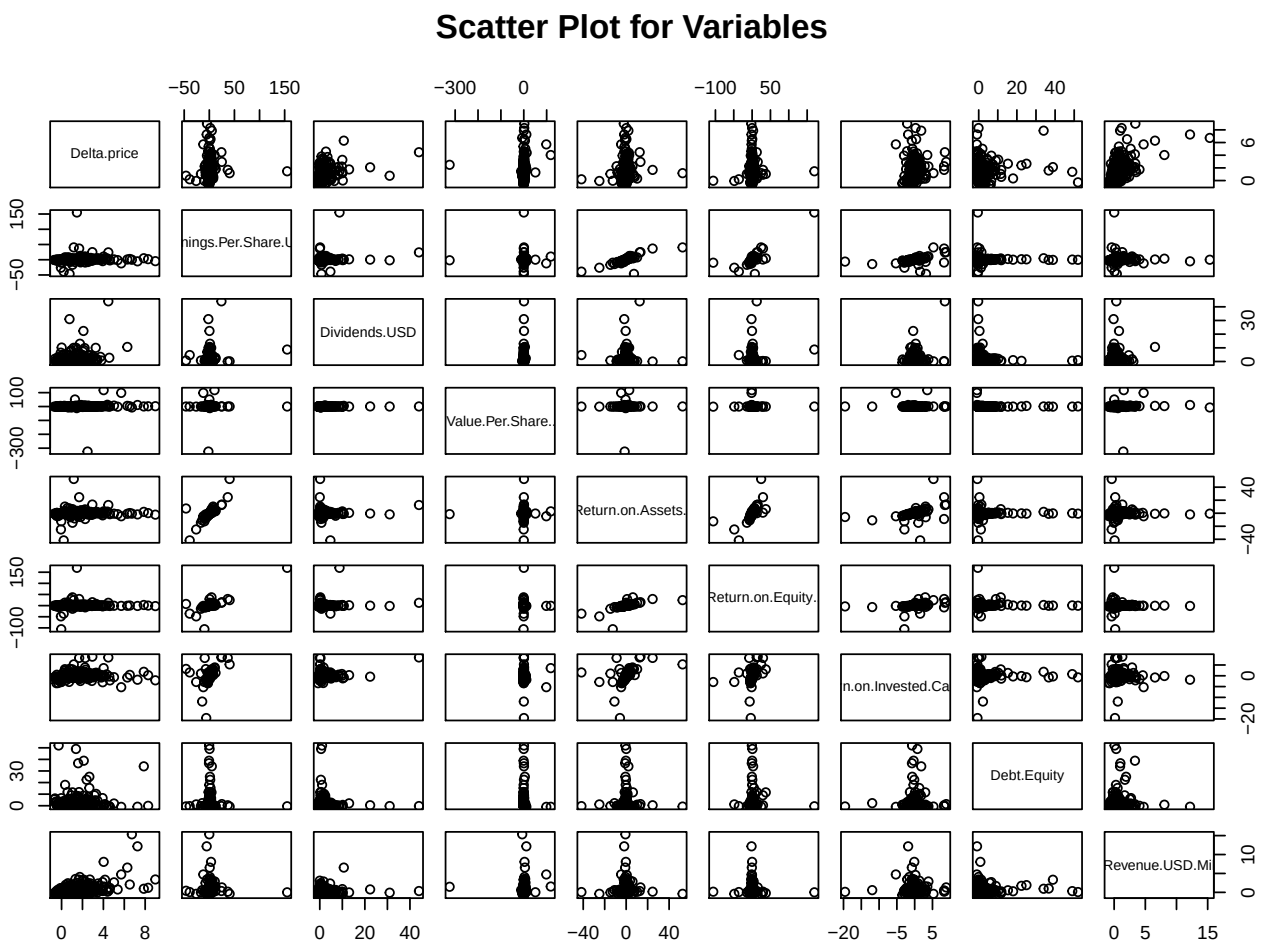
from this we can obtain a correlation matrix

	Delta.price	Earnings.Per.Share.USD	Dividends.USD
Delta.price	1.0000000	0.0825301	0.3127448
Earnings.Per.Share.USD	0.0825301	1.0000000	0.1615642
Dividends.USD	0.3127448	0.1615642	1.0000000
Book.Value.Per.Share…USD	0.0639443	0.0133526	0.0984901
Return.on.Assets..	0.0802322	0.7985285	0.0495701
Return.on.Equity..	0.0678005	0.8158156	0.1154512
Return.on.Invested.Capital..	0.1416728	0.4376127	0.2968282
Debt.Equity	0.1082277	-0.0128430	-0.0467704
Revenue.USD.Mil	0.5405257	0.0047657	0.0773277

	Book.Value.Per.Share…USD	Return.on.Assets..	Return.on.Equity..
Delta.price	0.0639443	0.0802322	0.0678005
Earnings.Per.Share.USD	0.0133526	0.7985285	0.8158156
Dividends.USD	0.0984901	0.0495701	0.1154512
Book.Value.Per.Share…USD	1.0000000	0.0131516	-0.0110597
Return.on.Assets..	0.0131516	1.0000000	0.6330586
Return.on.Equity..	-0.0110597	0.6330586	1.0000000
Return.on.Invested.Capital..	-0.0221520	0.4308342	0.3230085
Debt.Equity	-0.0475424	-0.0118170	0.0281459
Revenue.USD.Mil	0.0474088	-0.0003543	-0.0116261

	Return.on.Invested.Capital..	Debt.Equity	Revenue.USD.Mil
Delta.price	0.1416728	0.1082277	0.5405257
Earnings.Per.Share.USD	0.4376127	-0.0128430	0.0047657
Dividends.USD	0.2968282	-0.0467704	0.0773277
Book.Value.Per.Share…USD	-0.0221520	-0.0475424	0.0474088
Return.on.Assets..	0.4308342	-0.0118170	-0.0003543
Return.on.Equity..	0.3230085	0.0281459	-0.0116261
Return.on.Invested.Capital..	1.0000000	0.0074943	-0.0171131
Debt.Equity	0.0074943	1.0000000	0.0540909
Revenue.USD.Mil	-0.0171131	0.0540909	1.0000000

and scatterplot.



Why we chose this data

We chose this project due to the vast amounts of data that the stock market encompasses. The stock market is a very broad topic and this project can be applied to many different situations.

With our project we hope to find which predictor variables affect the response variable (share price) the most. We are excited to see what our models come up with. The findings will be interesting as there are many different theories as to what is the most important predictor variables for a companies share price.

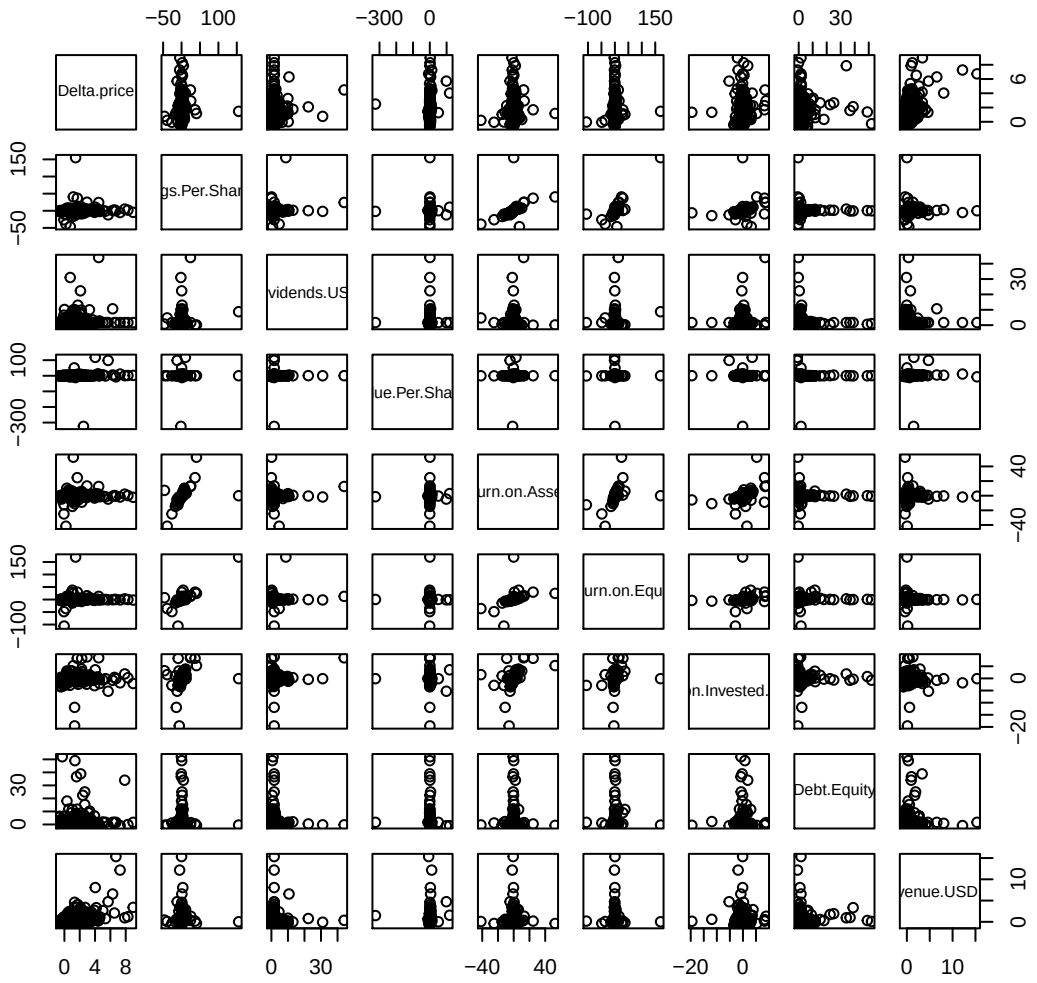
Current Progress

Before we were able to run a multiple linear regression we had to clean our data and replace the `NA`'s. For the predictor variable `Dividends` the `NA`'s were replaced with 0 because this meant that the company (stock) did not have a dividend payment. For the other predictor variables, the `NA` was replaced with means. For this cleaned data we obtain the correlation matrix and scatterplot.

	Delta.price	Earnings.Per.Share.USD	Dividends.USD
Delta.price	1.0000000	0.0818584	0.1866160
Earnings.Per.Share.USD	0.0818584	1.0000000	0.1557679
Dividends.USD	0.1866160	0.1557679	1.0000000
Book.Value.Per.Share…USD	0.0631350	0.0133508	0.0050527
Return.on.Assets..	0.0790602	0.4553726	0.0447495
Return.on.Equity..	0.0642827	0.8152669	0.0973074
Return.on.Invested.Capital..	0.0990169	0.2446333	0.1529477
Debt.Equity	0.0888932	-0.0127845	-0.0348071
Revenue.USD.Mil	0.5313520	0.0047124	0.0334177

	Book.Value.Per.Share…USD	Return.on.Assets..	Return.on.Equity..
Delta.price	0.0631350	0.0790602	0.0642827
Earnings.Per.Share.USD	0.0133508	0.4553726	0.8152669
Dividends.USD	0.0050527	0.0447495	0.0973074
Book.Value.Per.Share…USD	1.0000000	0.0131484	-0.0046222
Return.on.Assets..	0.0131484	1.0000000	0.4084428
Return.on.Equity..	-0.0046222	0.4084428	1.0000000
Return.on.Invested.Capital..	-0.0091943	0.4246314	0.2075853
Debt.Equity	-0.0201462	-0.0116337	0.0245200
Revenue.USD.Mil	0.0463499	-0.0003542	-0.0096492

	Return.on.Invested.Capital..	Debt.Equity	Revenue.USD.Mil
Delta.price	0.0990169	0.0888932	0.5313520
Earnings.Per.Share.USD	0.2446333	-0.0127845	0.0047124
Dividends.USD	0.1529477	-0.0348071	0.0334177
Book.Value.Per.Share…USD	-0.0091943	-0.0201462	0.0463499
Return.on.Assets..	0.4246314	-0.0116337	-0.0003542
Return.on.Equity..	0.2075853	0.0245200	-0.0096492
Return.on.Invested.Capital..	1.0000000	0.0081731	-0.0140577
Debt.Equity	0.0081731	1.0000000	0.0404446
Revenue.USD.Mil	-0.0140577	0.0404446	1.0000000



After this, our initial attempt was to take our data and run a multiple linear regression on it. The summary is as follows.

```
##
## Call:
## lm(formula = Delta.price ~ ., data = clean_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4984 -0.6388 -0.1045  0.3713  6.6183
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   0.888564   0.061799  14.378  < 2e-16 ***
## Earnings.Per.Share.USD        0.001641   0.010075   0.163   0.8707
## Dividends.USD                 0.065480   0.016208   4.040 6.25e-05 ***
## Book.Value.Per.Share...USD    0.003022   0.002900   1.042   0.2978
## Return.on.Assets..            0.011051   0.014615   0.756   0.4500
## Return.on.Equity..            0.002052   0.008244   0.249   0.8035
## Return.on.Invested.Capital.. 0.049108   0.034103   1.440   0.1505
## Debt.Equity                   0.018887   0.009769   1.933   0.0538 .
```

```
## Revenue.USD.Mil              0.536356   0.039114  13.713  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.054 on 467 degrees of freedom
## Multiple R-squared:  0.3267, Adjusted R-squared:  0.3151
## F-statistic: 28.32 on 8 and 467 DF,  p-value: < 2.2e-16
```

We recorded an Rsquared value of 0.3266805 and had multiple predictors that were deemed insignificant by minimizing AIC via R's `step` function. Minimizing AIC will optiminize quality and minimize the complexity. Due to this, we ran a stepwise function to limit our data to just the significant predictors. The significant predictors that we ended up with were `Dividends.USD`, `Return on Invested Capital`, `Debt/Equity`, `Revenue`. We then obtained the following reduced linear model.

```
## Start:  AIC=59.08
## Delta.price ~ Earnings.Per.Share.USD + Dividends.USD + Book.Value.Per.Share...USD +
##     Return.on.Assets.. + Return.on.Equity.. + Return.on.Invested.Capital.. +
##     Debt.Equity + Revenue.USD.Mil
##
##                                Df Sum of Sq    RSS     AIC
## - Earnings.Per.Share.USD        1     0.029 518.93  57.107
## - Return.on.Equity..            1     0.069 518.97  57.143
## - Return.on.Assets..            1     0.635 519.54  57.662
## - Book.Value.Per.Share...USD    1     1.207 520.11  58.186
## <none>                                      518.90  59.080
## - Return.on.Invested.Capital..  1     2.304 521.21  59.189
## - Debt.Equity                   1     4.153 523.06  60.875
## - Dividends.USD                 1    18.134 537.04  73.431
## - Revenue.USD.Mil               1   208.935 727.84 218.140
##
## Step:  AIC=57.11
## Delta.price ~ Dividends.USD + Book.Value.Per.Share...USD + Return.on.Assets.. +
##     Return.on.Equity.. + Return.on.Invested.Capital.. + Debt.Equity +
##     Revenue.USD.Mil
##
##                                Df Sum of Sq    RSS     AIC
## - Return.on.Equity..            1     0.389 519.32  55.464
## - Return.on.Assets..            1     0.722 519.66  55.769
## - Book.Value.Per.Share...USD    1     1.217 520.15  56.222
## <none>                                      518.93  57.107
## - Return.on.Invested.Capital..  1     2.324 521.26  57.234
## - Debt.Equity                   1     4.129 523.06  58.879
## - Dividends.USD                 1    18.618 537.55  71.886
## - Revenue.USD.Mil               1   209.098 728.03 216.266
##
## Step:  AIC=55.46
## Delta.price ~ Dividends.USD + Book.Value.Per.Share...USD + Return.on.Assets.. +
##     Return.on.Invested.Capital.. + Debt.Equity + Revenue.USD.Mil
##
##                                Df Sum of Sq    RSS     AIC
## - Book.Value.Per.Share...USD    1     1.204 520.53  54.566
## - Return.on.Assets..            1     1.339 520.66  54.690
## <none>                                      519.32  55.464
## - Return.on.Invested.Capital..  1     2.379 521.70  55.639
## - Debt.Equity                   1     4.223 523.55  57.319
```

```
## - Dividends.USD                 1    19.203 538.53  70.747
## - Revenue.USD.Mil               1   208.889 728.21 214.384
##
## Step:  AIC=54.57
## Delta.price ~ Dividends.USD + Return.on.Assets.. + Return.on.Invested.Capital.. +
##     Debt.Equity + Revenue.USD.Mil
##
##                                Df Sum of Sq    RSS     AIC
## - Return.on.Assets..            1     1.386 521.91  53.832
## <none>                                      520.53  54.566
## - Return.on.Invested.Capital..  1     2.326 522.85  54.689
## - Debt.Equity                   1     4.129 524.66  56.327
## - Dividends.USD                 1    19.246 539.77  69.849
## - Revenue.USD.Mil               1   210.837 731.36 214.439
##
## Step:  AIC=53.83
## Delta.price ~ Dividends.USD + Return.on.Invested.Capital.. +
##     Debt.Equity + Revenue.USD.Mil
##
##                                Df Sum of Sq    RSS     AIC
## <none>                                      521.91  53.832
## - Debt.Equity                   1     4.045 525.96  55.508
## - Return.on.Invested.Capital..  1     4.990 526.90  56.362
## - Dividends.USD                 1    19.015 540.93  68.866
## - Revenue.USD.Mil               1   211.115 733.03 213.521
##
## Call:
## lm(formula = Delta.price ~ Dividends.USD + Return.on.Invested.Capital.. +
##     Debt.Equity + Revenue.USD.Mil, data = clean_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5586 -0.6462 -0.0981  0.3765  6.6185
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   0.891452   0.061592  14.473  < 2e-16 ***
## Dividends.USD                 0.066261   0.015996   4.142 4.07e-05 ***
## Return.on.Invested.Capital.. 0.065402   0.030820   2.122   0.0344 *
## Debt.Equity                   0.018599   0.009734   1.911   0.0567 .
## Revenue.USD.Mil               0.538397   0.039006  13.803  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.053 on 471 degrees of freedom
## Multiple R-squared:  0.3228, Adjusted R-squared:  0.317
## F-statistic: 56.12 on 4 and 471 DF,  p-value: < 2.2e-16
```

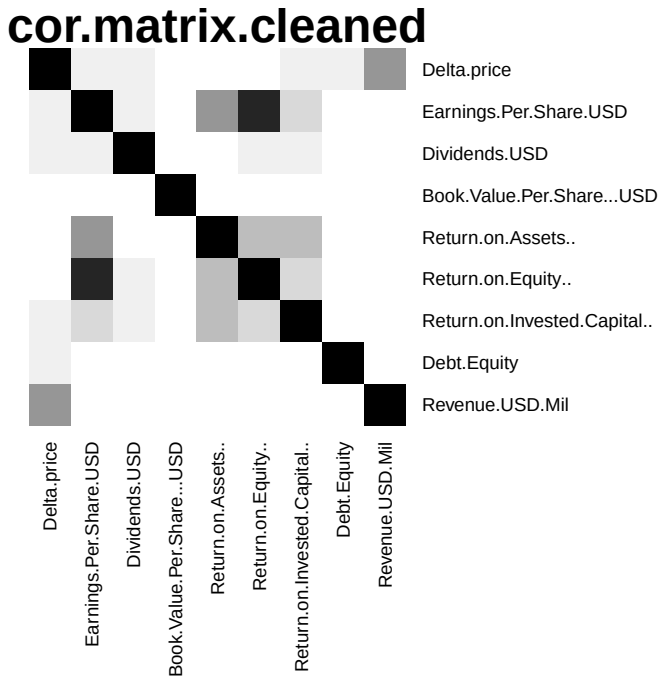
By looking at the correlations between the 4 predictors, they are very independent with each other. We can see this by looking at the correlation matrix and its relevant heat map.

	Dividends.USD	Return.on.Invested.Capital..
Dividends.USD	1.0000000	0.1529477
Return.on.Invested.Capital..	0.1529477	1.0000000

	Dividends.USD	Return.on.Invested.Capital..
Debt.Equity	-0.0348071	0.0081731
Revenue.USD.Mil	0.0334177	-0.0140577

	Debt.Equity	Revenue.USD.Mil
Dividends.USD	-0.0348071	0.0334177
Return.on.Invested.Capital..	0.0081731	-0.0140577
Debt.Equity	1.0000000	0.0404446
Revenue.USD.Mil	0.0404446	1.0000000

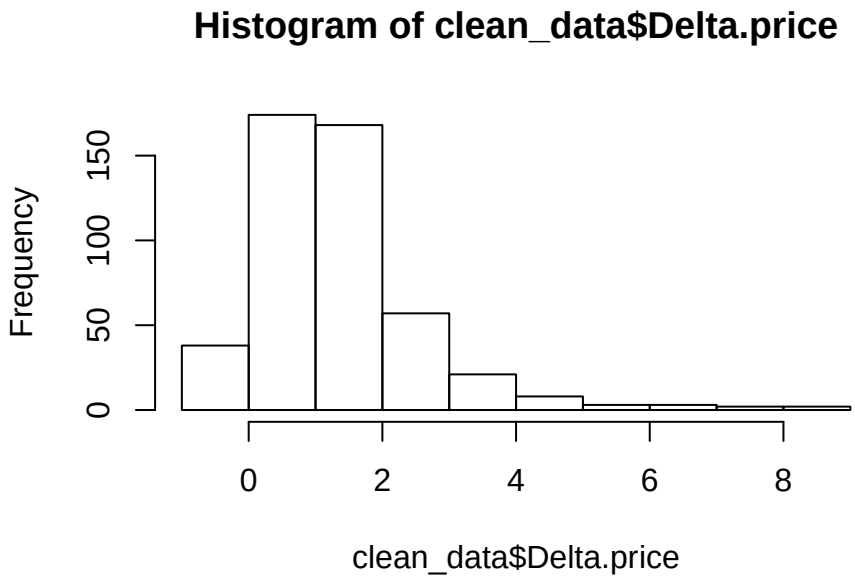
We also provide the heat map of this correlation matrix, with higher correlation corresponding to darker shades.



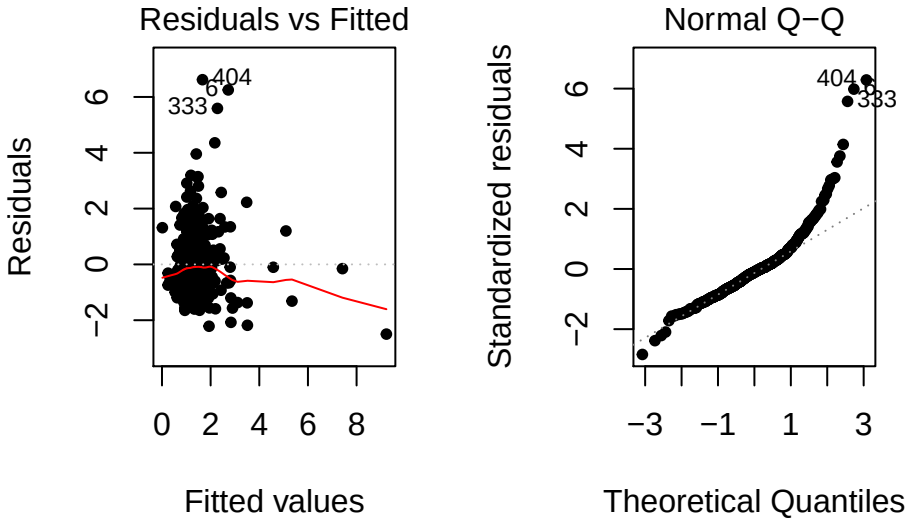
We will to continue to explore our data and see if we can potentially use different predictor variables and/or a polynomial regression to fit our data better. Additionally, we will explore other methods to replace the `NA`'s which could give our model a higher Rsquared. But first, we will see if there are any outliers.

Checking for outliers

In hopes of improving our model, we will set out to find a small subset of outliers we can throw out to drastically improve our model. First, let's take a look at the frequency histogram for our response variable `Delta.price`,



which shows that the values of `Delta.price` are skewed right and do not approximate a normal distribution. We can also inspect two diagnostic plots of our fitted model.

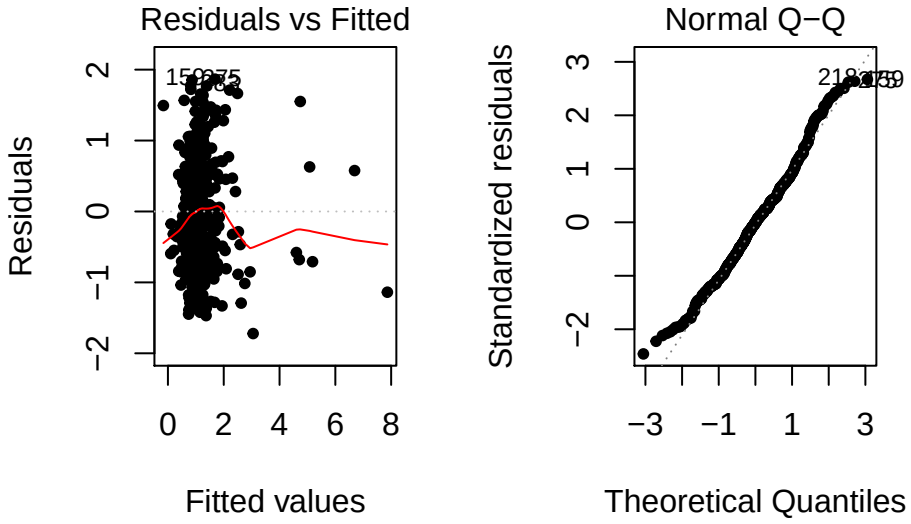


The plot on the right shows the model residuals vs the fitted values. We expect these residuals to be uniformly randomly distributed evenly both above and below mean 0. There are clearly outliers, as numbered. The plot on the right plots the standardized residuals against the theoretical values. This follow the normal line (dotted) closely. Clearly these same outliers , cases 404, 6 and 333, as well as many other values diverge from this line. We will take out these observations as well as others, for a total of 31 observations deleted.

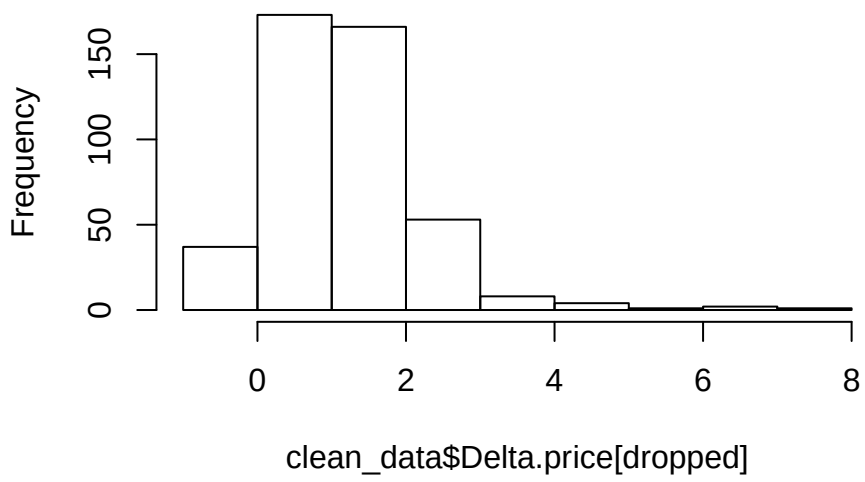
```
##
## Call:
## lm(formula = Delta.price ~ ., data = clean_data, subset = dropped)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.72171 -0.50695 -0.00751  0.45352  1.86579
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)                   0.7359615  0.0430701  17.088  < 2e-16 ***
## Earnings.Per.Share.USD       -0.0002269  0.0068198  -0.033   0.9735
## Dividends.USD                 0.0827521  0.0122492   6.756 4.56e-11 ***
## Book.Value.Per.Share...USD    0.0237253  0.0043952   5.398 1.11e-07 ***
## Return.on.Assets..            0.0112479  0.0098528   1.142   0.2542
## Return.on.Equity..            0.0031051  0.0055666   0.558   0.5773
## Return.on.Invested.Capital..  0.0529112  0.0232752   2.273   0.0235 *
## Debt.Equity                   0.0050248  0.0069370   0.724   0.4692
## Revenue.USD.Mil               0.4664434  0.0270679  17.232  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7091 on 436 degrees of freedom
## Multiple R-squared:  0.5092, Adjusted R-squared:  0.5002
## F-statistic: 56.55 on 8 and 436 DF,  p-value: < 2.2e-16
```

Now, given our R-squared of 0.5092, our response is much better explained by our variables. Furthermore, lookingat same diagnostic plots, we see that the normality of our standardized residuals vs theoretical values is much improved.




```
hist(clean_data$Delta.price[dropped])
```




```
reduced.model.drop <- step(full.model.drop, direction = 'backward')
```

```
## Start:  AIC=-297.08
## Delta.price ~ Earnings.Per.Share.USD + Dividends.USD + Book.Value.Per.Share...USD +
##     Return.on.Assets.. + Return.on.Equity.. + Return.on.Invested.Capital.. +
##     Debt.Equity + Revenue.USD.Mil
##
##                                Df Sum of Sq    RSS      AIC
## - Earnings.Per.Share.USD        1     0.001 219.21 -299.081
## - Return.on.Equity..            1     0.156 219.37 -298.764
## - Debt.Equity                   1     0.264 219.47 -298.547
## - Return.on.Assets..            1     0.655 219.86 -297.754
## <none>                                      219.21 -297.082
## - Return.on.Invested.Capital..  1     2.598 221.81 -293.838
## - Book.Value.Per.Share...USD    1    14.650 233.86 -270.294
## - Dividends.USD                 1    22.947 242.16 -254.780
## - Revenue.USD.Mil               1   149.300 368.51  -67.931
##
## Step:  AIC=-299.08
## Delta.price ~ Dividends.USD + Book.Value.Per.Share...USD + Return.on.Assets.. +
##     Return.on.Equity.. + Return.on.Invested.Capital.. + Debt.Equity +
##     Revenue.USD.Mil
##
##                                Df Sum of Sq    RSS      AIC
## - Debt.Equity                   1     0.266 219.48 -300.541
## - Return.on.Equity..            1     0.356 219.57 -300.358
## - Return.on.Assets..            1     0.675 219.88 -299.713
## <none>                                      219.21 -299.081
## - Return.on.Invested.Capital..  1     2.598 221.81 -295.838
## - Book.Value.Per.Share...USD    1    14.649 233.86 -272.294
## - Dividends.USD                 1    23.420 242.63 -255.910
## - Revenue.USD.Mil               1   149.330 368.54  -69.895
##
## Step:  AIC=-300.54
## Delta.price ~ Dividends.USD + Book.Value.Per.Share...USD + Return.on.Assets.. +
##     Return.on.Equity.. + Return.on.Invested.Capital.. + Revenue.USD.Mil
```

```
##
##                                Df Sum of Sq    RSS      AIC
## - Return.on.Equity..            1     0.378 219.85 -301.776
## - Return.on.Assets..            1     0.651 220.13 -301.223
## <none>                                      219.48 -300.541
## - Return.on.Invested.Capital..  1     2.597 222.07 -297.307
## - Book.Value.Per.Share...USD    1    14.482 233.96 -274.106
## - Dividends.USD                 1    23.276 242.75 -257.687
## - Revenue.USD.Mil               1   150.338 369.81  -70.359
##
## Step:  AIC=-301.78
## Delta.price ~ Dividends.USD + Book.Value.Per.Share...USD + Return.on.Assets.. +
##     Return.on.Invested.Capital.. + Revenue.USD.Mil
##
##                                Df Sum of Sq    RSS      AIC
## <none>                                      219.85 -301.776
## - Return.on.Assets..            1     1.226 221.08 -301.301
## - Return.on.Invested.Capital..  1     2.644 222.50 -298.456
## - Book.Value.Per.Share...USD    1    14.447 234.30 -275.455
## - Dividends.USD                 1    24.038 243.89 -257.601
## - Revenue.USD.Mil               1   150.165 370.02  -72.113
```

```
summary(reduced.model.drop)
```

```
##
## Call:
## lm(formula = Delta.price ~ Dividends.USD + Book.Value.Per.Share...USD +
##     Return.on.Assets.. + Return.on.Invested.Capital.. + Revenue.USD.Mil,
##     data = clean_data, subset = dropped)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.72819 -0.52212 -0.01017  0.44652  1.86475
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   0.743837   0.041516  17.917  < 2e-16 ***
## Dividends.USD                 0.083364   0.012033   6.928 1.53e-11 ***
## Book.Value.Per.Share...USD    0.023526   0.004380   5.371 1.27e-07 ***
## Return.on.Assets..            0.014030   0.008966   1.565    0.118
## Return.on.Invested.Capital.. 0.053346   0.023217   2.298    0.022 *
## Revenue.USD.Mil               0.467103   0.026975  17.316  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7077 on 439 degrees of freedom
## Multiple R-squared:  0.5078, Adjusted R-squared:  0.5022
## F-statistic: 90.58 on 5 and 439 DF,  p-value: < 2.2e-16
```

```
anova(reduced.model.drop)
```

```
## Analysis of Variance Table
##
## Response: Delta.price
##                                Df  Sum Sq Mean Sq  F value    Pr(>F)
```

```
## Dividends.USD                  1  36.087  36.087  72.0576 3.250e-16 ***
## Book.Value.Per.Share...USD     1  34.713  34.713  69.3151 1.076e-15 ***
## Return.on.Assets..             1   3.830   3.830   7.6475  0.005925 **
## Return.on.Invested.Capital..   1   2.015   2.015   4.0231  0.045495 *
## Revenue.USD.Mil                1 150.165 150.165 299.8471 < 2.2e-16 ***
## Residuals                    439 219.853   0.501
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
full_data$Company[-1*dropped]
```

```
##  [1] "ABMD" "NVDA" "STZ"  "ADSK" "HD"   "ALGN" "AAL"  "DAL"  "MAS"  "SBAC"
## [11] "TTWO" "ARNC" "FISV" "AOS"  "ALK"  "EA"   "ANDV" "ATVI" "NFLX" "NOC"
## [21] "ADBE" "CTAS" "V"    "C"    "DXC"  "ALL"  "SHW"  "AIG"  "BSX"  "MHK"
## [31] "IDXX"
```