

Finding dataset suitable for our topic

In this project, our objective is to validate that Value Investing is a decent investment strategy which will bring investors a long-term profit. A Value Investing Strategy demonstrates that value investors should choose undervalued stocks. As a result, we wish to compare if the undervalued stocks will bring investors a relatively larger return.

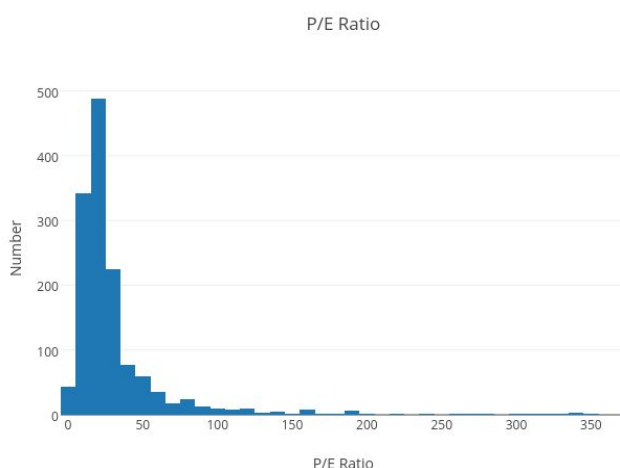
An undervalued stock is believed to be priced too low based on current indicators, such as those used in a valuation model. Should a particular company's stock be valued well below the industry average in terms of fundamental index, it may be considered undervalued. In these instances, value investor may focus on acquiring these investments as a method of pulling in reasonable returns for a lower initial cost. Typically, value investors select stocks with lower price-to-book, price-to-earnings ratios (P/E ratios), higher dividend yields and larger market capitalization etc.. Investors invest if the comparative value is high enough. Also, value investors desire the stocks with a low earning multiple, which is a high earnings yield.

Keeping the objective in mind, our dataset will contain all public companies with their price-to-sales, price-to-book ratios, price-to-earning, Price/Earnings to growth (PEG ratios), dividend yields, industry classification market capitalization and stock exchange place as features.

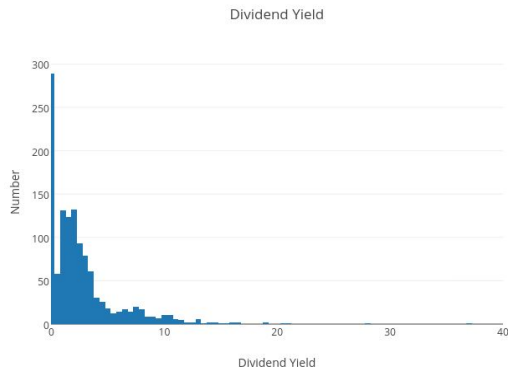
For a primary approach to our objective, we collected all the stocks traded in Nasdaq and AMEX, which sum into around 3700 stocks and we extracted their current corresponding index mentioned above. To measure the stocks performance, we calculated their price level one year ago and today, smoothed by moving average method.

Descriptive Statistics

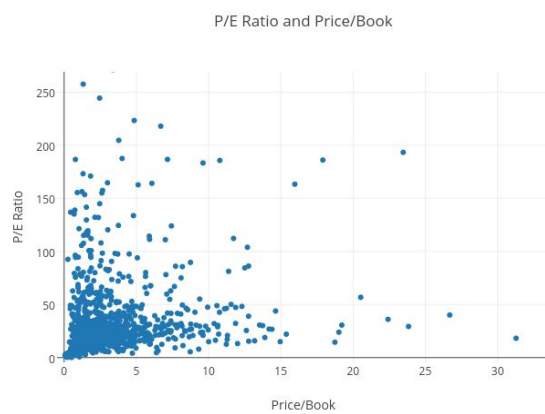
We first decided to have a look at our data:



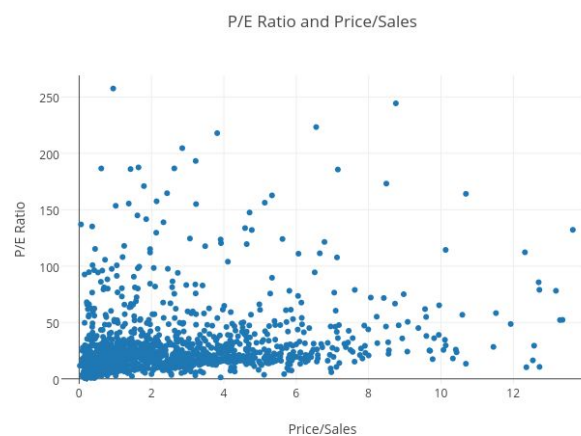
Here we see the range of the price per earnings ratio of the companies we are looking at, one of the key measures in value-based investing.



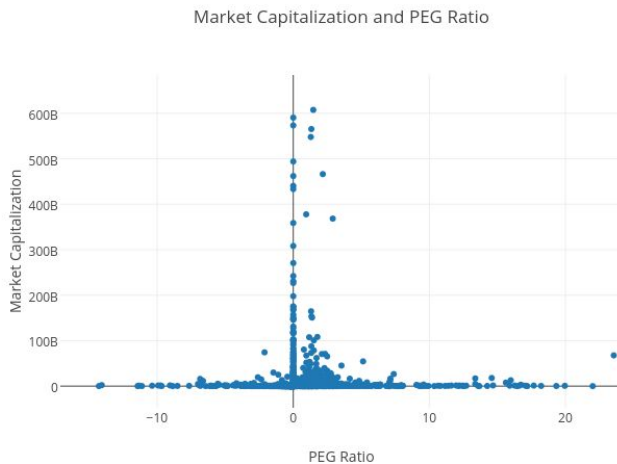
This shows the distribution of the dividend yields for the companies we are looking at, expressed as a percentage of the share price.



This scatter plot shows the P/E ratio and the P/B ratio. A lower P/B ratio could mean that a stock is undervalued, an indicator to a value-investor that it could be time to buy. On visual inspection, it looks like these two could be slightly positively correlated.



Price/Sales reflects how “valuable” each dollar of a company’s revenue may be. This shows a possible slight positive correlation as well.



Market capitalization is the total value of shares for a company, and PEG ratio combines price per earnings with the growth rate over our time period. Interestingly, These two seem to show some sort of gaussian relationship, something that may require further investigation.

Having had a basic understanding on the distribution of the fundamental index of the selected stocks, we decided to dig deeper to understand the relations between them and find out whether the return of a stock is dependent on whether its fundamental indexes are favored by the value investing theorem.

Discovering relation via linear regression

Before starting our regression, we checked and reformed our data: all stocks with entries missing are excluded from the incoming regression. All stocks are grouped by their sectors: Basic Industries, Capital Goods, Consumer Durables, Consumer Non-Durables, Consumer Services, Energy, Finance, Health Care, Miscellaneous, Public Utilities, Technology, or Transportation. Then we calculated the sector average for all those fundamental index listed above. If any stock miss one or more of those index, we will not include that stock in the corresponding index mean calculation, but will include it in other index mean calculation.

We did two linear regressions with respect to our collected dataset.

The first regression is regressing the return rate for the past one year on its Price-Sales, Price-Earning, Price-Book, Price-Earning-Per-Growth ratio, Dividend Yield and Market Capitalization. After excluding stock which misses index value, we used 1421 stocks for this regression.

The result is:

```
Call:
lm(formula = Growth ~ Price.Sales + P.E.Ratio + Price.Book +
    PEG.Ratio + Dividend.Yield + Market.Capitalization, data = Regression)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.51151 -0.04775 -0.00117  0.04251  1.23983
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.658e-02  3.489e-03  10.486 < 2e-16 ***
Price.Sales    4.999e-05  8.710e-05   0.574  0.56612
P.E.Ratio      6.720e-05  2.224e-05   3.022  0.00256 **
Price.Book     -2.620e-07  1.539e-05  -0.017  0.98642
PEG.Ratio      3.002e-07  2.122e-05   0.014  0.98871
Dividend.Yield -4.724e-03  1.321e-03  -3.576  0.00036 ***
Market.Capitalization 3.438e-14  8.017e-14   0.429  0.66814
```

```
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.1038 on 1414 degrees of freedom
Multiple R-squared:  0.0158,    Adjusted R-squared:  0.01162
F-statistic: 3.782 on 6 and 1414 DF,  p-value: 0.0009643
```

As we can see, Price-Sales, PE, PEG ratio and Market Capitalization do contribute to a higher return. To our surprise, Price-Book value has a negative coefficient but the t-value is very small, which means the overall negative coefficient for Price-Book ratio is not a counterexample of value investing theorem. Besides that, this regression is more of a reference than any direct evidence for or against value investing theorem as the fundamental ratios here is their absolute value, not comparative value, as needed in the value investing theorem.

The second regression is regressing the return of each stock based on the difference of their fundamental index and the sector average level. The result we got is:

```
Call:
lm(formula = Growth ~ Price.Sales + P.E.Ratio + Price.Book +
    PEG.Ratio + Dividend.Yield + Market.Capitalization, data = stockindudiff)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.48895 -0.04823 -0.00274  0.04241  1.24197
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.101e-02  3.563e-03   8.702 < 2e-16 ***
Price.Sales    4.171e-05  1.337e-05   3.119  0.001848 **
P.E.Ratio      6.392e-05  2.243e-05   2.849  0.004445 **
Price.Book     7.205e-07  1.525e-05   0.047  0.962335
PEG.Ratio      1.293e-06  2.124e-05   0.061  0.951475
Dividend.Yield -4.309e-03  1.260e-03  -3.419  0.000647 ***
Market.Capitalization 3.967e-14  8.057e-14   0.492  0.622577
```

```
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.1036 on 1414 degrees of freedom
Multiple R-squared:  0.01788,    Adjusted R-squared:  0.01371
F-statistic: 4.289 on 6 and 1414 DF,  p-value: 0.0002711
```

As we can see, the coefficient gets positive in this regression, though its t-value is still small. All of the Price-fundamental coefficients are positive, which is not quite what value investing theory predicted while the dividend yield having a negative effect on return rate. The market capitalization does have a positive effect on stock return (same as the value investing theory predicted). However, overall, the result of this regression cannot serve as a primary evidence of the value investing theorem, at least not in the time period of recent one year.

Further Development

First of all, we missed a very important stock exchange NYSE in our dataset due to some technical problem as very strange errors occurred when our code is extracting data from Yahoo Finance. We definitely need to fix this problem because stocks traded in NYSE represents most of the big successful companies in US and should not be missed if we want a convincing validation. Also, right now, our analysis is generally based on available stocks in the United States. And we wish to prove if the Value Investing Strategy also works in other countries. So, we might collect data in Europe, Asia, and Australia to do the same analysis.

Another variation we are going to try is the analysis within different time frames. We will try to collect the stocks' closing prices in the last 5 years and 10 years. Comparing the results with returns in a short time period (1-year investment). We wish to get some meaningful findings through these variations, because value investing theorem is more often applied to buy-and-hold (long term) strategy and may not be reflected in one year. Trying various time horizon can also generate simply more data to make our conclusion more representative.

Also, we may want to improve our data extraction tool as many values are missed in the results provided by Yahoo Finance. We had to delete almost half of total entries in our regression part because if one index is N/A, the whole entry cannot be used in the regression.

One last issue is on the parameter choice. Currently, we have seven parameters used in the regression, which may prove to be not flexible enough and may lead to underfitting. We should optimize our parameter selection in next part of our work so that we can explain most of the stock return while retaining the essence of the reasoning in value investing theorem.