Regression Analysis of the Equilibrium Valuation Model

Justin Dano

Yamin Tang

Abstract

This paper takes a look at a systematic equity valuation model and performs regression analysis to gain insights to whether or not the model has any predictive characteristics. The work here is a continuation of Boris Wang's paper[1]. His researched laid the ground work for the model and how build the features, but only involved a simple regression model in excel. Our intentions are to understand the underlying factors proposed in his model and see which factors are relevant, the validity of various types of regression, and measure the efficacy of these models. Wang's model is based off the discounted cash flow (DCF) theory which values an asset as a function of the income it generated, the uncertainty of that income, and the risk-free rate of the market. Our analysis follows in sync with Wang's, using a similar dataset of US companies in the S&P 500. We perform outlier analysis followed by the standard linear regression. We follow up with performing various types of regularized regression include Ridge, Lasso, and Elastic Net to determine which features have the highest corresponding importance.

1. Outline

Section 2 will introduce the underlying theory of financial valuation, and the theory behind the model. Section 3 gives a literature review, and section 4 our research thesis. Section 5 dives into the technical details of how the data was obtained and transformed in a way that could emulate Wang's Model. Section 6 dives into the model creation, deploying our original on using regression to build a predictive model. Section 7 covers the results of our models, and conclusion provided in 8. Following the conclusion we have an attached appendix for references and additional graphs.

2. Introduction

Valuation is the process of determining the economic value of a business or company, and understanding how to derive that economic value is the primary business of most financial institutions. By taking a look at the sheer size of the financial industry, we gain insight in just how relevant the ability to understand the worth of something really is. Our analysis looks to extend the current understanding of the valuation models, and more specifically, the equilibrium valuation model. Equilibrium, in the economic sense, is the state of the market where supply and demand are equal[2]. By making this brass assumption and we can construct a framework for understanding how a company can be valued using metrics such as cash flow, EBITDA, and operating expenses.

A large amount of research on valuation has already been done, and since the theory of the equilibrium model is not the subject of this paper, we will only summarize its main parts relevant to the analysis performed. It suffices to say that the model is derived from the discounted cash flow model (DCF Model) which is well known and established in the industry. The underlying economics of each factors is presented in the original paper by Wang.

2.1 Theory

The equilibrium valuation model is a combination of three factors; the income generated by the asset, the probability or risk that the income generated will change in the future, and the risk-free rate of the market.

$$\text{Equity Value} = f\,(\text{Income, Risk, Risk-Free Rate})$$

Income of an asset can be quantifiable in a few ways. The one we use here is the free cash flow (FCF) of the asset, which reflects the amount of cash a company can return to its shareholders after paying any operational expenses. The risk factor can be a function of the expectation or anticipation of how FCF will change in the future (TG). Finally, as a way to calibrate our model with the market, we take into consideration US Treasury Bills to represent the Risk Free Rate (RF) of the model. The result of these three factors give us the Firm Value (FV).

$$FV = f\,(FCF, TG, RF)$$

2.2 Proposed Model

Our model has three main components used to model firm value, FCF, TG, and RF. Each of these components can have varying degree of factors, but will be designed in a way that emulates the original paper by Wang[1]. The terms *n* are measured quarterly.

1) Free Cash Flow (FCF) = f (REV_n, $EBITDA\%_n$, $TAX\%_n$, $CAPEX\%_n$, $\Delta WC\%_n$)

   REV_n = Revenue for period *n*
   $EBITDA\%_n$ = EBITDA margin in period *n*
   $TAX\%_n$ = Tax Margin in period *n*
   $CAPEX\%_n$ = Capital Expenditures margin in period *n*
   $\Delta WC\%_n$ = Change in Working Capital margin in period *n*

2) Expected FCF Growth Rate (TG) = f (RG_n, $EG\%G_n$, $CG\%G_n$)

   RG_n = % Revenue growth between *n* and *n – 1* periods
   $EE\%G_n$ = % EBITDA expense margin growth between *n* and *n – 1* periods
   $C\%G_n$ = % Capital expenditure margin growth between the *n* and *n – 1* periods

3) Risk Free Rate (RF) = f ($3M_n$, $10Y_n$)

   $3M_n$ = 3 Month US Treasury yield in period *n*
   $10Y_n$ = 10 Year US Treasury yield in period *n*

The three components in combination form the ten factors of the equilibrium valuation model. The way the factors are chosen is much of an art form. This is related to the bias-variance trade-off that needs to be taken into consideration. Our work here in choosing the factors are to replicate the work of the original model by Wang. The complete factor model is:

$$FV_n = f(REV_n, EBITDA\%_n, TAX\%_n, CAPEX\%_n, \Delta WC\%_n, RG_n, EE\%G_n, C\%G_n, 3M_n, 10Y_n)$$

Firm Value (FV) itself is derived by a combination of factors including non-operating assets (which itself is part of derived calculations), and market capitalization. We include the formula below:

$$FV_n = NOA_n - MC_n$$

NOA_n = Non-operating assets in period *n*

MC_n = Market capitalization in period *n*

Here is a summary of the main fields we use in the model and their relation the generic Equilibrium model discussed in previous sections.

Field	Model Variable	Equilibrium Component
Revenue	REV_n	Income
EBITDA Margin	$EBITDA\%_n$	Income
TAX Margin	$TAX\%_n$	Income
CAPEX	$CAPEX\%_n$	Income
Change in Working Capital	$\Delta WC\%_n$	Income
Revenue Growth	RG_n	Risk
EBITDA Growth	$EE\%G_n$	Risk
Capital Expenditure Growth	$C\%G_n$	Risk
3 Month T Bills	$3M_n$	Risk-Free Rate
10 Year T Bills	$10Y_n$	Risk-Free Rate
Non-operating Assets	NOA_n	Value
Market Capitalization	MC_n	Value
Firm Value	FV_n	Value

3. Literature Review

This paper is largely a continuation of the research done by Boris Wang in his paper Equilibrium Valuation Model: A Regression-Based Fundamental Equity Valuation Model**.** The main reason we chose to follow Wang's design was because it was freely available, with intuitive research and diagrams that made it reproducible. The one area posed for improvement in his work is the explanation of how he designed the regression, and details of how the regression performed. Wang's model makes lots of assumptions, which are present in our work as well. These assumptions are:

1. It's a competitive market
2. No transaction costs
3. No restrictions on short-selling or borrowing
4. Investors seek to maximize economic returns
5. Investors are rational and risk-neutral
6. All assets are perfectly liquid
7. All information is available to all market participants simultaneously

4. Research Hypothesis

Our main objective in this paper is to build the equilibrium model presented by Wang and do a comprehensive analysis on the regression coefficients presented in the original paper. Our paper aims to answer two main questions:

1. How relevant are the proposed model features in determining a firm's value?
2. Do these features provide useful insights through linear regression, and other types of regression?

5. Methods of Data Collection and Transformation

Our data pipeline consists of three stages. The first stage involves the raw data collection from Bloomberg to our local computers. The second stage involves a basic transformation and cleaning to extract only the relevant data required for the model. The third and final stage is for creating the model features to be used for analysis. The data cleaning operation performed in this section is done in Python (while our analysis in section 6 is done in R).

5.1 Data Collection

Two sources of data were required for this analysis, the first being fundamental data from large companies, and the second being US treasury yields from the US. The first was sourced from Bloomberg, while the yields were downloaded from the St. Louis Fed website[3]. The three financial statements of nine blue-chip stocks were collected including data spanning 38 years from as early as 1st quarter of 1990 (or the earliest available data), to the 3rd quarter of 2018, with a term structure of quarter frequency. This provides approximately 120 different points in time where the metrics of each company is analysed. Data obtained from Bloomberg[4] had a constraint of 40 periods per download. The first sheet contains data from 1990-1998, the second 1998-2008, and final period is 2008-2018. The first transformation involves aggregating each of these three files into one dataset (for each company). For the equity data, Bloomberg provided excel files, each with five sheets pertaining to one of the main financial statements. Our first transformation was to extract the relevant columns and combine all data for one company in different .csv files for easy processing.

Original Excel Sheet Name	**CSV file created**
Income – Adjusted	income_statement.csv
Bal Sheet - Standardized	balance_sheet.csv
Cash Flow - Standardized	cash_flow.csv
Per Share	shares.csv
Stock Value	stock_values.csv

5.2 Data Cleaning Calculations

These calculations are required as the original research paper had them already extracted. These additional calculations were not included in the original paper, but needed to be done to match the data proposed in the model.

1. Operating Expenses = Cost of Revenue + Operating Expenses

2. Working Capital = Total Current Assets - Total Current Liabilities

3. Cash & Investments = Cash, Cash Equivalents & STI + LT Investments & Receivables

5. Debt = ST Debt + LT Debt

Each field also has an explicit variable name. These variable, while not seen when viewing the excel file, are necessary when extracting the data from a software program. The fields are included in the appendix.

Type of File	Bloomberg File – Metric	New CSV Header
Income Statement	Revenue	REVENUE
Income Statement	Cost of Revenue	OPEX
Income Statement	Operating Expenses	OPEX
Income Statement	EBITDA	EBITDA
Income Statement	Income Tax Expense	TAX_EXPENSE
Balance Sheet	Total Current Assets	CHNG_WC
Balance Sheet	Total Current Liabilities	CHNG_WC
Balance Sheet	Cash, Cash Equivalents & STI	CASH_INVESTMENTS
Balance Sheet	LT Investments & Receivables	CASH_INVESTMENTS
Balance Sheet	ST Debt	DEBT
Balance Sheet	LT Debt	DEBT
Balance Sheet	Minority/Non-Controlling Interest	NON_CON_INT
Balance Sheet	Preferred Equity and Hybrid Capital	PREF_SEC
Cash Flow Statement	Change in Fixed & Intang	FREE_CASH_FLOW
Shares	Diluted Weighted Avg Shares	WADS
Stock Value	Last Price	PRICE
3month tbill	3 Month Treasury	TB3MS
10yr_tbills	10 Year Treasury	DGS10

5.3 Creating Model Variables

These next calculations are organized separately because they are included in the original research paper by Wang. We followed his formula to be consistent. Before these calculations were done, we performed trailing twelve month transformation on each data column. For a given period n, it is calculated as the sum of (n – 1 + n – 2 + n – 3 quarters). This is help remove the seasonality of the data. For more information on how the data preparation for the features were done, it can be read in Wang's paper[1].

These transformations were done using a combination of Python and Pandas. After all the fields are extracted and placed in their respective .csv files we use another Python to create the *model.csv* file used in analysis for each company. Given the required effort for data cleaning, it is worth mentioning that our analysis is a small sample of the original paper's universe for data analysis, and is not entirely representative. Now that the scripts are built however, it would easier to extract larger datasets for future analysis. The only manual process is download the files from Bloomberg. Also the original paper excluded financial sector and some other industries that had did unique types of financial reporting.

The model proposed is:

$$FV_n = (Rev_n \times \beta_1) \times (1 + EBITDA\%_n \times \beta_2) \times (1 + Tax\%_n \times \beta_3) \times (1 + CAPEX\%_n \times \beta_4) \times (1 + \Delta WC\%_n \times \beta_5) \times (1 + RG\%_n \times \beta_6) \times (1 + EE\%G_n \times \beta_7) \times (1 + C\%G_n \times \beta_8) \times (1 + 3M_n \times \beta_9) \times (1 + 10Y_n \times \beta_{10})$$

More detail on why these features are chosen can be reviewed in section 2.2. Here we describe how the value of each one of these features are determined. As previously stated, these are largely replicated to match the values done in Wang's paper[1]. Deviations from his formula are mentioned explicitly.

- REV_n - The gross cash inflow received from operations
- $EBITDA\%_n$ - Percentage of REV_n retained as operating income
- $TAX\%_n$ - Percentage of REV_n paid out as tax expenses
- $CAPEX\%_n$, - Percentage of REV_n paid out as capital expenditures
- $\Delta WC\%_n$ - Percentage of REV_n paid out as working capital
- RG_n - Revenue Growth is calculated by taking change in Revenue from period *n* and *n-1*
- $EE\%G_n$ - EBITDA Growth is calculated by taking change in EBITDA Expense margin from period *n* and *n-1*
- $C\%G_n$ - CAPEX Growth is calculated by taking change in EBITDA Expense margin from period *n*
- $3M_n$ – Provided Straight from vendor
- $10Y_n$ – Provided Straight from vendor

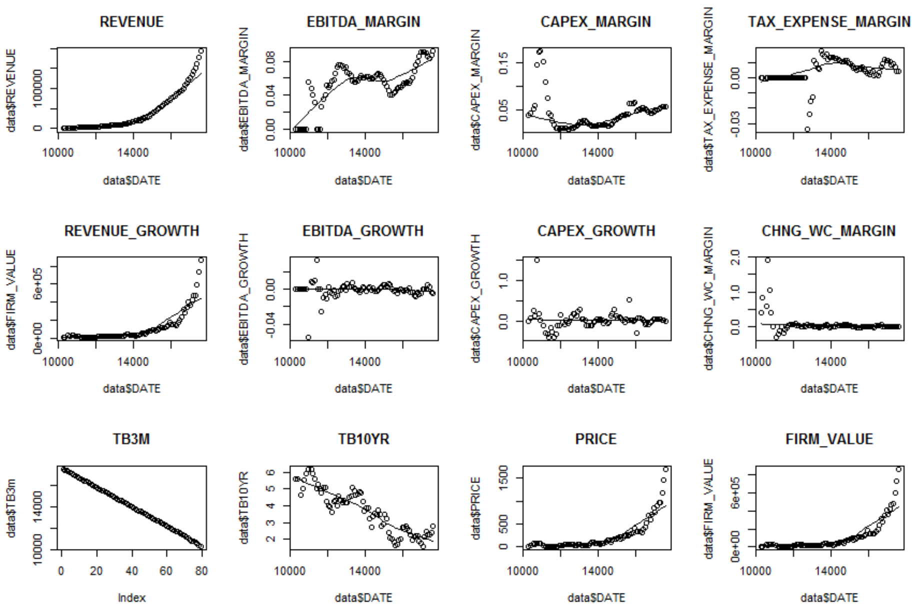
6. Analytic Methods

After the data cleaning and transformation we begin analysis on our nine companies. This is represented as a list of dataframes in R. The model features for each company is represented in its own dataframe for initial exploratory analysis.

```
> head(cleaned_model_data$apple)
           REVENUE EBITDA_MARGIN TAX_EXPENSE_MARGIN CAPEX_MARGIN CHNG_WC_MARGIN REVENUE_GROWTH EBITDA_GROWTH
2018-06-01  246777     0.3069087         0.06120911   0.05331129   0.0489713385     0.03454824   0.001971093
2018-03-01  238536     0.3082721         0.06881142   0.05085186   0.0369755509     0.04349196   0.002484207
2017-12-01  228594     0.3099863         0.06884695   0.05505831  -0.0001399862     0.02275991   0.006013660
2017-09-01  223507     0.3141110         0.07028415   0.05648145   0.0416004868     0.01383490   0.003098264
2017-06-01  220457     0.3162295         0.07162848   0.06068304   0.0422803540     0.01072355   0.005221212
2017-03-01  218118     0.3197810         0.07226364   0.05942655   0.0874526632     0.01407311   0.008262540
           CAPEX_GROWTH     TB3M    TB10YR FIRM_VALUE NON_OP_ASSETS     WADS  PRICE       DATE
2018-06-01   0.04836452 1.840000 2.920625   792842.7        129143 4926.609 185.11 2018-06-01
2018-03-01  -0.07639991 1.560000 2.758525   702676.5        145386 5068.493 167.78 2018-03-01
2017-12-01  -0.02519658 1.206667 2.371452   724003.1        162697 5157.787 169.23 2017-12-01
2017-09-01  -0.06923818 1.036667 2.241429   653409.9        153215 5183.585 154.12 2017-09-01
2017-06-01   0.02114353 0.890000 2.260952   599469.3        153177 5233.499 144.02 2017-06-01
2017-03-01  -0.05653116 0.590000 2.446613   606878.8        158319 5261.688 143.66 2017-03-01
> |
```

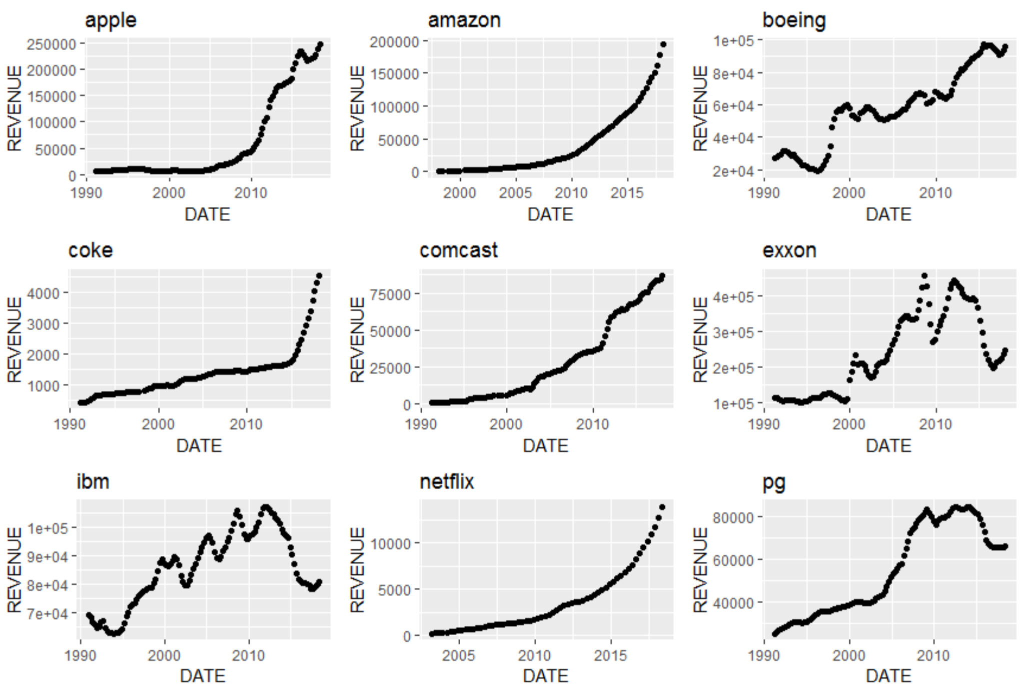
6.1 Outliers

First objective in our analysis was to consider the outliers of the dataset for each parameter of the model. A method was developed to plot each model feature for the nine companies, along with a method to plot all model features for one company. This way we can get a comprehensive view of our data by model feature, and by company. First a quick look at each feature visually will aid us in figuring out what outliers should be considered. We will take one sample company, Amazon.

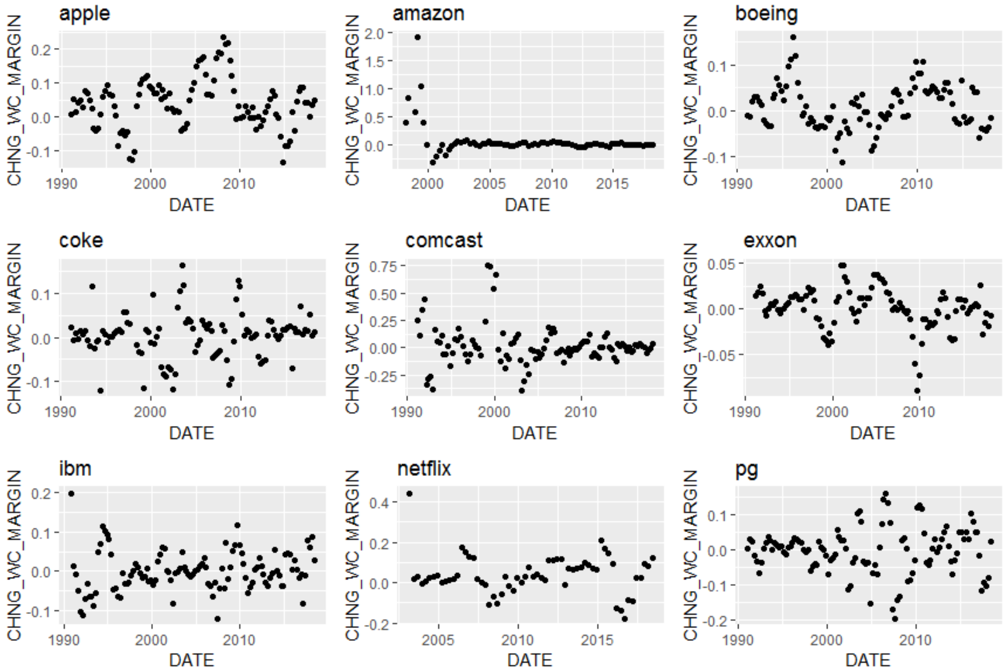


It is clear from the various scatterplots that some of our features do exhibit some outliers. Examples of some extreme outliers include the tax expense margin and change in working capital margin. In the next section we will take a look on how to handle these outliers. One thing to consider is perhaps we just had bad data on one of our companies. The next method used will extract the same feature for all nine companies and display it in a similar matter. This way we will be able to identify what features are more commonly prone to outliers instead of a single company. For comparing one feature for all companies, we provide two samples of features, revenue and Change in Working Capital Margin along with the intended predicted value (Firm Value). To see the graphs of all model features please the appendix.

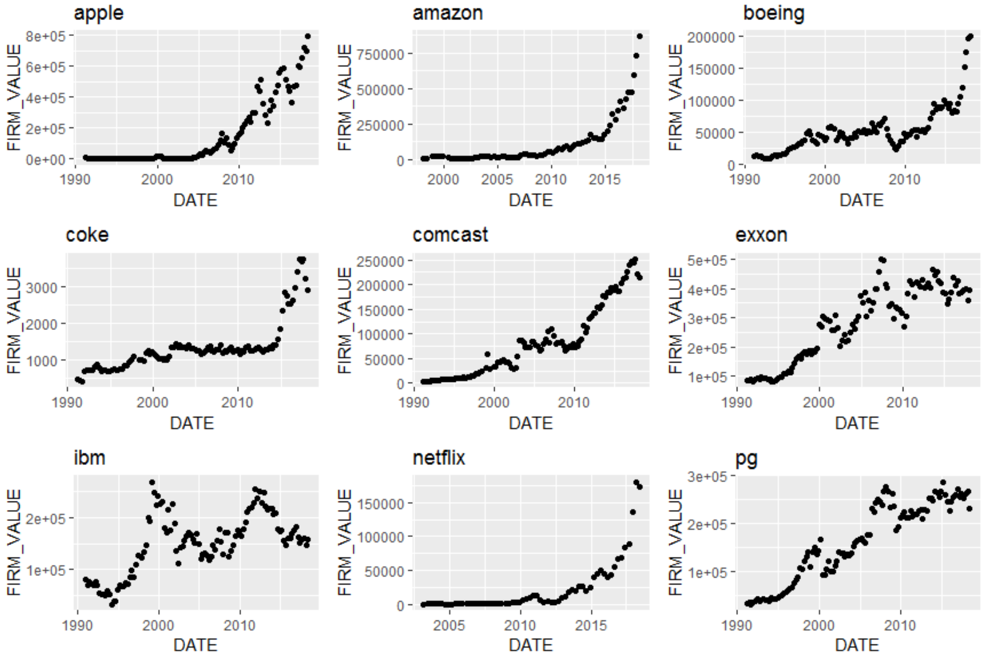
Revenue



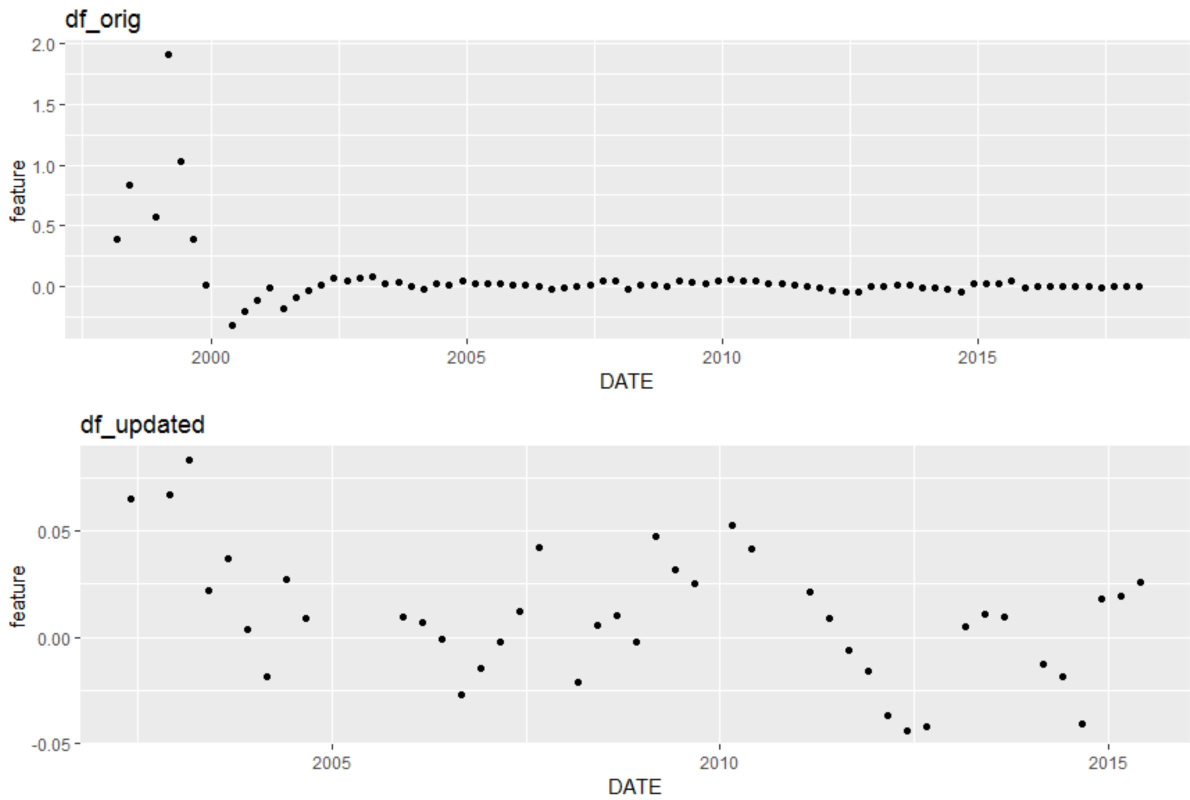
Change in Working Capital Margin



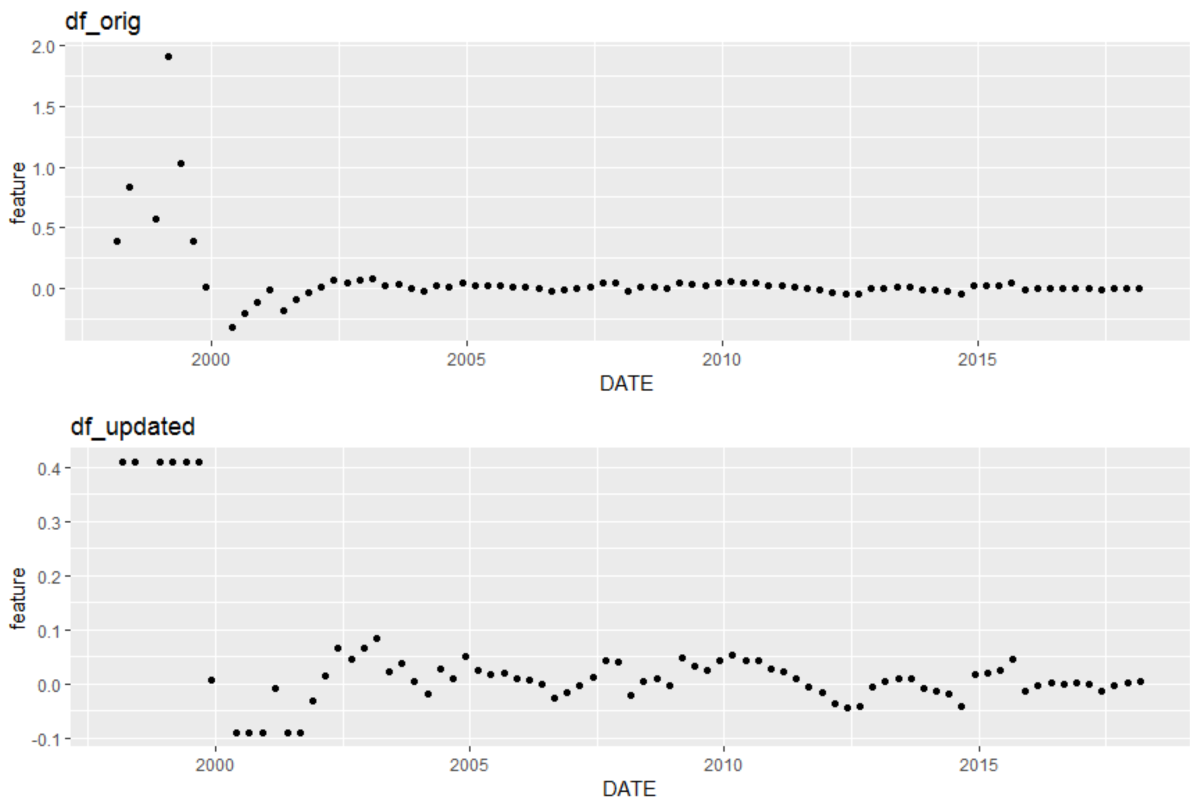
Firm Value



From the previous three graphs it is made quite clear that our data on revenue and calculated firm value is pretty solid, while our data on change in working capital is prone to some outliers (while being overall more volatile). After removing any outliers that are clearly just issues from egregious data, we have three options for handling outliers that actually impact the analysis of the regression. First, and the easiest option, is to simply ignore the outliers. The second option would be to remove any outliers, and the third is capping the outliers based on the 1st and 3rd quartile. First let us visualize how the data may change when removing outliers.



As you can see, the data is clearly transformed by removing the first few points. One thing to keep in mind is that outliers from all features are removed, not just this feature. So overall it means quite fewer data points. Our next method caps the outliers so they remain within the inter quartile range, as shown below.



Moving forward, the regression done will use data from the capped outlier method. Usually the no-outlier case may be optimal, however given the lower number of data points we have decided the method that keeps more data for the model. After the outliers are cleaned, we transform our list of dataframes into one dataset. As the figure shows, we have 904 rows of data (about 100 samples per company) ranging from the 4th quarter in 1990 to the 2nd quarter in 2018.

```
> df <- ldply(test_data, data.frame)
> summary(df)
      .id               REVENUE          EBITDA_MARGIN     TAX_EXPENSE_MARGIN   CAPEX_MARGIN      CHNG_WC_MARGIN
 Length:904         Min.   :   177.9   Min.   :0.00000   Min.   :-0.049807   Min.   :-0.01047   Min.   :-0.260152
 Class :character   1st Qu.:  5998.7   1st Qu.:0.09277   1st Qu.: 0.008707   1st Qu.: 0.01893   1st Qu.:-0.023558
 Mode  :character   Median : 42422.0   Median :0.14216   Median : 0.025867   Median : 0.03411   Median : 0.006874
                    Mean   : 67052.1   Mean   :0.16555   Mean   : 0.029606   Mean   : 0.04588   Mean   : 0.012492
                    3rd Qu.: 86590.0   3rd Qu.:0.22316   3rd Qu.: 0.048019   3rd Qu.: 0.05430   3rd Qu.: 0.040205
                    Max.   :455088.0   Max.   :0.45522   Max.   : 0.135176   Max.   : 0.27082   Max.   : 0.409558
 REVENUE_GROWTH      EBITDA_GROWTH        CAPEX_GROWTH          TB3M             TB10YR         FIRM_VALUE
 Min.   :-0.115630   Min.   :-2.890e-02   Min.   :-0.94776   Min.   :0.01333   Min.   :1.564   Min.   :   526.3
 1st Qu.: 0.000869   1st Qu.:-5.364e-03   1st Qu.:-0.05029   1st Qu.:0.15667   1st Qu.:2.744   1st Qu.:  9620.0
 Median : 0.018775   Median :-4.990e-04   Median : 0.00000   Median :1.72000   Median :4.267   Median : 69921.8
 Mean   : 0.029242   Mean   :-2.026e-05   Mean   : 0.01501   Mean   :2.40158   Mean   :4.322   Mean   :116660.5
 3rd Qu.: 0.050442   3rd Qu.: 3.801e-03   3rd Qu.: 0.04979   3rd Qu.:4.65000   3rd Qu.:5.616   3rd Qu.:176662.8
 Max.   : 0.351668   Max.   : 4.162e-02   Max.   : 0.97981   Max.   :6.99000   Max.   :8.406   Max.   :606878.8
 NON_OP_ASSETS          WADS              PRICE               DATE
 Min.   :-59943.0   Min.   :   8.365   Min.   :  0.4754   Min.   :1990-12-01
 1st Qu.:-11213.0   1st Qu.: 467.750   1st Qu.: 16.1323   1st Qu.:1999-03-01
 Median :  -878.1   Median :1804.900   Median : 45.0225   Median :2005-12-01
 Mean   :  -912.6   Mean   :2467.175   Mean   : 69.7741   Mean   :2005-06-20
 3rd Qu.:  1794.5   3rd Qu.:4717.095   3rd Qu.: 82.0575   3rd Qu.:2012-03-01
 Max.   :162697.0   Max.   :7049.000   Max.   :962.0150   Max.   :2018-06-01
> |
```

All data, from the original excel files provided from Bloomberg, to the .csv files generated from the data wrangling (and the scripts used to do it) are available on github.com/dano09.

7 Regression and Results

Our data is cleaned, and we have verified outliers. Now we are finally able to perform our analysis on the equilibrium valuation model. Wang's original paper did not go into detail about what kind of regression he did. The information provided was that he used excel solver to minimize the MAPE error. Here we will explore a couple of models and review how well each feature performs and their relevance on prediction of firm value.

Model 1: Linear Regression

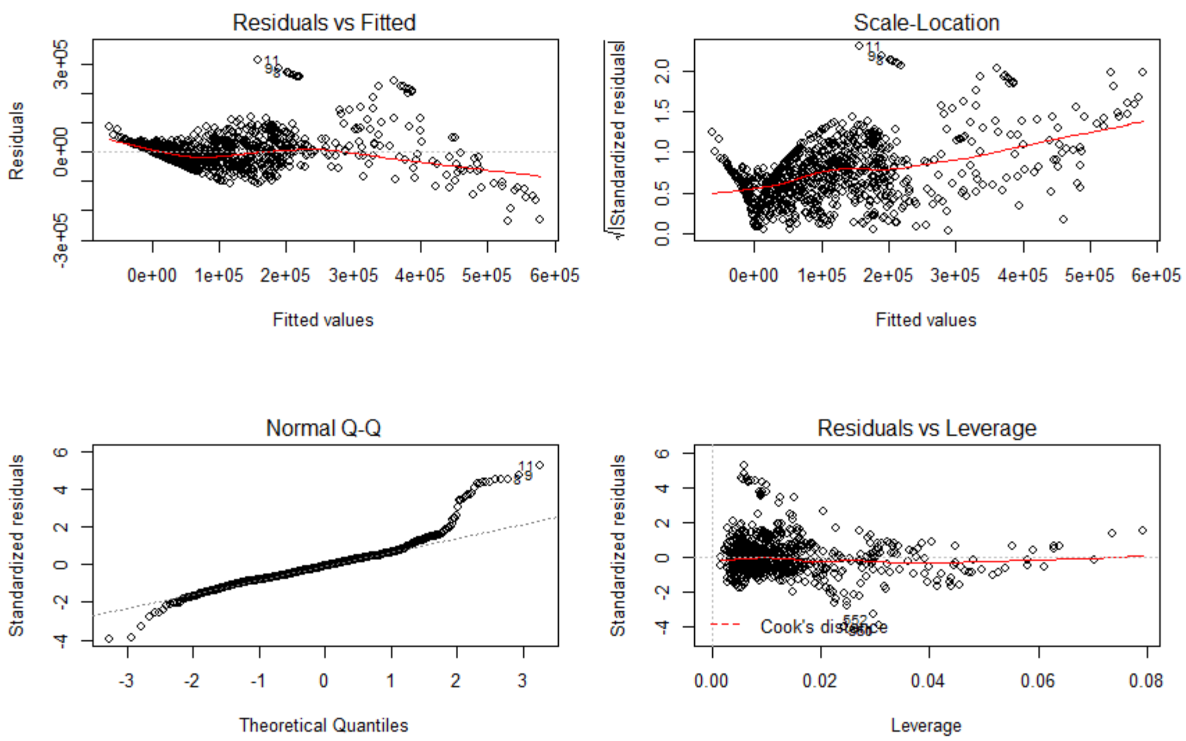
Our first model is the standard linear regression in R. We use the linear model function and create the model using the features described in previous sections.

```
# LINEAR REGRESSION
model <- lm(df$FIRM_VALUE ~ df$REVENUE + df$EBITDA_MARGIN + df$TAX_EXPENSE_MARGIN +
              df$CAPEX_MARGIN + df$CHNG_WC_MARGIN + df$REVENUE_GROWTH +
              df$EBITDA_GROWTH + df$CAPEX_GROWTH + df$TB3M + df$TB10YR)
summary(model)
layout(matrix(c(1,2,3,4),2,2))
plot(model)
```

The code to run the model is simple to implement. Next we will look at the diagnostics and summary of variables from the model.

```
> summary(model)

Call:
lm(formula = df$FIRM_VALUE ~ df$REVENUE + df$EBITDA_MARGIN +
    df$TAX_EXPENSE_MARGIN + df$CAPEX_MARGIN + df$CHNG_WC_MARGIN +
    df$REVENUE_GROWTH + df$EBITDA_GROWTH + df$CAPEX_GROWTH +
    df$TB3M + df$TB10YR)

Residuals:
    Min      1Q  Median      3Q     Max
-233761  -34999   -3866   25416  316894

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            8.794e+04  8.707e+03  10.100  < 2e-16 ***
df$REVENUE             1.105e+00  2.868e-02  38.522  < 2e-16 ***
df$EBITDA_MARGIN       3.127e+05  2.816e+04  11.102  < 2e-16 ***
df$TAX_EXPENSE_MARGIN  2.141e+05  1.072e+05   1.997   0.0462 *
df$CAPEX_MARGIN       -2.035e+05  5.159e+04  -3.944 8.64e-05 ***
df$CHNG_WC_MARGIN     -6.395e+04  2.782e+04  -2.299   0.0218 *
df$REVENUE_GROWTH      6.843e+04  3.887e+04   1.760   0.0787 .
df$EBITDA_GROWTH      -4.727e+05  1.910e+05  -2.475   0.0135 *
df$CAPEX_GROWTH       -1.453e+04  9.855e+03  -1.475   0.1406
df$TB3M                1.232e+04  1.977e+03   6.231 7.15e-10 ***
df$TB10YR             -2.885e+04  2.436e+03 -11.844  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 59990 on 893 degrees of freedom
Multiple R-squared:  0.7947,    Adjusted R-squared:  0.7924
F-statistic: 345.6 on 10 and 893 DF,  p-value: < 2.2e-16
```

Revenue, EBITDA, CAPEX Margin seemed to be the most significant features for our initial analysis, along with the two risk-free rates. Looking particularly as the Multiple R-squared, and Adjusted R-square and seeing these values of approx. 79% give an impression that there is some correlation between our model and the firm value.

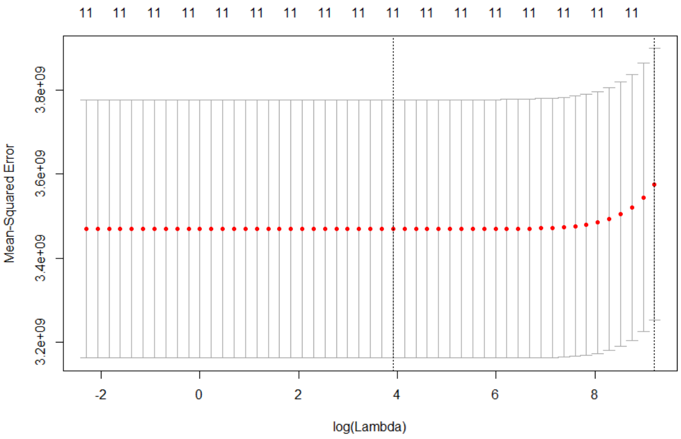


The Residuals vs Fitted chart sheds light on our underlying factors. It gives insight in that there may be a non-linear relationship between our predictor variables and the outcome variable. The residuals shown from Normal Q-Q plot appear to behave linearly except towards the end of the distribution, it almost looks as if it follows some sort of Poisson process. Checking for Homoscedasticity, we view the Scale-Location graph. While somewhat linear, there is a clear pattern of the majority of the variance taking place early on followed by a dispersion. And since we already took care of the majority of our outliers, the Cook's distance presented in the Residuals vs Leverage diagram does not provide much useful insight.

Model 2 – Ridge Regression

Our next three models will try to build out robust models using regularization[5]. For ridge regression, we check various regularization values (lambda) to identify the best fit. We will accomplish this by using the cross validation library in R. These models apply regularization terms to penalize models that have large numbers of coefficients. There are plenty of online resources for discovering the different implementation details of these models.

```
# RIDGE REGRESSION
set.seed(123)
ridge_model <- cv.glmnet(df_matrix, firm_value, lambda=10^seq(4, -1, -.1), alpha=0)
best_ridge_lambda <- ridge_model$lambda.1se
ridge_coef <- ridge_model$glmnet.fit$beta[, ridge_model$glmnet.fit$lambda == best_ridge_lambda]
```

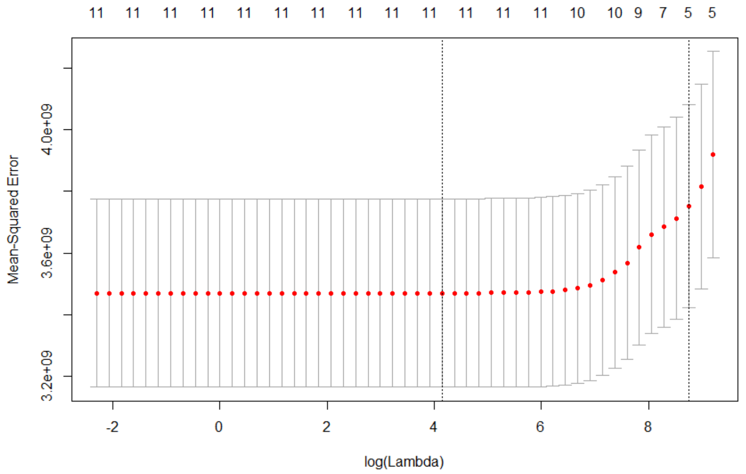


Model 3 – LASSO Regression

Similar to Ridge, we can perform LASSO and elastic net regression by changing the alpha parameter.

```
# LASSO REGRRSSION
set.seed(123)
lasso_model <- cv.glmnet(df_matrix, firm_value, lambda=10^seq(4, -1, -.1), alpha=1)
# Lambda that gives the best error
best_lasso_lambda <- lasso_model$lambda.1se
lasso_coef <- lasso_model$glmnet.fit$beta[, lasso_model$glmnet.fit$lambda == best_lasso_lambda]
```

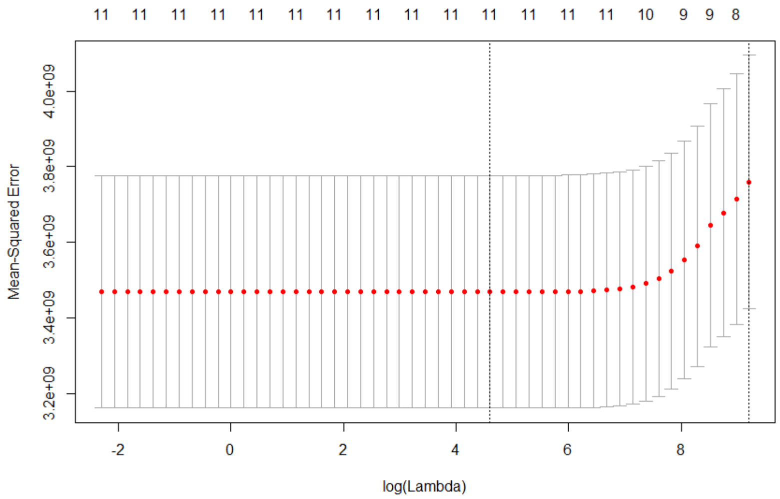
The graph below shows for LASSO lambda is best at around 4 and starts increase the error after it grows past 6.



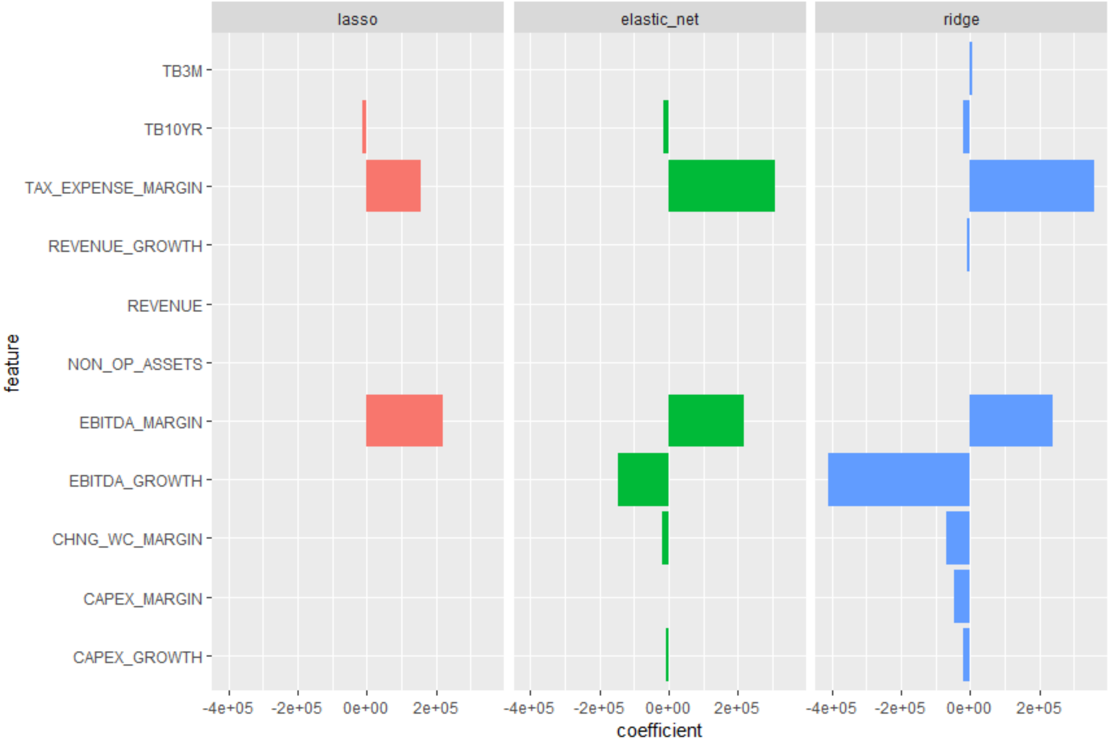
Model 4 – Elastic Net Regression

Finally we take a look at a combination of both L1 and L2 regularization which is called elastic net.

```
# ELASTIC-NET
set.seed(123)
elastic_net <- cv.glmnet(df_matrix, firm_value, lambda=10^seq(4, -1, -.1), alpha=0.3)
best_en_lambda <- elastic_net$lambda.1se
net_coef <- elastic_net$glmnet.fit$beta[, elastic_net$glmnet.fit$lambda == best_en_lambda]
```



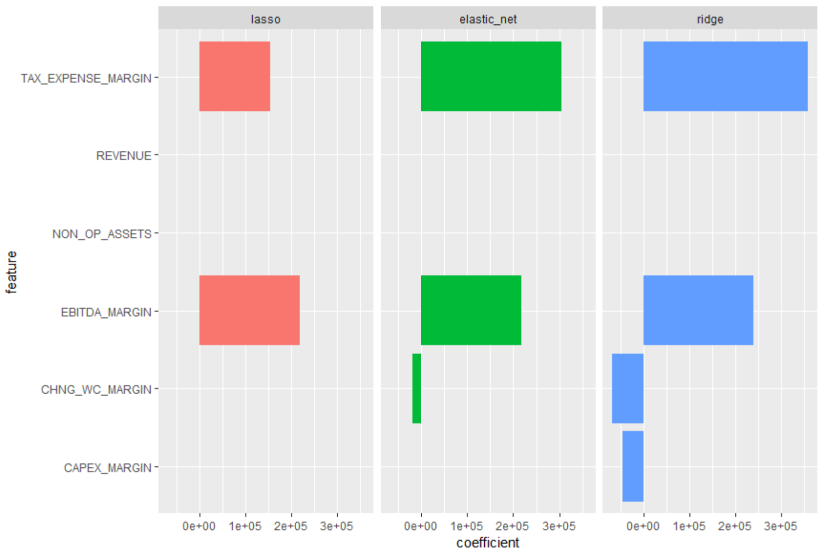
We will now compare the coefficient values for each of regularization models to determine which features are most useful for the equilibrium valuation model.



Here we get a snapshot showing how each model identifies features. Don't let the size of coefficients throw you off, they are a function of the size of the feature. The important thing to note is while Ridge regression makes use of all features, Lasso regression only showed that the 10 year treasury yield, Tax Expense, and EBITDA margin are the most important factors. Elastic net does a combination of both models loss terms so it includes a few more. Let us break down each coefficient graph by their three main components, value, growth, and risk-free rate.

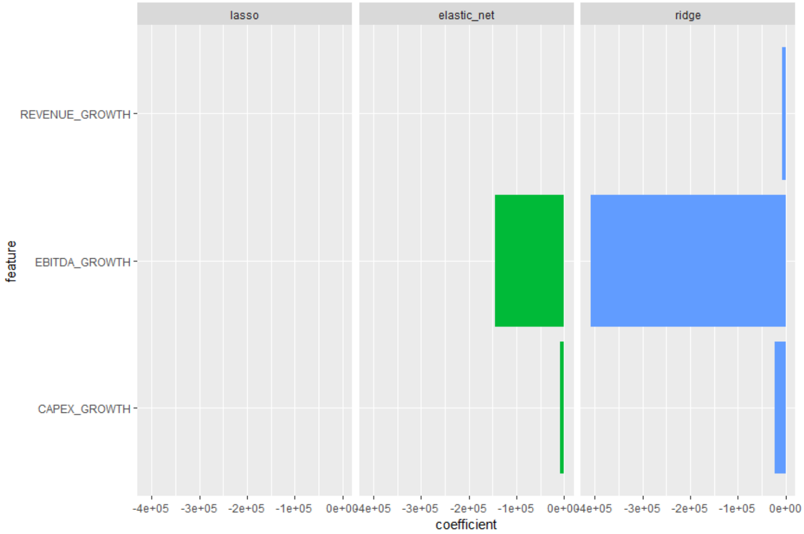
Price Factors

For price factors, we see all factors being used except Revenue and Operating Assets for ridge regression. This may be because these two features are the only ones that are representative of actual value, not a margin calculation.



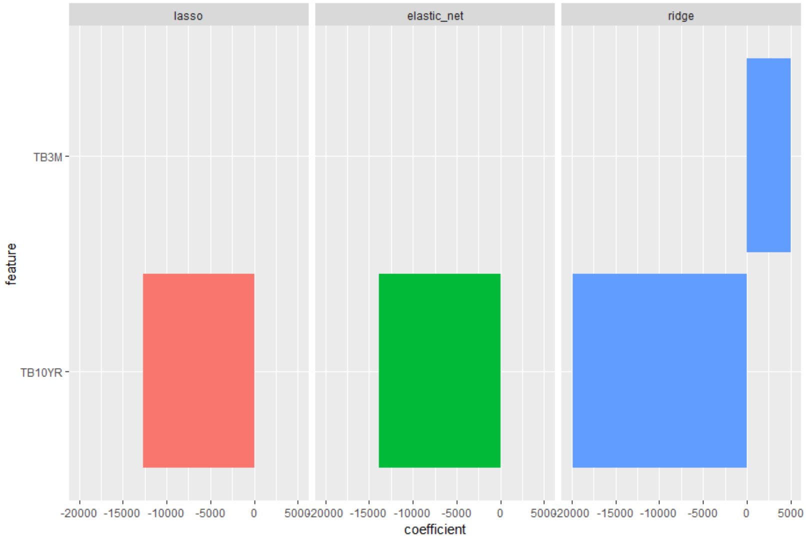
Growth Factors

For growth factors, Ridge regression focuses on EBITDA growth, Elastic Net using the same, but with a lower weighting, and lasso completely removing all growth factors.



Risk-Free Rate

It is clear that the 10 Year Treasury Rate has an impact on the equilibrium model. This comes almost obviously since most things in finance are price in relation to the risk free rate. The thing that is interesting is how much more the longer term horizon rate (10 Years) appeared relevant compared to the shorter term treasury yield.



Conclusion

Our analysis has taken a look at the equilibrium valuation model from Wang. From extracting the data to replicating the features and finally performing regression on the model. Our findings show that the model does have potential to be used in a predictive manner, and that some of the features may appear to have some potential to be indicators on the firm value. For determining the accuracy of this model, future work could be done to apply these models to test datasets of new financial statements.

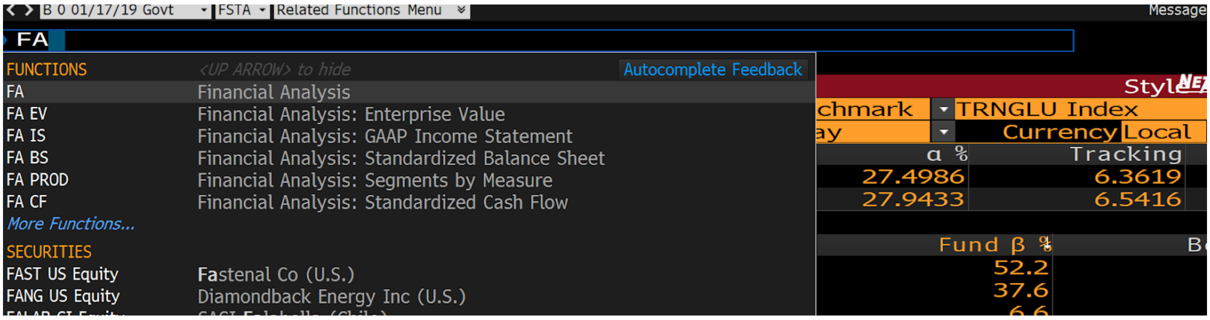
Appendix

[1] Equilibrium Valuation Mode: A Regression-Based Fundamental Equity Valuation Model

[2] https://www.investopedia.com/terms/e/equilibrium.asp

[3] https://www.stlouisfed.org/

[4] Using FA (financial Analysis function)



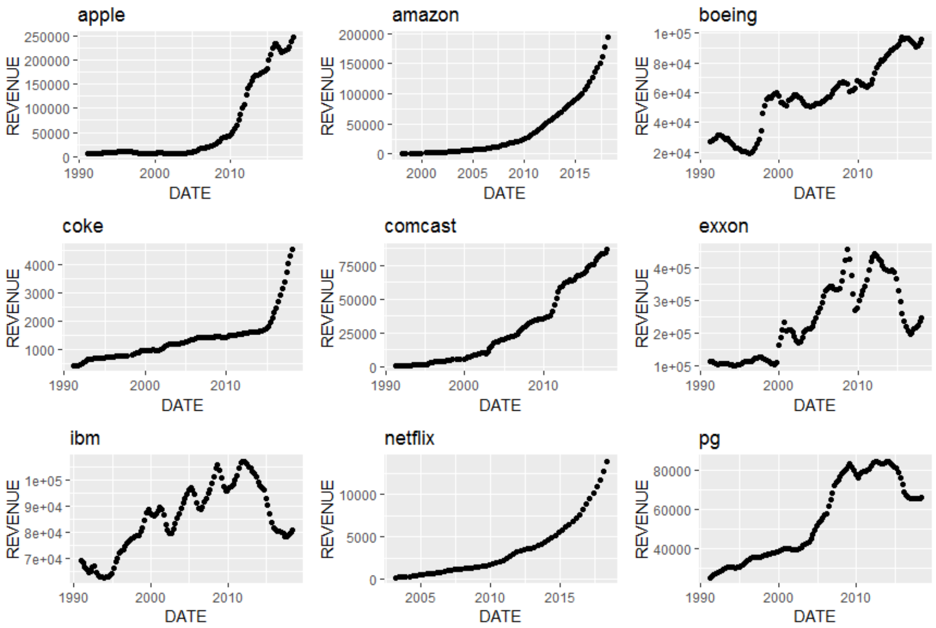
[5] https://www.youtube.com/watch?v=FWCPFUwZkn0

Revenue	SALES_REV_TURN
Cost of Revenue	IS_COGS_TO_FE_AND_PP_AND_G
Operating Expenses	IS_OPERATING_EXPN
EBITDA	EBITDA
Income Tax Expense	IS_INC_TAX_EXP
Total Current Assets	BS_CUR_ASSET_REPORT
Total Current Liabilities	BS_CUR_LIAB
Cash, Cash Equivalents & STI	CE_AND_STI_DETAILED
LT Investments & Receivables	BS_LT_INVEST
ST Debt	BS_ST_BORROW
LT Debt	BS_LT_BORROW
Minority/Non Controlling Interest	MINORITY_NONCONTROLLING_INTEREST
Preferred Equity and Hybrid Capital	BS_PFD_EQTY_&_HYBRID_CPTL
Change in Fixed & Intang	CHG_IN_FXD_&_INTANG_AST_DETAILED

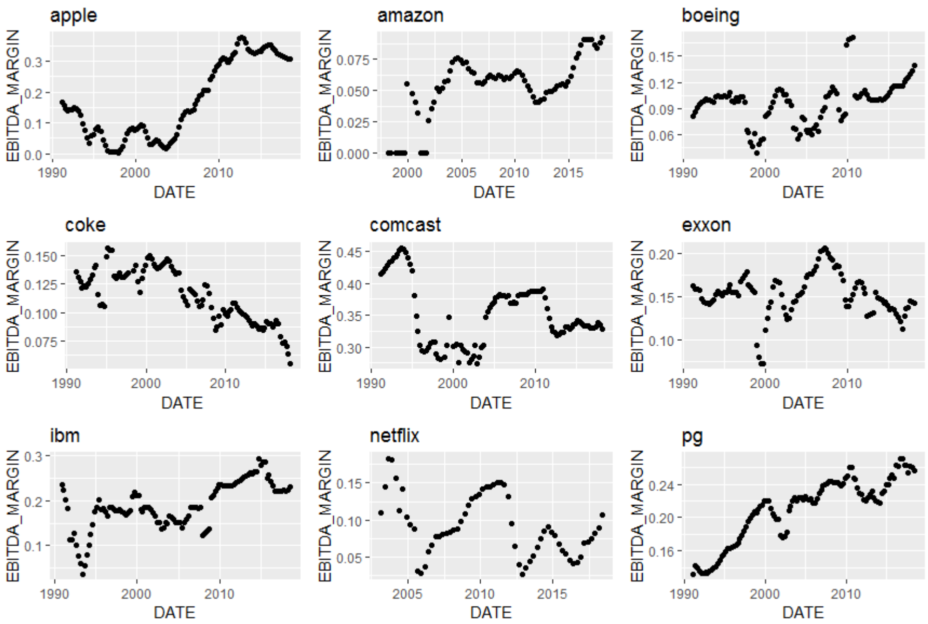
Diluted Weighted Avg Shares	IS_SH_FOR_DILUTED_EPS
Last Price	PX_LAST
3 Month Treasury	TB3MS
10 Year Treasury	DGS10

Additional Graphs:

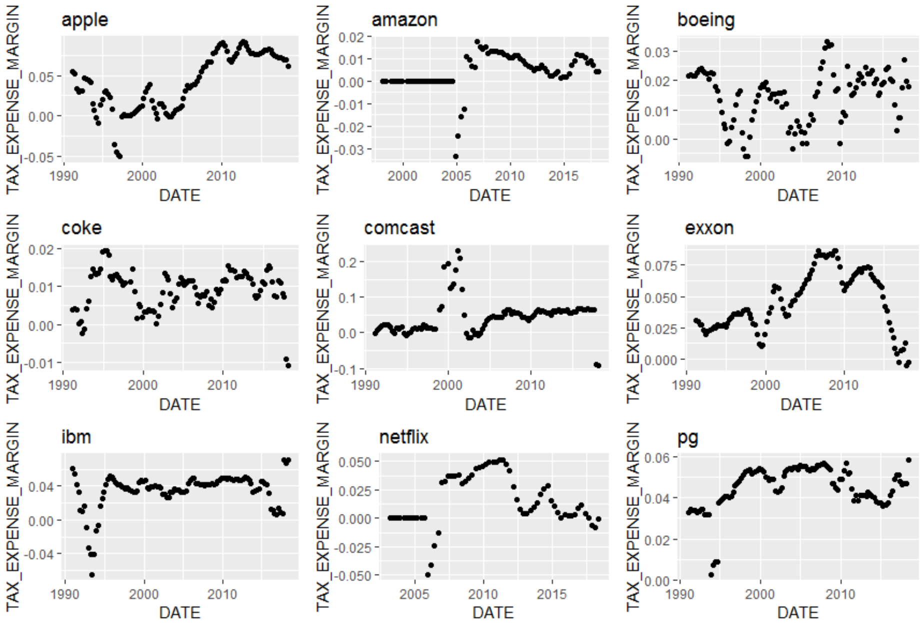
Revenue



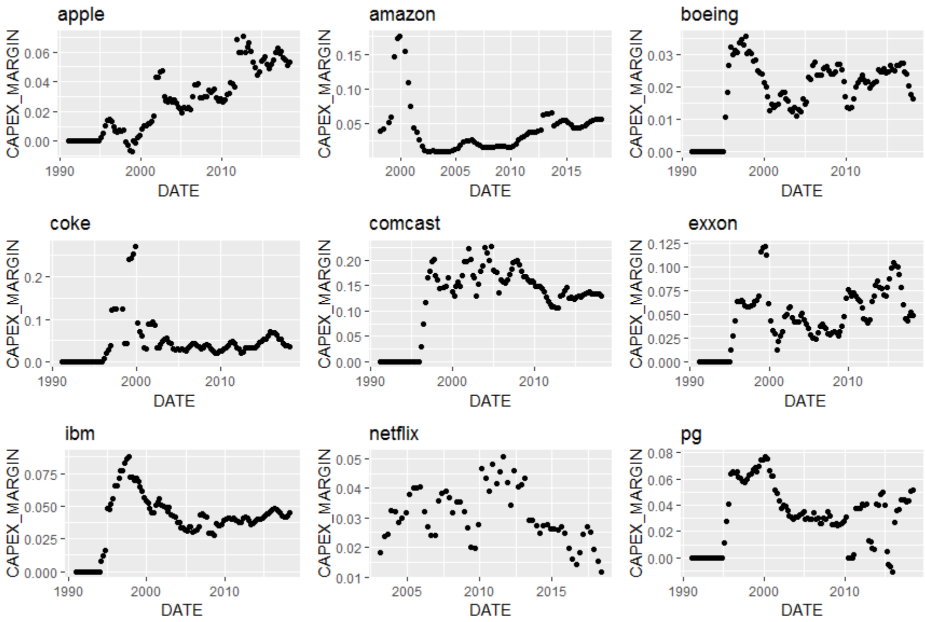
EBITDA MARGIN



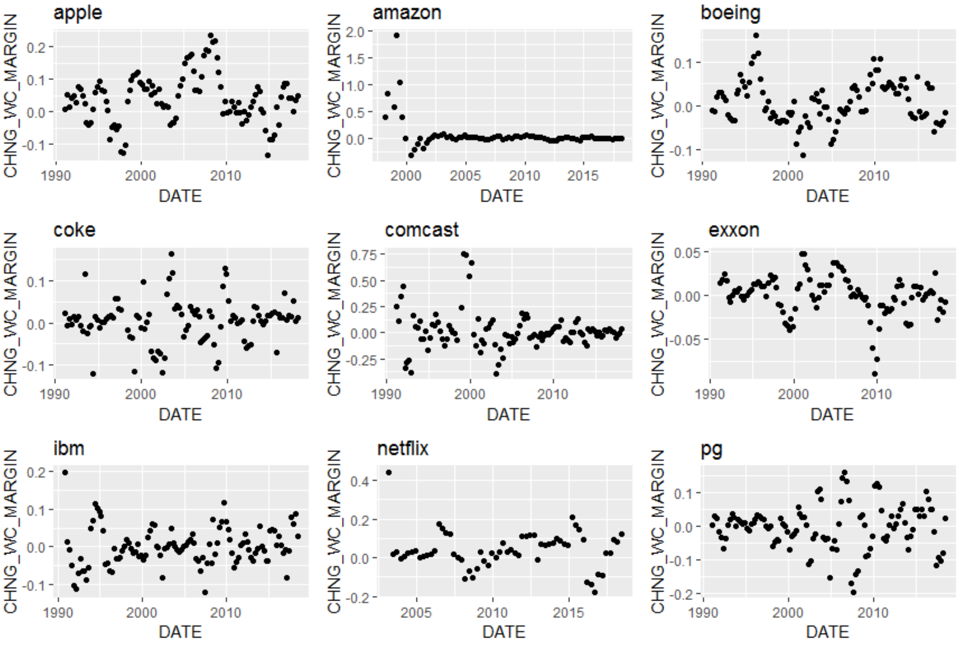
TAX EXPENSE MARGIN



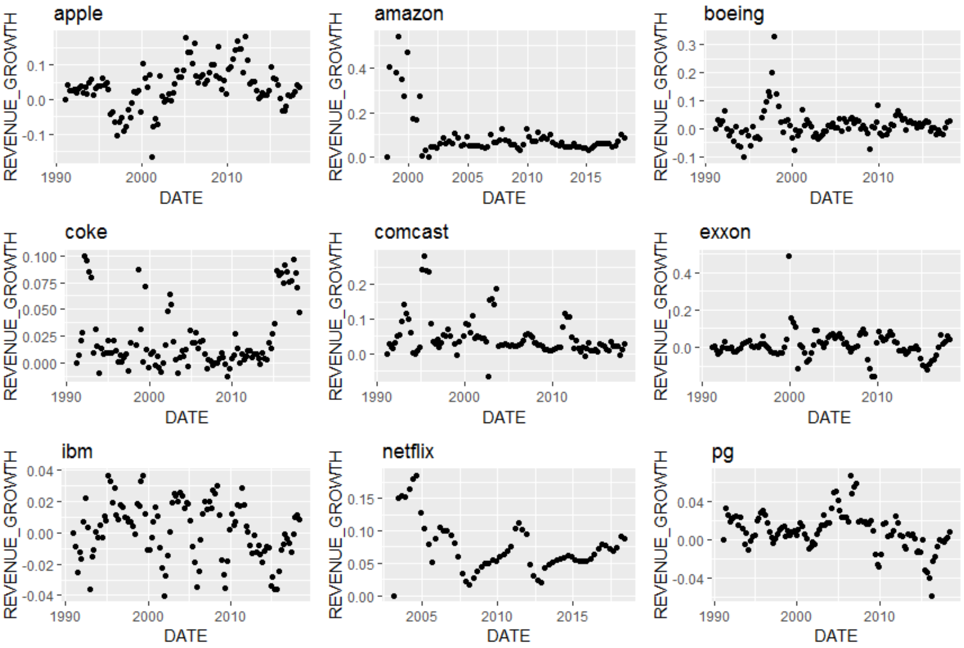
CAPEX MARGIN



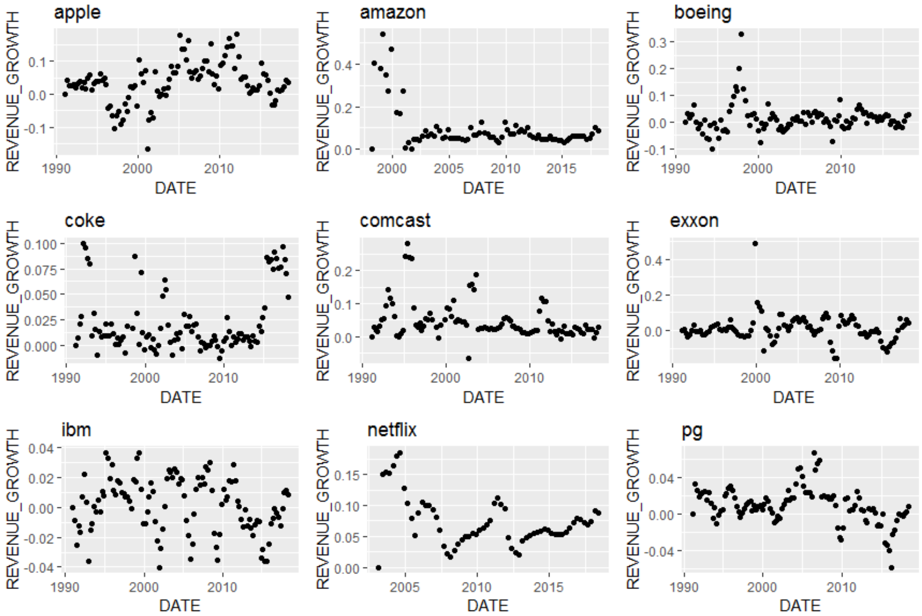
CHNG WC MARGIN



REVENUE GROWTH



EBITDA GROWTH



CAPEX GROWTH