1-1-2019

Corporate Equity Performance and Changes in Firm Characteristics

Cole McLemore
Mississippi State University

Corporate Equity Performance and Changes in Firm Characteristics

By

Cole McLemore

Table of Contents

1. Introduction....1
2. Background and Related Literature....6
3. Empirical Design....10
    - 3.1. Sample and Data....11
    - 3.2. Research Design....12
    - 3.3. Preliminary Analysis....15
    - 3.4. Factor Analysis....16
    - 3.5. Empirical Approach....17
    - 3.6. Return Analysis....18
    - 3.7. Sample Characteristics....19
    - 3.8. Industry Analysis....21
    - 3.9. Alternate Dependent Variables....23
4. Conclusion....26

References....28

Tables and Figures....29

- Table 1: Factor Model Empirical Design....29
- Table 2: Summary Statistics....30
- Table 3: Five-Factor Model Alternatives, Significance Evaluation....31
- Table 4: Alternative Measures of Firm Value and Performance....32
- Table 5: Industry Analysis....33
- Figure 1: Industry as a Percent of Total Observations....34
- Figure 2: Histogram of Percent Change in Adjusted Price....35

Abstract

While prior equity performance research analyzes portfolio characteristics using multi-factor models, portfolio groups are typically utilized to explain average returns. Instead, I explore annual firm-level data to compare with annual percentage changes in firm characteristics, emphasizing model predictive power and individual variation. My analyses result in a significant link between individual firm equity returns with percentage changes in total assets, book-to-market ratios, current ratios, and shares outstanding, in addition to historical returns and average market returns. My findings affirm prior work illustrating the importance of profitability, size, liquidity, momentum, and market returns, though I observe minimal evidence of the importance of investment in capital expenditures. I also perform these analyses at the industry level and note differences across industries, including the cyclical nature of business equipment and consumer durables industries in contrast to the utilities and energy sectors. Overall, I contribute to the understanding of corporate characteristics and equity performance.

1. Introduction

Financial literature from figures such as the notable Fama and French (2015) show that returns are dependent on five separate factors. These factors are size, liquidity, profitability, investment patterns, and market performance, which include those from the Fama and French (1993) original three-factor model in addition to research from Carhart (1997) who introduces momentum. These researchers perform their analyses on portfolio groupings to assess average measures. My research aims to both corroborate the significance of these models in addition to utilizing a novel framework: my analyses all utilize annual percent changes, similar to the data structure. Each observation in this research measures in terms of percent change over a year's time. This is because returns that represent the percent change in value of a firm stock price. Therefore, since I am seeking to explain a percent change, I also use percentage changes in related corporate characteristics, which further separates my work from prior research. While portfolio-based regression methods have innate weighting and grouping by size, percent change standardizes all data such that original size does not matter. The only information that affects the models are firm characteristic changes affect analyses. These changes further increase the weighting of small companies as they have the largest propensity to achieve abnormal percentage changes in their accounting measures. To perform generalizable research, I require minimal data constraints. These inclusions can affect data distributions. However, the role of extreme values is not substantive.

A consequence of using raw data is low r-squared values. The highest r-squared value that this research achieves in a significant and fair way is when comparing *ADJUSTED RETURN* to the utilities industry (r-squared of 0.1959). Meanwhile, Fama and French (2015) created 5 x 5 sorted portfolio models that achieve average r-squared of between 0.91 and 0.93. The *ADJUSTED RETURN* model in my research only achieves an r-squared of 0.0585 compared to the whole universe of firms. The drastically smaller r-squared values are due to the nature of observing twenty-eight years of data across all public firms and attempting to draw patterns from their changes. Such a huge collection of varied data, especially when studying a notoriously volatile variable, *STOCK RETURN*, will most likely not agree enough for a model to produce an r-squared anywhere near 0.90. Portfolio-based variables lump like firms together and removes much of the market noise through diversification.

This brings up an important distinction between the results of this research and that of previous literature. The models produced here indicate that as the change in factors changes, returns should move accordingly. This is an observation of the second derivative of accounting items with respect to time as much as it is an observation of the first derivative with respect to time. The low r-squared values could also be coming from this distinction. These models seem to show that the relation of the movement of changing accounting variables is much less responsible for movement of returns than the simple movement of nominal accounting items. This dataset varies from previous literature in three key ways: lack of portfolio modeling, studying change in variables rather than nominal values, and the inclusion of financial firms, utilities firms, and small firms. This means that the consistently lower power of prediction from this research's models could be due to any of these three differences or a combination thereof. However, the lack of portfolio modeling is absolutely causing some degree of drop in r-squared as a regression model can much more easily match the variance of the data it is modelling if the data has fewer points that do not conflict. Without creating portfolios, the dataset reaches a total of 168,861 observations. With these many observations, data contradiction and overlap are bound to occur and harshly limits the upper bound of the accuracy of a regression model.

Since this research does not use portfolio grouping, prior research does not directly translate into this research's models. Each factor from the Fama and French (2015) and Carhart (1997) model have the possibility to be expressed by multiple accounting items. Size can be described by *TOTAL ASSETS*, *TOTAL EQUITY*, or possibly *COMMON SHARES OUTSTANDING*. Liquidity can be described by *CURRENT ASSETS*, a *RATIO OF CURRENT TO TOTAL ASSETS*, and the *CURRENT RATIO*. Profitability can be described by *REVENUE*, *NET INCOME*, *EARNINGS BEFORE INTEREST AND TAXES*, and the *BOOK-TO-MARKET RATIO*.[1] Investment can be described by *CAPITAL EXPENDITURES* and acquisitions expense. Market performance is described by *VALUE-WEIGHTED S&P 500 RETURN*.[2] Momentum is analogous to the returns accrued over the past twelve months. Each combination of variables is tested in multivariate regression models here each model consists of one item from each list of possible variable substitutes for established factors against returns. Each dependent variable is still in terms of percent change over one year.

This combinatorial model process produces six significant variables: *VALUE-WEIGHTED S&P 500 RETURN*, *PRIOR RETURN*, *TOTAL ASSETS*, the *CURRENT RATIO*, the *BOOK-TO-MARKET RATIO*, and *TOTAL EQUITY*. These six variables represent two variables that represent size, one variable to represent *MARKET RETURN*, liquidity, momentum, and profitability, and no variable to represent investment. Since two variables are found to be significant for size, *TOTAL EQUITY* will be thrown out for future model testing as *TOTAL ASSETS* has higher average t-values and less correlation with other variables. The only factor that does not match up is

[1] Return on assets and return on equity are also considered, but in a later stage of model testing. This stage attempts to keep modelling simply by examining firm characteristics that only rely on one accounting item at a time. The book-to-market ratio is an exception in this step due to its documented significance to returns (Fama and French 1993).

[2] Equal-weighted returns have been observed prior and found to have lower significance when predicting returns. Value-weighted market returns have also been established as significant in predicting returns (Fama and French 1993).

investment. The variable under consideration in this research for investment is *CAPITAL EXPENDITURES*. Unfortunately, the firm-level yearly percent change method of observing financial data does not prove *CAPITAL EXPENDITURES* to be significant in the prediction of *STOCK RETURN*. This could mean that investment is important for the calculation of *STOCK RETURN* in a nominal manner, such that the level of investment matters far more than the change in investment. Once the appropriate level of investment is achieved for a firm, further changes do not seem to significantly impact *STOCK RETURN*.

At this stage, a core five-variate regression model is made for relating firm *STOCK RETURN* to changes in firm characteristics. In order to consider all possible influencers to *STOCK RETURN*, the remaining 95 other variables available in the dataset are added to the core model, one at a time, keeping the most significant additions for a final round of model testing. The resulting six additional variables are tested alongside the five-core variables individually by industry. Industries are determined according to SIC code and based on Fama and French's twelve industries. The variable that is most often significant against the twelve industries is weighted-average common shares outstanding (*SHARES*). This is selected as the final variable to be added to the regression model.

Since *STOCK RETURN* does not take stock splits into account, an *ADJUSTED RETURN* variable is created by applying the cumulative adjustment factor to returns. *ADJUSTED RETURN* is modeled against the same six-variable model and experience an increased r-squared in the *ADJUSTED RETURN* model compared to *STOCK RETURN*. This *ADJUSTED RETURN* model is then broken down by industry and reexamined. Overall, *MARKET RETURN* is the best predictor, significant in every industry. *PRIOR RETURN* is significant to all but the chemical industry and energy. The *CURRENT RATIO* holds significance in only two industries, manufacturing and consumer non-durables. Utilities has the greatest r-squared of any industry (0.1959), while energy experiences the lowest r-squared. Consumer durables industry holds few significant factors, while the consumer non-durables industry shows significant relations for every variable in the model. Overall, this research contributes to the understanding of equity performance and changes in corporate characteristics. The remainder of the paper is organized as follows: sections 2 and 3 describe the related literature and empirical methodology, while section 4 concludes.

2. Background and Related Literature

Fama and French (1993) propose a three-factor model for describing stock returns. The three factors used in this model rely size, market returns, and profitability (B/M):

$$R_{it} - R_{Ft} = \alpha_i + \mathrm{b}_i(R_{Mt} - R_{Ft}) + s_i SMB_t + h_i HML_t + e_{it} \quad [1]$$

R_{it} refers to returns of firm *i* at time *t*. R_{Ft} refers to the risk-free rate at time *t*. R_{Mt} refers to the market returns at time *t*. SMB_t refers to the size factor, small minus big. HML_t refers to the profitability factor, high minus low. e_{it} is the error term for the model. This three-factor model is later tested by Fama and French in 1997 against 48 different industries. These industries are selected and set apart by Standard Industrial Classification (SIC) codes. This research found that dividing data by industry yields high levels of error for cost of equity calculations. Cost of equity is analogous to a theoretical return demanded by the market given a firm's characteristics. Thus, it can be said that this research produces imprecise results when observing the three-factor regression model at the industry level. Fama and French attribute the high variation of risk loadings and the high variation in factor risk premiums as well. These variations over time feed the regression models inconsistent data that produces imprecise results.

A later Fama and French (2006) study on the same model allocates data by taking the universe of firm data from 1963-2003 available on the Center for Research in Security Prices (CRSP) database. Notably, this data is trimmed by excluding all financials firms and all firms with *TOTAL ASSETS* under \$25 million or have a book equity of less than \$12.5 million. This model is later improved upon again to include two more factors in Fama and French (2015)'s five-factor model. This model adds factors of profitability and investment:

$$R_{it} - R_{Ft} = \alpha_i + \mathrm{b}_i(R_{Mt} - R_{Ft}) + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA + e_{it} \quad [2]$$

Each of these models relies on portfolio analysis. Data are not observed at a firm level of observations, but rather by portfolios of similarly classified data. These portfolios are created by finding the intersections of n-dimensional matrices. Each dimension is created by taking a characteristic, such as firm size, B/M, and operating profitability, and creating value buckets for data to fall. The bounds of each bucket are determined based on the NYSE medians and percentiles. Each category that relies on median breakpoints is a combination of two portfolios separated at the median. This data categorization is core to the analyses, as the observed factors rely on the comparisons of different buckets within each category. For instance, the SMB factor, small minus big, is created by taking characteristics of the small portfolios and subtracting the same characteristics of the big portfolio. Other researchers have also made their own changes to the Fama and French models to account for different variables that the researcher may deem important for return regression. Carhart (1997) takes the three-factor model and adds a momentum factor:

$$R_{it} = \alpha_i + \mathrm{b}_i(R_{Mt} - R_{Ft}) + s_i SMB_t + h_i HML_t + p_i PR1YR_t + e_{it} \qquad [3]$$

Stock price momentum is analogous to the first derivative of returns with respect to time. This factor is constructed by creating a 2 x 2 x 2 matrix based on size, book-to-market equity, and one-year momentum on stock returns. Similar to French and Fama (1993), each bounded space in the matrix represents a portfolio of firms. More recently, Asness, Frazzini, and Pedersen (2018) create their own set of factors: Quality Minus Junk (QMJ). QMJ assesses the quality characteristics of a firm irrespective of value. QMJ is reliant on factors of firm profitability, growth, and safety. Profitability is defined in this research as profits per unit of book value. Growth is defined as the five-year growth of profitability measures. Safety is defined as the size of a firm's required rate of return where a low required rate of return is deemed safer than a higher required rate of return. Asness, Frazzini, and Pedersen found that high-quality firms tend to have marginally higher stock values. Since high quality is defined to be reliant on a high amount of safety, thus low risk, high quality firms experience high risk-adjusted returns.

Furthering the work by Fama and French (2006), Aharoni, Grundy, and Zeng (2013) revisit the model and analyze it with respect to the Miller and Modigliani (1961) model. The Miller and Modigliani (1961) model relates market value of a firm's equity to the firm's expected earnings, expected book value, and the market discount rate. This model is based heavily in financial theory and is considered a tautology. Aharoni et al. (2013) claim that Fama and French's model is built on a per share level of observation rather than a firm level. Adjusting the model to a firm level reveals new analyses, including an important note on *COMMON SHARES OUTSTANDING*. Aharoni et al. conclude that a change in *COMMON SHARES OUTSTANDING* likely mitigates the correlation between expected change in investment and expected returns. This correlation comes from previous research from Lyandres, Sun, and Zhang (2008) that an issuance of new equity is related to an immediate increase in investment. This mitigation that comes from the usage of *COMMON SHARES OUTSTANDING* may contribute to the lack of investment significance in my six-factor model that includes weighted average common shares outstanding (*SHARES*) as a significant factor.

In response to Fama and French's (2006) finding that creating portfolios split by profitability yields weak returns relative to the other portfolios examined, Robert Novy-Marx (2013) suggests that profitability can produce equal to or greater returns when observed as gross profits instead of the *BOOK-TO-MARKET RATIO*. Gross profits in this research indicate firm *REVENUE* less the cost of goods sold. Novy-Marx (2013) is suggesting that exchanging the *BOOK-TO-MARKET RATIO* for gross profits produces a stronger model. My research examines a similar path by testing the *BOOK-TO-MARKET RATIO* against *NET INCOME* and *REVENUE*, which are related to gross profit. Thus, according to Novy-Marx (2013), these other profitability factors should produce either better or comparable significance or accuracy to the regression model against returns. However, I find that, when observing a dataset that is not in a portfolio setting and in percent change over one year, the *BOOK-TO-MARKET RATIO* produces significantly better results than any other tested profitability measure.

3. Empirical Design

The data for this research is obtained through the Warton Research Data Service (WRDS). Using WRDS, balance sheet and income statement items are collected along with company identifiers for all publicly traded companies on the Compustat database. The data pulled is yearly starting from January 1990 until September 2018.[3] In total, 96 financial variables are put into this original data pull with a total of 353,239 observations of firm-year data. WRDS also has access to CRSP, which is where the stock price data is pulled from. This data is from the same time period and is analyzed at the monthly level. It contains 2,622,514 observations of data with firm identifiers, date, and price listed.

The two datasets have firm identifiers, but neither has permanent identifiers in common. A merging of the two is done based on CUSIP and date in SAS. CUSIPs are unique identifier codes for financial securities. CUSIPs can be reassigned over time, but a CUSIP will be unique at any one given date. Date is only recognized for the merge on a year and month basis in order to account for data that occurs at the same time but is submitted a few days apart. The only possible issue with this method is if a firm collects another firm's CUSIP in the same month that the original forfeits their old CUSIP, but this occurrence in the dataset proves minimal. Secondly, there are 122 observations without CUSIP. These observations must be thrown out as they cannot be matched up across data sets. However, this only accounts for less than one hundredth of a percent of total data and is negligible enough to not consider. The ending observation count comes out to be 168,861.

[3] September 2018 is the most recent date for data available from Compustat. January 1990 is chosen as the lower bound of date in order to keep the file sizes minimal while still collecting data from an ample amount of time.

3.1. Sample and Data

At this point, the data sets are both merged together, but the data has not been standardized. Data standardization is crucial, as firm size can vary drastically in the entire universe of recorded firms over the past thirty years. Comparing a nominal value of the balance sheet items of a relatively small company against that of a large company diminishes the impact of the smaller company's data on the model. As financial investment is scalable and looked at in terms of returns, the percent increase in price over time, small firm data should impact the model similar to large firm data. Common methods of standardization in financial literature is to divide financial information by *TOTAL ASSETS*. However, this research aims to approach financial standardization from the point of view of year over year change. Thus, each piece of numeric data is transformed by subtracting and then dividing by the previous year's observation.

The most immediate drawback to this method is the loss of an entire year's worth of data, as this method produces $n-1$ results for a starting n number of years collected due to the first year of data not having a previous year to compare to. The largest benefit of this method compared to the total asset view is the isolation of variables. If a firm experiences significant change in balance sheet items but a singular variable maintains its value, the percent of *TOTAL ASSETS* value of the variable will change even though the variable itself has not. In the year over year percent change method, the only changes that go into the multivariate regression model are from changes in the observed variables themselves and no other direct source. This also eliminates nominal values of data by classifying every observation as a percent change, thus accomplishing the original goal of standardization.

This method also equal weights firms regardless of their size. Fama and French specifically remove firms with under $25 million in *TOTAL ASSETS* resulting in models without the smallest of firms in the universe. This is due to the increased opportunity for volatility when a firm has low starting values. However, including these firms, of which there are few, should not adversely affect this percent change model. A higher volatility is still valid data as a large percent increase in an accounting item that also largely, or minimally, affects *STOCK RETURN* is useful information to throw into a model. In this way, extremes are still preserved. The one variable that is not converted to yearly percent change via the same formula is *MARKET RETURN*. *MARKET RETURN* refers to the value-weighted index returns, excluding dividends, on the S&P 500. These returns are retrieved monthly. The year over year percent change are calculated as follows:

$$R_{My,t} = \prod_{i=t-11}^{t}(R_{Mm,i} + 1) - 1 \qquad [4]$$

where $R_{My,t}$ refers to the yearly value-weighted index return at time t, and $R_{Mm,i}$ refers to monthly *VALUE-WEIGHTED S&P 500 RETURN* in month *i*. Industry is the final variable category added to this dataset. This is done according to Fama and French's twelve industry classifications based on SIC code. The occurrence of observations in each industry can be viewed in Figure 1.

3.2. Research Design

Now that the data set is established, work can now be done towards creating a multivariate regression model. The regression models in this research looks to test Fama and French (2015) five-factor model in a percent change format rather than a portfolio format. Since the current data does not establish matrix grouping, none of the previously mentioned factors can be used. Instead, accounting items are substituted for the portfolio-based factors of size, liquidity, profitability, investment, and market returns.

Since these are broad categories, multiple variables are tested per category. The size variables are *TOTAL ASSETS*, *TOTAL EQUITY*, and *COMMON SHARES OUTSTANDING*. The liquidity variables are *CURRENT ASSETS*, *RATIO OF CURRENT ASSETS TO TOTAL ASSETS*, and *CURRENT RATIO* defined by current assets divided by current liabilities. Profitability variables are *NET INCOME*, *REVENUE*, and *EARNINGS BEFORE INTEREST AND TAXES*. CAPITAL EXPENDITURES is the variable considered for investment. The variable considered at this stage for market returns is *VALUE-WEIGHTED S&P 500 RETURN*. It is very important to note that each of these variables are not nominal but in terms of percent change over time. These variables are more clearly seen below in Table 1.

A notable deviation is the lack of consideration to *MARKET VALUE* for size. *MARKET VALUE* is defined as the price of the stock of a firm multiplied by the number of *COMMON SHARES OUTSTANDING*. Since this research's primary model uses yearly returns as the dependent variable, adding an independent variable to the regression model that is directly built from returns would run against the point of this research. There is little point in predicting returns if the prediction relies on already knowing returns. In order to avoid this circular reference, *MARKET VALUE* is left out as an independent variable. Instead, it is later used as a replacement for returns as a dependent variable for the regression model. In its place, *COMMON SHARES OUTSTANDING* is singularly analyzed in order to not dilute the model's relevance.

Since prior research primarily uses the *BOOK-TO-MARKET RATIO* for profitability due to its proven strong relation to returns, *BOOK-TO-MARKET RATIO* is left out of the profitability variables and put in a category of its own. This keeps the ending model close in line with the five-factor model while still having the opportunity to include other profitability measures. Carhart's (1997) four-factor model also includes momentum in order to create a regression model for returns. In place of a momentum variable, *PRIOR RETURN* is considered in the list of independent variables. All possible combinations of variables are used, one from each category, to create multiple multivariate regression models. The base model can be summarized as follows:

$$R_t = \alpha + \beta_1 Size_t + \beta_2 Liquidity_t + \beta_3 Profitability_t + \beta_4 Investment_t + \beta_5 \frac{Book}{Market_t} + \beta_6 R_{M,t} + \beta_7 R_{t-1} + \epsilon_t \qquad [5]$$

Each β_n refers to the regression coefficient for the variable that it is next to. Size most commonly refers to how much a company possess in either held assets or assigned value. Liquidity is a measure of how quickly a company can turn holdings into cash. An item is typically considered liquid if it can be turned into cash within one year's time. Profitability refers to how a firm operates is assets or services and uses them to create capital. Investment refers to a firm's redirection of capital back into the firm.

Performing these procedures produces t-values and p-values for each variable in each iteration of regression. The mean t-values and mean p-values for each tested variable can be seen in Table 3. While non-weighted mean values can be misleading due to extreme values, the relatively small amount of data produced allows for a full manual review and no significant deviation is seen across model combinations for the same variables. The least significant variables in increasing order of average t-value are *REVENUE*, *EARNINGS BEFORE INTEREST AND TAXES*, *CURRENT ASSETS*, *RATIO OF CURRENT ASSETS TO TOTAL ASSETS*, and *NET INCOME*. Not only are each of these variables significant on average for the 27 combinations of models, each has a p-value greater than 0.5, apart from *NET INCOME* which is 0.3744. It can be said with confidence that these variables are not significant predictors of *STOCK RETURN* and are thus removed from further testing. It is worth noting that all tested measures of profit are deemed insignificant. It also goes against intuition as one would expect a similar change in profits to mean a similar change in company worth, and thus an increased expected *STOCK RETURN*.

However, it could be this two-degree removal from direct effect that causes these profit measures to be poor predictors of *STOCK RETURN*.

3.3. Preliminary Analysis

The most significant variables in decreasing order of the absolute value of average t-value are *VALUE-WEIGHTED S&P 500 RETURN* (30.79),[4] *TOTAL ASSETS* (8.68), *CURRENT RATIO* (8.25), *PRIOR RETURN* (8.59), *TOTAL EQUITY* (5.97), and *BOOK-TO-MARKET RATIO* (5.45). These are presented in order of the absolute value of average t-values, as all but *BOOK-TO-MARKET RATIO* has average p-value of less than or equal to 0.0001. While all six of these variables are significant, *TOTAL EQUITY* are not included in further models. Since both *TOTAL EQUITY* and *TOTAL ASSETS* measure firm size, it would be redundant to put the two together. *TOTAL ASSETS* has the higher t-value, it will be used as the stronger size factor. Also, *TOTAL EQUITY* has too great of a correlation coefficient with *BOOK-TO-MARKET RATIO* at 0.592. Thus, including *TOTAL EQUITY* alongside the *BOOK-TO-MARKET RATIO* would over fit the model with data that is too similar. These remaining five variables will remain in the multivariate model through the next testing phase. This base model can be described as follows:

$$R_t = \alpha + \beta_1 R_{M,t} + \beta_2 R_{t-1} + \beta_3 TA_t + \beta_4 \frac{CA}{CL_t} + \beta_5 \frac{Book}{Market_t} + \epsilon_t \qquad [6]$$

where TA refers to *TOTAL ASSETS* and $\frac{CA}{CL}$ refers to the current ratio. This model above yields an r-squared of 0.0303, an adjusted r-squared of 0.0302, and each variable is significant at an $\alpha = 0.01$.

[4] T-values for each variable are given in parentheses.

3.4. Factor Analysis

Five variables have been deemed reliable factors, but the data set still has other variables that have not been considered yet. In total, there are 95 other variables that are collected from Compustat or created from Compustat data. Each variable is tested individually; a multivariate regression model is created and run for each of the 95 variables in addition to the factors in the five-variate base model. Of these models, 14 have high significance from t-value that pass $\alpha = 0.05$. These 14 are then trimmed by removing the variables whose models use the lowest percent of total observations. The ratio of used observations to total observations available for the ending six variables have a tight range of 0.177 to 0.2170. These six variables, listed in decreasing order of their respective model's r-squared value, which are listed in parentheses, are weighted-average number of common shares outstanding in a year (*SHARES*) (0.051), sale of common and preferred stock (0.0481), income taxes paid (0.0427), income before extraordinary items (0.0318), return on assets (0.0305), and total liability (0.0303). Again, all these variables are in the format of percentage change from the previous year and are not in nominal terms.

Now that the variable consideration pool has been slimmed to six, models can efficiently be run on the industry level as well as the entire market. Thus, each of the six is run twelve times, one for each industry as defined by Fama and French's 12 industry classification. This will ensure that a holistic view of each variable can be evaluated. A variable that works well with the few largest industries and not at all with the rest may score well against the whole universe of firms. Such a variable would be exposed in this separate industry analysis.

The industries I examine include the following: consumer nondurables, consumer durables, manufacturing, energy, chemical, business equipment, telecommunications, utilities, retail shops, healthcare, finance, and other. Consumer nondurables consists of products that either expire, wear down quickly, or are frequently consumed such as food, alcohol, clothes, and fabrics. Consumer durables are items consumed at a much slower pace that are not repurchased quickly such as expensive electronics, furniture, and cars. Manufacturing firms create products that are assembled, created, or processed in order to increase value such as paper, planes, and machinery. Energy firms deal in the creation or distribution of energy sources such as oil, gas, and coal. Chemical firms deal in the creation, implementation, or handling of chemicals. Business equipment handles items that add value to businesses such as computers, software, and specialized equipment. Telecommunication specialize in communication via telephone and television. Utilities companies operate and distribute utilities for businesses and households. Retail shops sell products or services directly to consumers via wholesale, retail outlets, and laundromats. Healthcare oversees the creation and administration of health services and products such as the creation of medical devices or drugs. Finance specializes in holding, distributing, and transforming cash flows for households and businesses. Other consists of all firms not captured by the previous eleven industries such as mining, constructing, transportation, hotels, and entertainment.[5]

3.5. Empirical Approach

Of the remaining variables under consideration, *SHARES* came out as having the most instances of significance out of all six variables observed. *SHARES* (i.e., *CSHPRI*) is described by Compustat as "Common Shares Used to Calculate Earnings Per Share/Basic." After further investigation, this variable most commonly equates to a weighted average of common shares outstanding in a year. *SHARES* has a significance level that beats an $\alpha = 0.05$ in every single

[5] Industry examples and SIC classifications are courtesy of Ed deHaan at the University of Washington via http://faculty.washington.edu/edehaan/pages/Programming/industries_ff12

industry besides telecommunications. Furthermore, the models that use *SHARES* has a relatively high average adjusted r-squared of 0.0852. These are high adjusted r-squared values compared to those calculated with the previous iterations of modelling. *SHARES* also have a high average observation use rate. On average between the twelve industries, the *SHARES* model makes use of 29.56% of all data observations available. The final model resulting from using *SHARES* can be denoted as follows:

$$R_t = \alpha + \beta_1 R_{M,t} + \beta_2 R_{t-1} + \beta_3 TA_t + \beta_4 \frac{CA}{CL_t} + \beta_5 \frac{Book}{Market_t} + \beta_6 SHARES + \epsilon_t \quad [7]$$

The adjusted r-squared of this model comes out to 0.0508 and each variable has a significant p-value that passes at an $\alpha = 0.01$ level. The parameter estimates for this model are below in the first column of Table 4.

3.6. Return Analysis

A criticism of the current model is that it does not currently take stock splits into account. If a company decides to double the number of current shares outstanding, firm price will instantaneously be halved such that firm *MARKET VALUE* is unchanged. While it is correct to say that this results in a decrease in price by a factor of 50%, all shareholders are then in possession of double their previous number of shares. This means that while technically stock price experiences large negative returns, the practical return is actually zero. The same is true for reverse splits as well. For this reason, the same model has been run against *ADJUSTED RETURN*, instead of simple *STOCK RETURN*. These adjustments are made by the following equation:

$$R\_adj_t = (R_t + 1) * \frac{CFACPR_{t-1}}{CFACPR_t} - 1 \quad [8]$$

where R_adj_t refers to the *ADJUSTED RETURN* during period t, $CFACPR_t$ refers to the cumulative factor to adjust price in time t, and $CFACPR_{t-1}$ refers to the cumulative factor to adjust price in the period directly prior to t. The *CFACPR* is designed to be the divisor for price in order to generate adjusted price. However, as this data set is using *STOCK RETURN* instead of price, equation [8] is created algebraically to transform *STOCK RETURN*. Similar r-squared results are achieved when *STOCK RETURN* is swapped out for *ADJUSTED RETURN* in the final model. Keeping the independent variables of the multivariate regression the same and setting it equal to *ADJUSTED RETURN* yields even stronger results than those with *STOCK RETURN*. The adjusted r-squared rises to 0.0571 and the right-hand side variables are significant at the $\alpha = 0.01$ level. The parameter estimates for this model are given below in Table 4.

3.7. Sample Characteristics

Now that that all relevant variables are determined and interactions can be seen between one another, we shall observe each individual variable to get an idea of its distribution and overall characteristics. Figure 2 displays a histogram of adjusted prices. Adjusted prices have a visible right skew which is caused by percent change variables having a lower bound at -1 and no upper bound. In fact, the maximum of adjusted price can be seen in Table 2 and is 178.613. In fact, all variables besides *MARKET RETURN* exhibit abnormal maxima that far exceed that of the third quartile. The extreme values most likely come from small firms who have small holdings but can achieve abnormal percent increases when they suddenly move to be more capitalized or successful firms.

Most medians seem to be just above zero, apart from *BOOK-TO-MARKET RATIO* and the *CURRENT RATIO* which are slightly negative. All variables seem to be having positive yearly change across the dataset due to both the positive medians and the greater magnitude of the upper quartile than the magnitude of the lower quartile. The only exception to this is the *CURRENT*

RATIO. This is expected for these variables as the overall market has had longer periods of bear markets than bull markets in the past thirty years. This means that *MARKET RETURN*, *MARKET VALUE*, and *STOCK RETURN* should all be expected to be more positive than negative. The *CURRENT RATIO*, as it is a ratio of assets held, is not affected by the state of the market in the same way. A firm increasing in value will immediately affect the *CURRENT RATIO* unless firm directors choose for a change in asset allocation following a change in value. Therefore, the *CURRENT RATIO* should not be expected to have a positive median in this dataset. However, this also does not mean the *CURRENT RATIO* should be expected to be negative. It is simply not immediately affected by the rise in *MARKET VALUE*.

A clear outlier in the data group is the minimum of the *BOOK-TO-MARKET RATIO*. This is the only value whose minimum is below -1. This is because the book value of a firm is a calculated value and can drop below zero. This is a rare occurrence, which is why the minimum value for the change in book-to-value can be so low while the median stays relatively close to the rest of the variables observed. The *BOOK-TO-MARKET RATIO* is also an outlier because of its low count of observations. This is due to one of its components, *TOTAL EQUITY*, has an observation count of 47,212. *BOOK-TO-MARKET RATIO* can only produce a non-null value when a firm-year contains instances of both *TOTAL EQUITY* and *MARKET VALUE*, which has an observation count of 75,105. *TOTAL EQUITY* and *MARKET VALUE* are values pulled directly from Compustat, so the low count of *BOOK-TO-MARKET RATIO* data is from sourcing issues and not from manipulation.

On a higher level of analysis, it is important to discuss the normality of this data when considering the validity of the multivariate regression models that are being run. One would assume that percent change over time of accounting items would hold a normalized distribution over time when market movement is held constant. However, markets are not held constant over time and is why the center of most variable distributions is marginally greater than zero. The right and left bounds of each interquartile range are close to equidistant from the median, as is shown in Table 4. Exact normality distributions for each variable would be cumbersome to add to this paper, so the *ADJUSTED RETURN* distribution is listed in Figure 2. The shape of this distribution is similar in shape to the distributions of the other variables in the utilized dataset.

The models run cannot be confirmed to have normal error terms. With the high magnitude of data being observed and a wide array of price and factor changes, the models are not able to produce easily read residual charts. When observing such a figure on such a mass of data with over 160,000 firm-year observations, no clear shape or characteristics can be significantly determined. Further, due to such larger power, the statistical test is unlikely to be useful to assess the problem. Thus, normality of the multivariate regression models cannot be confirmed. However, given that this dataset consists of mostly normally distributed variables and that the observed data is a population and not a sample, the regression models can be trusted to give accurate information as far as their variable significance and model accuracy is reported. The regression models observed are not being run against market-like data or ideal subsets, but the full reporting of market data for the past twenty-eight years.

3.8. Industry Analysis

MARKET RETURN has the highest regression coefficient within each industry. At the same time, *MARKET RETURN* also consistently has the highest t-values for each industry. This variable is obviously the strongest predictor of *STOCK RETURN*, regardless of which industry is being observed. This makes sense as *MARKET RETURN* is composed of select firm returns that attempt to mimic the larger firm universe as fairly as possible. The industry that saw the lowest coefficient (0.462) and second t-value (5.79) for *MARKET RETURN* is utilities. Utilities is a highly regulated industry and is not allowed to vary its operations, subsequently restricting the variation of its returns, and is most likely why *MARKET RETURN* is weakest in this industry.

The *CURRENT RATIO* fails to be significant in ten of the twelve industries. This is surprising to see as the *CURRENT RATIO* has previously done well in the models that include the full universe of firms. One might assume that the *CURRENT RATIO* is a great predictor for most data, but not many industries. However, the largest industry represented, finance, only produces a t-value of -0.27 for the *CURRENT RATIO*. The two industries that it is significant in are manufacturing and consumer non-durables. These industries rely on the quick turnover of inventory in order to maximize profit. If an event arises with an immediate need for an influx of capital, a more liquid firm will be able to see to the need and return operations to normal in a timely fashion. This may be why the *CURRENT RATIO* is significant in these two industries.

PRIOR RETURN fails to hold significance in chemical and energy sectors. These industries are most likely driven by supply of raw product available and thus hold a volatility that is not associated with *PRIOR RETURN*, but some exogenous factor. For the energy sector, this could include crude oil prices and government action. The utilities industry experiences the highest r-squared out of any other industry observed. This may be due to utilities not having much volatility in its accounting items or valuations due to its strict government regulations and boundaries. On the other hand, energy has the lowest r-squared out of the observed industries. This is most likely due to exogenous variables as discussed earlier. Its volatility comes from external factors instead of the accounting items observed.

Consumer durables experiences only one variable, market returns, as *MARKET RETURN*, is significant to the 1% threshold and one, previous returns, as *PRIOR RETURN*, that is significant at the 10% threshold. Low significance to other variables, coupled with the relatively high r-squared value of 0.0723, may indicate that the two significant variables are most of what is needed to explain the variations in *STOCK RETURN* for consumer durables. Thus, consumer durables industry is market-driven and relies heavily on the state of other firms and slightly on the expectations from *PRIOR RETURN*.

The finance sector shows a remarkably low observation count of 991 observations read when compared to its total observation count in the dataset of 58,084. This means that equation [7], when observed against *ADJUSTED RETURN,* is due to the previously discussed low occurrence of *TOTAL EQUITY* throughout the data. In the finance industry, around 17% of all finance observations contain values for their change in *TOTAL EQUITY*. Since the *BOOK-TO-MARKET RATIO* relies on the occurrence of *TOTAL EQUITY*, a 17% observation usage rate is the maximum available for this model. The less than 100% occurrence rate of the other tested variables throughout the data only brings the used observation rate down further, along with the yearly change transformation eliminating more than one thirtieth of the full dataset. The compounding of these data imperfections, most notably the lack of *TOTAL EQUITY*, is what causes the unfortunately low observation rate for finance.

3.9. Alternate Dependent Variables

A model that has been fully developed and analyzed for studying *STOCK RETURN* may also favorably model other related variables. The variables that will be put under consideration alongside *STOCK RETURN* are *TOBIN'S Q* and *MARKET VALUE*. These two variables are similar to *STOCK RETURN,* because they are thought of as measurements of performance. *MARKET VALUE* is calculated as stock price multiplied by *COMMON SHARES OUTSTANDING*. Thus, a change in *MARKET VALUE* should follow the same changes in price when *COMMON SHARES OUTSTANDING* are held constant. However, shares do not hold constant in reality and can fluctuate with ease over the course of a year. This is where any differentiation will be seen between yearly changes in *MARKET VALUE* and yearly *STOCK RETURN*. *TOBIN'S Q* is calculated in the following manner:

$$Q = \frac{Total\ Assets - Total\ Common\ Equity + Price * Common\ Shares\ Outstanding}{Total\ Assets} \quad [9]$$

TOBIN'S Q is analogous to the market-to-book value of assets. Essentially, the markets evaluation of a company with respect to its actual holdings. A higher Q ratio means that the market is putting a premium on the value of a firm while a low Q ratio means that the market is valuing the firm at a discount. The coefficients, significance, and t-values for each variable in these two new models are reported in Table 4.

One of the easiest differences to notice in *TOBIN'S Q* model versus all other models ran is that the coefficient for *MARKET RETURN* is less than one. When considering both Table 5 and Table 5, this is the second lowest instance of *MARKET RETURN*'s coefficient. This could be related to *TOBIN'S Q* heavy reliance on *TOTAL ASSETS*. *TOTAL ASSETS* tend not to fluctuate with the market,[6] which is why *MARKET RETURN* would lose strength when predicting *TOBIN'S Q*.

The one variable whose coefficient is not significant for an $\alpha = 0.01$ is the *CURRENT RATIO*. This is less surprising given Table 4 in which the *CURRENT RATIO* is insignificant in

[6] When observing the correlation coefficient of market returns and *TOTAL ASSETS* in this data set, the p-value for the null hypothesis that the correlation is zero versus the alternative hypothesis that the correlation coefficient is not equal to zero came out to be 0.7874. The two are almost certainly not related when compared one-to-one.

many industries. However, the *TOBIN'S Q* model observes the entire universe of firms. This seems to indicate that change in firm liquidity simply does not play a pivotal role in determining the change in premium/discount that a market uses to value a firm.

Overall, the *TOBIN'S Q* model yields an adjusted r-squared of 0.0271. This is a very low number in comparison to the other models' r-squared values in Table 5. When switching the model form *ADJUSTED RETURN* to *TOBIN'S Q*, the r-squared drops by over half. *MARKET VALUE* is far more in line with the *ADJUSTED RETURN* model than the *TOBIN'S Q* model is. Every variable from the model yields a significant p-value and the r-squared rose to 0.0935. This large difference from the *ADJUSTED RETURN* model's r-square of 0.0585 is most likely due to the *BOOK-TO-MARKET RATIO* and weighted-average common shares outstanding. In the *BOOK-TO-MARKET RATIO*, the denominator of the ratio is simply *MARKET VALUE*. Thus, the *MARKET VALUE* model has *MARKET VALUE* on both sides of the equation to some extent. Technically, this is not a direct conflation of data as the transformation of nominal data to yearly percent change affects the ratio in such a way that an $x\%$ increase in *MARKET VALUE* would not yield an $\frac{1}{x}\%$ increase in *BOOK-TO-MARKET RATIO*. The algebra involved in taking two changing variables and creating a percent change ratio protects the model from having too much direct affect when a variable is in one form as a dependent variable and in another form as an independent variable.

Similarly, *MARKET VALUE* is just a ratio with *COMMON SHARES OUTSTANDING* in its denominator. Thus, weighted average common shares outstanding is most likely interacting with *MARKET VALUE* in the same way that *MARKET VALUE* interacts with the *BOOK-TO-MARKET RATIO*. In fact, the weighted average common shares interaction may be the stronger of the two as that variable has a large t-value of 33.16.

4. Conclusion

Fama and French (2015) and Carhart (1997) have well-crafted and celebrated models to predict firm returns using portfolio-based variables. In the conversion of portfolio-based variables to raw data, market returns, momentum, liquidity, profitability, and size all maintain their significance in the form of *VALUE-WEIGHTED S&P 500 RETURN*, PRIOR RETURN, *CURRENT RATIO*, *BOOK-TO-MARKET RATIO*, and *TOTAL ASSETS,* respectively. Five of the six variables agree, for the most part with previous literature. Size, liquidity, momentum, market returns, and profitability all represent in this research's empirical model. The only factor that did not match up is investment. The only variable considered in this research for investment is *CAPITAL EXPENDITURES*, as it is often the most straight-forwardly related variable to firm investment. Unfortunately, the firm level yearly percent change method of observing financial data does not prove *CAPITAL EXPENDITURES* to be significant in the prediction of *STOCK RETURN*. This could mean that investment is important for the calculation of *STOCK RETURN* in a nominal manner, such that the level of investment matters far more than the change in investment. In future research, research and development as well as acquisitions expense could be considered as alternatives for the investment factor.

Even though the models produced are not able to account for the same amount of variation as previous techniques have been able to, it does show that observing the full universe without portfolios still result in almost all the same factors being significant. The regression model search process is not built solely on finding a relation to returns that would have the lowest possible error. While purposefully minimized as much as possible, significant model error is unavoidable with this structure of data. Instead, the goal of these models is to create regressions that primarily find literature-like factors in non-portfolio terms that can match *STOCK RETURN* and secondarily to keep explanation of variance as high as the data would allow. This goal is achieved. The Fama and French (2015) factors of size, liquidity, profitability, and market returns along with Carhart's (1997) momentum are all verified to be significant with this different data structure.

Looking forward, this research could be augmented by expanding the dataset observed. While twenty-eight years of data are used in this model creation, many more years of data are available and could add more clarity to the strength of variables. At the same time, it would be interesting to observe how the significance of these factors changes over time by observing the model against decades of data, rather than by industry. In the same spirit, including foreign firms to the data set and observing the model by country could also yield telling results on the importance of basic firm characteristics to returns according to geography. Perhaps a strong predictability in utility returns is an American phenomenon and would not be replicated in foreign markets or changes in the *CURRENT RATIO* turns out to be much more significant in an Asian country. The analysis needed for these hypothetical claims would require a significant allocation of data but would hold satisfying results once the data is obtained.

References

Aharoni G., Grundy B., Zeng Q., 2013. Stock returns and the Miller Modigliani valuation formula: Revisiting the Fama French analysis. Journal of Financial Economics 110, 347—357.

Asness S., Frazzini A., Pedersen L., 2018. Economics 116, 1–22. Quality minus junk. Review of Accounting Studies 24, 34–112.

Carhart, M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57-82.

Fama, E., French, K.,1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E., French, K., 1997. Industry costs of equity. Journal of Financial Economics 43, 153-193.

Fama, E., French, K., 2006. Profitability, investment, and average returns. Journal of Financial Economics 82, 491–518.

Fama, E., French, K., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1—22.

Lyandres E., Sun L., Zhang L., 2008. The new issues puzzle: testing the investment-based explanation. The Review of Financial Studies 21, 2825—2855.

Novy-Marx, R., 2013. The other side of value: the gross profitability premium. Journal of Financial Economics 108, 1—28.

Table 1: Factor Model Empirical Design

Table 1 displays the 27 combinatorial options for creating firm-level, rather than portfolio-level, models. Each of the categories in the first row are the factors that Fama and French (2015) sought to explain with their portfolio-based variables. Under each are accounting variables that are considered indicators of their above categories. A model is created for each combination of variables possible, taking one from each column with each model. The variables deemed significant enough to be implemented into further models are italicized below. The *CURRENT RATIO* refers to the ratio of current assets to current liabilities.

Size	Liquidity	Profit	Investment	Market Returns
TOTAL ASSETS	CURRENT ASSETS	REVENUE	CAPITAL EXPENDITURES	*VALUE-WEIGHTED S&P 500 RETURN*
TOTAL EQUITY	RATIO OF CURRENT TO TOTAL ASSETS	EARNINGS BEFORE INTEREST AND TAXES		
COMMON SHARES OUTSTANDING	*CURRENT RATIO*	NET INCOME		

Table 2: Summary Statistics

Table 2 shows the summary statistics for each of the variables used in the primary analyses. The lower quartile refers to the 25th percentile of data and the upper quartile refers to the maximum value of the 75th percentile of data. Due to the extreme maxima of the dataset, median and quartile information are given alongside the mean and standard deviation. These values better represent the characteristics of most of the data. All variables besides the *BOOK-TO-MARKET RATIO* have a left bound of -1, which is why so many minima exist at or around this point.

Variable	N	Mean	Std Dev	Lower Quartile	Median	Upper Quartile	Minimum	Maximum
RETURN	147,207	0.179	1.646	-0.214	0.024	0.274	-0.996	172.091
ADJUSTED RETURN	147,188	0.141	0.944	-0.194	0.037	0.292	-1.000	178.613
TOBIN'S Q	124,315	0.083	6.269	-0.132	0.001	0.128	-0.998	2,166.174
MARKET VALUE	75,105	0.767	66.359	-0.223	0.059	0.393	-0.998	15,517.582
MARKET RETURN	164,051	0.097	0.154	0.054	0.119	0.204	-0.440	0.494
BOOK-TO-MARKET	40,754	0.080	12.358	-0.245	-0.033	0.253	-1,840.325	1,298.077
PRIOR RETURN	147,206	0.179	1.646	-0.214	0.024	0.274	-0.996	172.091
TOTAL ASSETS	125,635	1.106	293.691	-0.035	0.058	0.193	-1.000	10,3908.00
CURRENT RATIO	99,470	0.266	9.955	-0.193	-0.017	0.165	-1.000	1,550.332
SHARES	125,096	0.134	1.264	0.000	0.012	0.095	-0.997	354.000

Table 3: Five-Factor Model Alternatives, Significance Evaluation

Table 3 displays the average absolute coefficients and t-values (in parentheses) for each variable on the left throughout the 27 combinations of models tested in section 3.2. The average t-value is the average of the absolute values of all t-values. Each model is run with *RETURN* as the dependent variable. The Percent Positive column displays the percent of coefficient instances that were positive. The Percent Significant columns indicate what percentage of instances were significant at the respective level and not significant at other levels. Robust two-tailed t-statistics are presented in parentheses below the coefficients. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively. Significance is based on average p-value.

		Percentage of Models with			
	RETURN	Positive Coefficients	Coefficients Significant at 0.01	Coefficients Significant at 0.05	Coefficients Significant at 0.10
Constant	0.048*** (8.07)	100%	100%	0%	0%
MARKET RETURN	1.397*** (40.83)	100%	100%	0%	0%
BOOK-TO-MARKET	0.005*** (7.24)	0%	100%	0%	0%
PRIOR RETURN	0.025*** (7.46)	0%	100%	0%	0%
TOTAL ASSETS	0.05*** (10.13)	100%	100%	0%	0%
CURRENT ASSETS	0.024 (4.82)	33%	33%	0%	0%
NET INCOME	0.0000 (1.34)	100%	20%	0%	0%
REVENUE	0.0000 (0.58)	0%	0%	0%	0%
EARNINGS BEFORE INTEREST AND TAXES	0.0000 (0.78)	0%	0%	0%	0%
CURRENT ASSETS-TO-TOTAL ASSETS	0.034 (2.87)	40%	40%	0%	0%
CURRENT RATIO	0.011*** (4.22)	100%	83%	17%	0%
TOTAL EQUITY	0.006*** (7.59)	100%	100%	0%	0%

Table 4: Alternative Measures of Firm Value and Performance

Table 4 presents the results of running the previously created *ADJUSTED RETURN* model against *MARKET VALUE* and TOBIN'S Q as well. Robust two-tailed t-statistics are presented in parentheses below the coefficients. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively. The adjusted r-squared is given for each model at the bottom of the table. All variables are in the form of annual percent changes.

	RETURN	*ADJUSTED RETURN*	*MARKET VALUE*	*TOBIN'S Q*
MARKET RETURN	1.504***	1.398***	1.536***	0.741***
	(27.54)	(41.84)	(36.77)	(26.74)
PRIOR RETURN	-0.015***	-0.027***	-0.047***	0.024***
	(-2.89)	(-8.49)	(-11.80)	(-9.10)
BOOK-TO-MARKET	-0.003***	-0.003***	-0.003***	-0.002***
	(-3.50)	(-5.45)	(-5.02)	(-4.70)
CURRENT RATIO	0.044***	0.010***	0.041***	-0.004*
	(10.87)	(3.88)	(13.25)	(-1.89)
TOTAL ASSETS	0.050***	0.024***	0.078***	-0.024***
	(9.84)	(7.79)	(19.86)	(-8.11)
SHARES	-0.538***	0.088***	0.516***	0.051***
	(-26.41)	(7.04)	(33.16)	(4.98)
Observations	31,999	31,984	32,085	32,054
Adjusted R-Squared	0.0508	0.0585	0.0935	0.0271

Table 5: Industry Analysis

Table 5 presents the results of performing multivariate regression against *ADJUSTED RETURN* by industry. Robust two-tailed t-statistics are presented in parentheses below the coefficients. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively. The adjusted r-squared is given for each model at the bottom of the table. All these variables are still in the form of yearly percent change, except for returns which are already a measure of a change in price.

	Business Equipment	Chemical	Consumer Durables	Energy	Finance	Health	Manufacturing	Consumer Non-Durables	Other	Retail	Telecom	Utilities
MARKET RETURN	1.550***	1.095***	1.898***	0.896***	1.065***	1.484***	1.358***	1.155***	1.161***	1.355***	2.176***	0.462***
	(22.66)	(8.15)	(8.09)	(3.85)	(4.32)	(14.70)	(19.63)	(13.06)	(18.58)	(14.93)	(10.39)	(5.79)
PRIOR RETURN	-0.037***	-0.009	-0.043*	-0.016	-0.038*	-0.036***	-0.042***	-0.059***	-0.023***	-0.033***	-0.031**	-0.134***
	(-6.00)	(-0.55)	(-1.66)	(-0.98)	(-1.66)	(-4.94)	(-3.85)	(-3.10)	(-3.27)	(-2.78)	(-2.28)	(-4.53)
BOOK-TO-MARKET	-0.004***	-0.0776***	-0.001	-0.041***	-0.088***	-0.013***	-0.001**	-0.017***	-0.021***	-0.015***	-0.002	-0.092***
	(-2.62)	(-6.13)	(-0.23)	(-3.44)	(-4.25)	(-4.72)	(-2.03)	(-3.63)	(-7.14)	(-3.40)	(-0.88)	(-10.61)
CURRENT RATIO	0.019	-0.029	0.137	0.017	-0.002	-0.008	0.120***	0.066**	0.004	0.012	0.012	-0.003
	(1.38)	(-0.67)	(1.40)	(1.17)	(-0.27)	(-1.01)	(3.87)	(2.27)	(1.31)	(0.49)	(0.34)	(-0.17)
TOTAL ASSETS	0.188***	0.186***	0.147	0.214***	0.302***	0.106***	0.452***	0.326***	0.001	0.184***	0.115*	0.055**
	(11.81)	(3.51)	(1.42)	(3.66)	(3.75)	(9.43)	(11.63)	(7.84)	(0.26)	(4.40)	(1.88)	(1.96)
SHARES	0.090**	-0.022	0.19697	-0.326***	0.006	0.100***	0.178***	0.079*	0.068**	0.092	0.072	0.053
	(2.16)	(-0.34)	(1.27)	(-3.06)	(0.10)	(4.45)	(5.90)	(1.96)	(2.42)	(1.40)	(1.63)	(1.45)
Observations	6,960	998	912	2,077	991	4,404	3,605	1,833	5,012	3,379	1,001	812
Adj R-Squared	0.0979	0.106	0.0723	0.0209	0.0485	0.0818	0.1491	0.0505	0.08	0.0745	0.1065	0.1959

Figure 1: Industry Composition

Figure 1 displays the relative frequency of industries across all observations in the data set. The left axis indicates the percentage of the total data observations that belong to each industry. *FINANCE* refers to financial services firms, *BUSEQ* refers to business equipment firms, *HEALTH* refers to healthcare firms, *MANUF* refers to manufacturing firms, *SHOPS* refers to retail establishments, *NONDUR* refers to consumer nondurables firm, *ENERGY* refers to firms in the energy sector, *TELECO* refers to telecommunications firms, *UTILS* refers to utility companies, *CONSDU* refers to consumer durables firms, and *CHEM* refers to chemical companies, while *OTHER* includes the remaining firms. *FINANCE* is most notably the most common industry in the dataset.

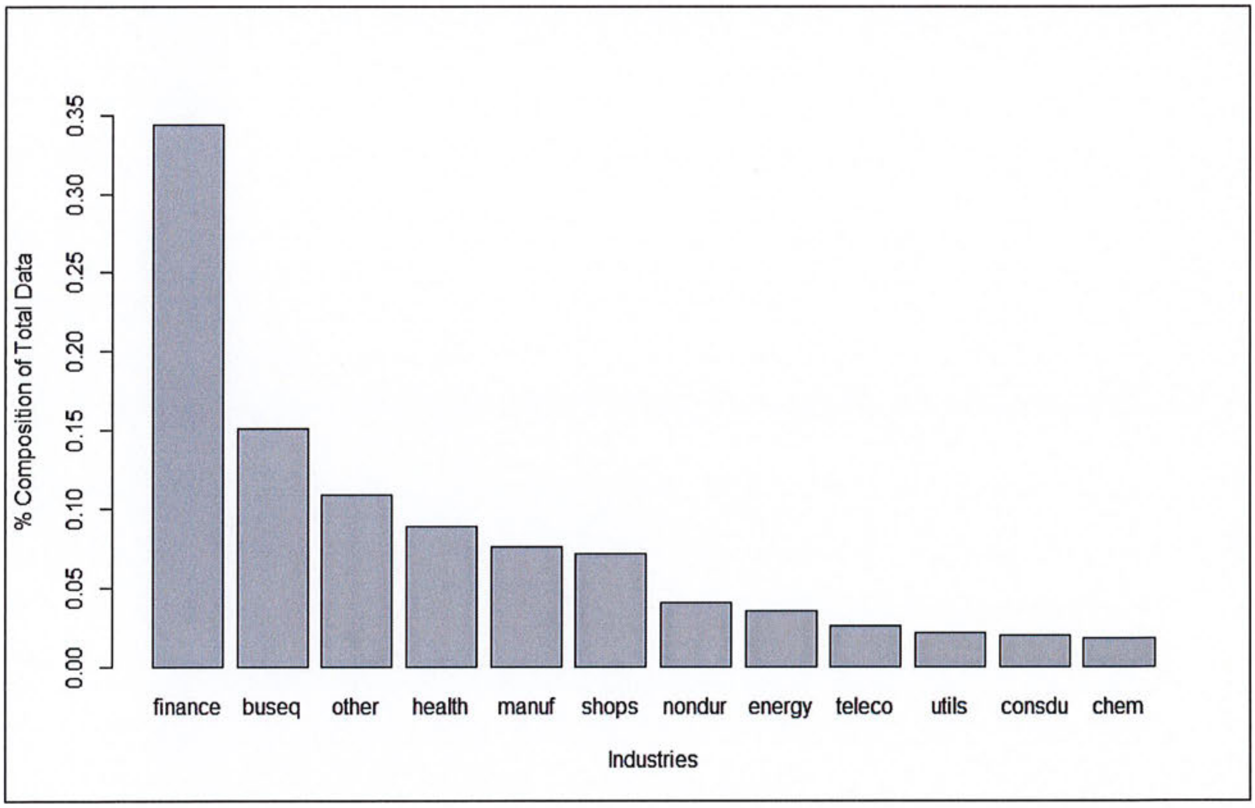

Figure 2: Distribution of Percentage Changes in Adjusted Price

Figure 2 displays a histogram for the values of adjusted price. For readability, the rightmost bin does not share the same width as the rest, but instead accounts for all values greater than 1.8. The distribution is right skewed as the variable has a left limit of -1 but no such right bound. The most that a stock price can lose is all its value but can theoretically gain ad infinitum. The data is centered around zero, if not slightly to the right of zero, and is relatively evenly distributed on either side, apart from the skewness.