Board Characteristic Dynamics:

An Investigation on the effect of Composition of a Company's Board of Directors on Resulting Performance

Joseph Leung, Gabriel Smith, Samuel Hsiung

HLS GROUP L.L.C.
4355 Imperial Way
Provo, UTAH, U.S.A
December 12, 2019

Contents

1 Introduction

This project aims to explore the effect that those most involved in the strategic decision-making process, namely the CEO and their fellow board members, have on the overall performance of a company. Though this is a question that would pique the interest of many companies on an international scale, relatively little research has been dedicated to understanding that relationship, in large part because of the enormity of the question and the lack of data to support strong analyses beyond specific case studies on certain companies. Useful data is further limited by the often private nature of company board meetings. However, digital resources allow us better understanding of the company metrics that can be used to tackle this question empirically. Though board members and a company's CEO cannot be judged in great detail individually, measurable characteristics of that board and/or CEO allows us to observe how a company's performance can be swayed by those characteristics.

1.1 Background

Of the research that does exist on this question, most fail to use much statistical sophistication. A paper[1] cited from the Harvard Business Review has suggested that for the last 60 years,the impact of a CEO has increased until recently where it has leveled off, though fails to explain why.[2] They do this through an ANOVA analysis and observe how variables such as return on sales vary as a result of the "CEO effect," which they isolate by comparing a company's performance against other similar companies. Of the valid resources we came across, this was the only study that used some statistical tool to analyze the question at hand. Other resources we found have suggested other interesting ideas, such as a "negative and significant relationship with the independence of boards",[3] where the independence of boards suggests no material interest in a company. A consulting firm recently analyzed whether the tenure of a given company's CEO helped determine its success, suggesting that executives staying in their position longer than 15 years tended to associate negative outcomes with external factors, often leading to poor performance in the future.[4] Last but not least, an article in the Wall Street Journal found that the optimal board membership should be roughly seven people, where membership greater than nine people can make the board too big to function.[5]

While these are interesting possibilities that may truly affect different indicators of performance of a company, they often rely on case-studies, or general summary statistics.

[1] "Has the "CEO effect" increased in recent decades?" (https://tinyurl.com/yx2gmdkc)
[2] https://hbr.org/2014/03/research-ceos-matter-more-today-than-ever-at-least-in-america
[3] https://www.tandfonline.com/doi/full/10.1080/1331677X.2018.1436454
[4] https://clarkstonconsulting.com/insights/impact-ceo-tenure-companys-success/
[5] https://www.investopedia.com/articles/analyst/03/111903.asp

This project considers many of those same variables like board independence or board size and their effect on a company's performance, but attempts to analyze them more rigorously with regression analyses and see what other potential variables ought to be taken into consideration to understand the relationship in question. Though there are recognizable limitations with such a method, it does allow this study to build off the given domain knowledge without losing sight of the question at hand, as there is still much ambiguity with the overall relationship between those involved in the strategic decision-making process and a company's performance.

1.2 Project Purpose

It follows that the purpose of our project can be summarized as an effort to determine the effect of a company board's culture on a company's performance, with the culture being identified by characteristics that describe the board.

While many indicators of company performance could be used, we look specifically at a company's revenue, profit, their respective growth, and return to investors over the last 10 years. Much of these dependent variables were selected for us, as data on a company's performance is limited. Nevertheless, these are measurements that are often of most interest to investors to a company, and provide a general range of that company's overall health.

This project extends into a larger idea of how the social dynamic, rather than just characteristics, of those involved with strategic decision making affect a company's resulting performance. By analyzing how different characteristics affect different company indicators, we uncover information that can help provide parameters to develop an algorithm predicting how the decision itself is formed in based on who makes up the board membership given the context of the firm's goals. As this end goal would require far more in-depth Machine Learning tools, we focus just on the general relationship between the characteristics of a company board and that company's performance.

1.3 Hypothesis

We test the effect of various board characteristics such as the ranking of a company's CEO and the Board's diversity on various metrics representing the performance of a company, such as profit or revenue. This can be represented by the following model:

$$CompanyIndicators_i = \beta_0 + \sum_{j=1}^{n} \beta_j BoardCharacteristics_{i,j} + \sum_{k=i} \beta_k \eta_k + \epsilon_i$$

Where η_k represents the different variables which need to be controlled for, such as a company's industrial sector or geographical location corresponding to β_k. More information on this can be found in part A of the appendix and under the Feature Engineering section of this paper.

We have provided an example (Figure 1) of the types of regression we are running, where in this case we have regressed the Median Age (*Board Characteristcs*) with yearly Profit Growth (*Company Indicators*).

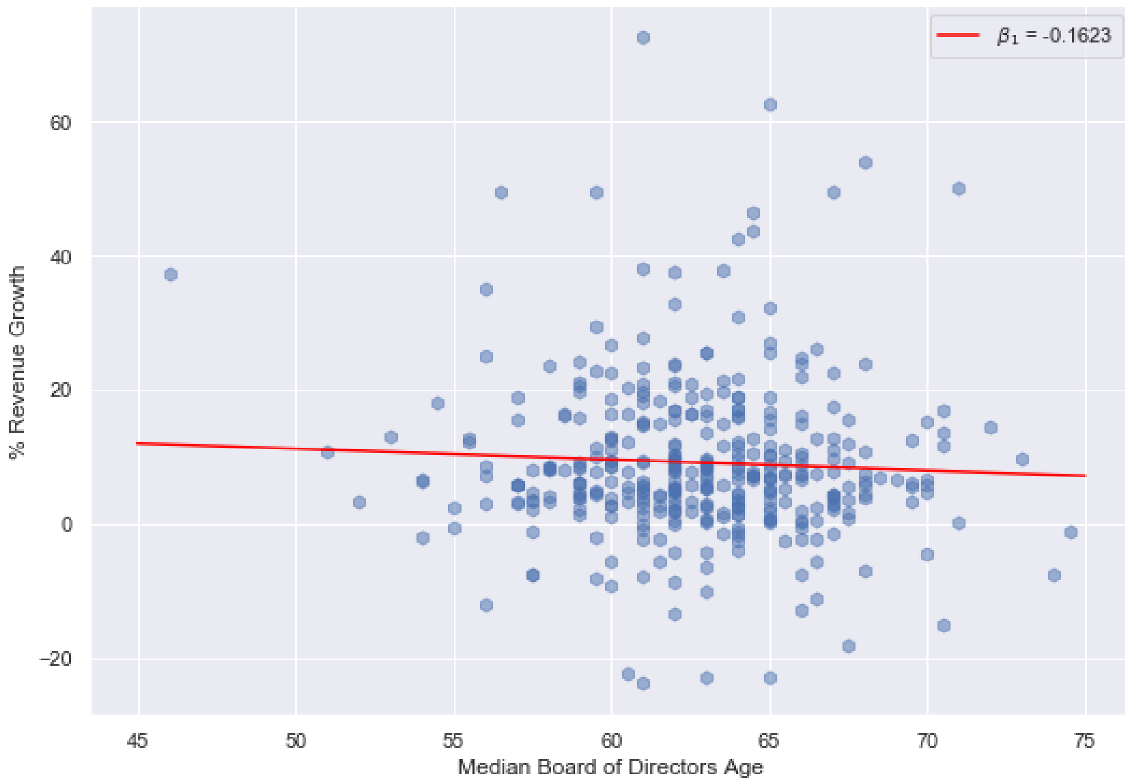

Figure 1: As the Median Age of Boards increase Revenue Growth declines

2 Data

2.1 Overview of Data Origin

The data comes from a variety of sources which were scraped from the web. Details on the companies came from the 2019 Fortune 500 list[6]. This source provided this year's Fortune 500 companies' revenue, profit, return on investment and various other company statistics. As this is often the key reference that most businesses use to evaluate successful companies, we feel that it provides valid information for our specific project. We also collected

[6]https://fortune.com/fortune500/2019/search/

information on board members from an interactive graphic found on the Wall Street Journal[7], a news source whose quality has often been esteemed. Finally, we obtained CEO rankings from two different popular and accredited business sites, the Harvard Business Review[8] and Glassdoor.com[9]. While we had sought more potential rankings, we noticed that almost all sources actually referred to these two. A useful contrast that came into play with these two ranking styles was the traditionalism of HBR, and the new digital interface of Glassdoor. As we will see later, being ranked on either of these sources offer interesting insights. These sources allow us to analyze our question in sufficient detail.

2.2 Data Collection and Cleaning

Due to the uniqueness of the data and their considerably well-renowned sources, much of the difficulty in organizing the data came from scraping it from their given websites, and then cleaning them to be in a comparable format. We had ensured to follow the websites' legal requirements for web scraping according to their robots.txt files, but scraping was not a convenient process. Fortunately, this implied that little needed to be done in terms of feature engineering as much of the data we sought was already provided by our various sources, and other potentially interesting information could not be provided by the scraped information regardless.

2.3 Scraping

We primarily relied on data scraping to retrieve our data. We built via Python's selenium package a web crawler that would click through each of the 1000 company profiles from the Fortune 500 list (Fortune 500 is actually a list of the largest 1000 companies based on their annual total revenue). Each page contains the CEO name, company profits, revenues, growth rates, market valuations, geographical location, and more. See the following code block to view our scraping procedure to obtain the Fortune 500 data, and part C of the Appendix to observe select pieces of code used to scrape data from other resources.

Code used to Scrape Data from Fortune 500 Companies

```
'''
This block of code collects the company data as found on the fortune 1000 company. It needs to
    ↪ cycle through 10 different tables (which was formatted in a sort of HTML web tool),
    ↪ collects the hyperlinks and data from those tables, and then goes into the hyperlinks to
    ↪ collect more detailed but necessary information.
```

[7]http://graphics.wsj.com/boards-of-directors-at-SP-500-companies/
[8]https://hbr.org/2019/11/the-ceo-100-2019-edition
[9]https://www.glassdoor.com/Award/Top-CEOs-LST_KQ0,8.htm

```
'''
## Execute our webdriver object for chrome and initialize our base URL
browser = webdriver.Chrome('chromedriver.exe')
base_url="https://fortune.com/fortune500/2019/search/"

page_data = [] #data found on each page
CEO_data = [] #data found on the page describing the CEO and further details
try:
    browser.get(base_url)
    next_page = browser.find_elements_by_xpath('//button[@type="button"]')
    sleep(5)
    for p in range(10):
        print(p)
        browser.get(base_url)
        next_page = browser.find_elements_by_xpath('//button[@type="button"]')
        sleep(5)
        for i in range(p):
            #find the next button, and move to the page we want to be on
            next_page = browser.find_elements_by_xpath('//button[@type="button"]')
            next_page[-3].click()
        #find the hyperlinks
        hl = browser.find_elements_by_class_name('searchResults__cellWrapper--39MAj')
        hyperlinks1 = [w.get_attribute('href') for w in hl]
        hyperlinks2 = []
        for h in hyperlinks1:
            if h not in hyperlinks2:
                hyperlinks2.append(h)

        page_data.append([w.get_attribute('text') for w in hl])
        #getting the CEO data
        for h in hyperlinks2:
            browser.get(h)
            sleep(5)
            stuff = browser.find_elements_by_class_name('dataTable__row--34F3j')
            CEO_data.append([s.get_attribute('innerHTML') for s in stuff])
```

The data from the Wall Street Journal (WSJ) includes a collection of 500 companies' board of directors information including characteristics such as median age, percent female, median salary, and number of directors. This data is contained within an interactive web tool that stores the underlying data, which extracted by isolating the data sources and scraping using the XPaths. The essentials of the code to implement this task can be found in part C of the appendix.

Harvard Business Review (HBR) and Glassdoor (GD) CEO rankings consisted of the CEO's name, the company they belonged to, their ranking, and their overall approval rating. The information is provided all in one page but not in a downloadable format so we proceeded to build a scraper to pull that information. The code for implementation can also be found in part C of the appendix.

2.4 Cleaning

2.4.1 Fortune & WSJ Data

After pulling all of the data using our code above, we dropped a few unneeded columns with repeated information such as CEO names, unclear indices and unnamed columns. As we intended to merge the data together, company names needed to not just be cleaned but also normalized to merge all tables into one master file. To do this, we removed special characters, spaces, and put all the names as lower case using regular expressions (regex). There were difficult exceptions, as on occasion parent companies would take the place of a sub-company in one list, but not the other. This issue could not be resolved easily, but as it was not a very common occurrence, we took care of these issues manually. Another issue that was resolved computationally were slightly different forms of a company's name. For example, in one ranking a company may have been called "Company [insert name here]" whereas in the other it was just identified by "[insert name here]". To treat this, we carefully worked to remove these kinds of filler words, making sure that doing so would not erase the actual name. A final detail to clean the data was to fill in null values with zeros, as most of these null values were quantitative variables and implied things such as no change in value. Additionally we scaled the value "Median Pay" to be in thousands of dollars and turned variables like "Portion of Women on Board" into a percentage value to bring units closer together for easier direct comparison.

2.4.2 Glassdoor and Harvard Business Review

Due to our scraping code, each value often contained extra characters unneeded for our computation and analysis. We used regular expressions to delete all unnecessary characters from our data set, similar to our cleaning process for the WSJ data set. We then cleaned the company names as they appeared in these rankings to match with our normalized company names as previously designed to consistently merge each data set.

2.5 Feature Engineering

As a result of a comprehensive data scraping and cleaning portion, slight feature engineering proved necessary to coalesce the data for proper analysis. It was clear several categorical variables needed to be one-hot encoded. A particular issue that drew concern was the fact that much of a company's performance could very easily be biased by external factors such as the geographic location a company was located in, the sector they belong to, and other exogenous details. For example, financial companies in New York or tech companies on the Pacific Coast have seen explosive growth that could have been immensely affected simply by the fact that a company is a tech company or located in Manhattan. One example can be seen with Figure 2 that depicts the mean profit of a company by geographic division, as defined by the U.S. Census Bureau[10].

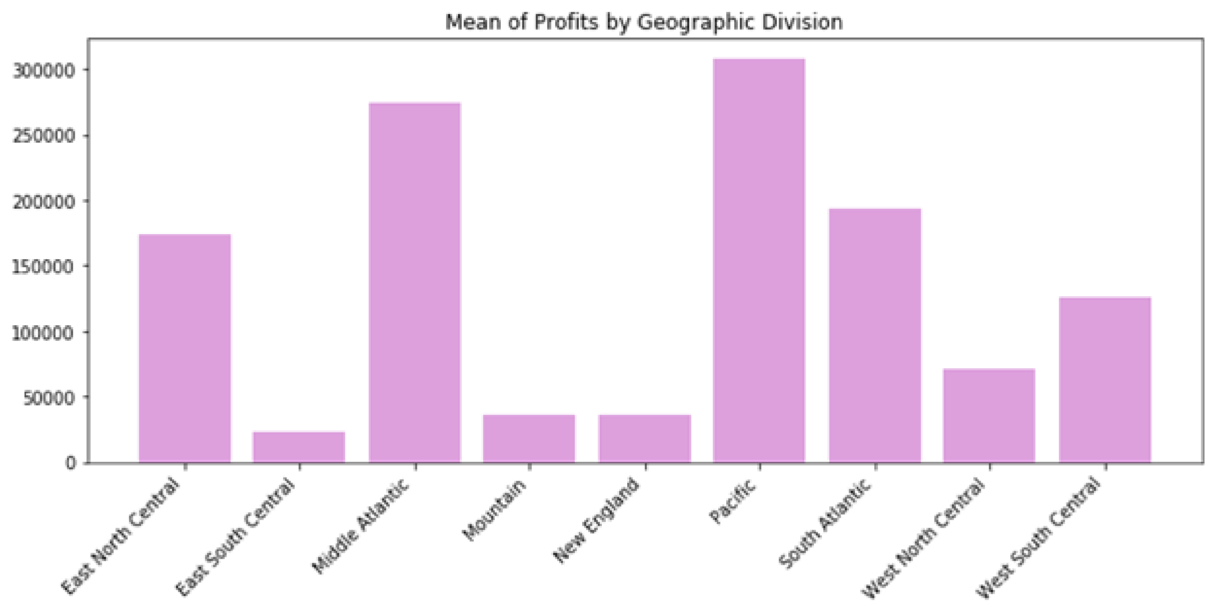

Figure 2: Distribution of Mean Profit by Geographic Division

In Figure 2 we note that companies in the Pacific or the Middle Atlantic United States have significantly higher mean profit, thus if we were to analyze a company's board characteristics in relation to that company's profit, without accounting for where that company is located, our analysis could be quite skewed. Figures 3-6, as found in part B of the appendix, show other visualizations and offer descriptions that help us understand what variables we will want to contain in our analysis to avoid omitted variable bias, though the reasoning is the same. Ultimately, we chose to control for a company's industry sector, their geographic division, and overall market size which was broken into categories of <\$250 million, \$50 million ∼ \$1 billion, \$1 ∼ \$5 billion, \$5 ∼ \$50 billion, and \$50 Billion+ USD.

While adding all these control variables, the issue of over-fitting may occur instead.

[10] https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf

While this is certainly a possibility, it seems unlikely to be a significant problem. There are countless variables that could be incorporated to answer our proposed question, and we are ultimately only looking at a few of them as we determine what board characteristics may have the greatest impact on a company's performance. In other words, we have a very noisy problem, and it is unlikely that accounting for variables such as geographic location will fully account for that noise. Thus, we feel justified with our large selection of control variables.

One additional variable which we did engineer was whether a CEO was ranked on both HBR and Glassdoor. While is reputable for a CEO to be found on a top 100 ranking at all, it felt significant to see if a CEO was found on both rankings.

Our final data can be seen below with two main categories of data: Company Performance[11] and Board Characteristics[12]

	Revenues ($M)	Revenue Growth ($M)	Profits($M)	Profit($M) Growth	ROI($M) (2018)	ROI($M) (5 year)	ROI($M) (10 year)
count	990.0	987.0	989.0	863.0	934.0	849.0	747.0
mean	15455.765	11.037	1253.994	179.835	-11.874	3.685	13.32
std	33479.02	17.62	3693.792	2514.758	26.825	13.691	10.502
min	1955.5	-34.3	-22355	-5538.6	-93.6	-55.0	-58.6
25%	3049.475	2.7	119.7	-30.0	-29.275	-3.6	7.9
50%	5599.95	7.6	363.4	12.2	-13.35	5.0	13.7
75%	12918.975	15.65	1059.3	53.3	4.1	12.6	19.45
max	514405.0	222.8	59531.0	66600.0	153.2	54.3	67.2

Table 1: Summary of Company Performance Indicators

From Table 1, we see that companies in our data set are on average larger companies with annual revenues of $15+ billion. Hence, our data analysis may not necessarily apply directly to smaller companies or startups. Despite that, the majority of our observations are household-name companies with large impacts on the economy of the United States, thus allowing for a qualified analysis.

	No. Directors	Med. age	Board Ind.	Med. Tenure	Med. pay	%Women	HBR Ranked	GD Ranked	GD&HBR Ranked
count	500.0	500.0	500.0	500.0	500.0	500.0	1262.0	1262.0	1262.0
mean	11.018	62.732	0.81534	7.608	265662	0.19814	0.079239	0.079239	0.005547
std	2.104884	4.786857	0.107376	3.614619	120042.7	0.088607	0.270219	0.270219	0.074299
min	5.0	-1.0	0.39	-1.0	-1.0	0.0	0.0	0.0	0.0
25%	10.0	60.5	0.75	5.0	220700.0	0.13	0.0	0.0	0.0
50%	11.0	63.0	0.83	7.5	255100.0	0.2	0.0	0.0	0.0
75%	12.0	65.0	0.9	10.0	294700	0.25	0.0	0.0	0.0
max	24.0	76.0	1.0	22.5	1000000.0	0.54	1.0	1.0	1.0

Table 2: Summary of Board Statistics

These summary statistics on the board membership in Table 2 do provide a few general details that give us an general picture of a typical board of directors. Board of directors on average have about 11 people, are male dominated, are comprised by roughly

[11] Table 1
[12] Table 2

80% independent members and are around 60 years-old. Interestingly, few CEOs of these companies are ranked in either HBR or Glassdoor.

3 Analysis

We recall our original hypothesis as modeled by the equation below:

$$CompanyIndicators_i = \beta_0 + \sum_{j=1}^{n} \beta_j BoardCharacteristcs_{i,j} + \sum_{k=i} \beta_k \eta_k + \epsilon_i$$

To analyze this, we use OLS regression analysis to observe the coefficients that would help us evaluate the effect of the board characteristics on a company's indicator. The final results of our regression analysis can be found in table 3, showing the effects of board characteristics on various company variables indicative of performance. A few considerations were taken when performing the regression, mainly having to do with the order of the variables if we were to omit a characteristic such as the percentage of women on the board. As we see in table 5, we ultimately see that once we add a few characteristics, the coefficient remains more or less the same, and measure of information content like AIC are relatively the same. This led us to conclude that enough of an effect on a company indicator was absorbed by the controls, and thus table 3 is the most appropriate analysis for our question.

	Profits	Profit Growth	Revenue	Revenue Growth	ROI (2018)	ROI (5 years)	ROI (10 year)
const	-6.451e+02	1.386e+01	-4.677e+04	**3.344e+01*** **	5.760e-01	1.073e+01	**2.169e+01*** **
	(4.948e+03)	(3.802e+02)	(3.855e+04)	(1.164e+01)	(2.160e+01)	(1.050e+01)	(7.857e+00)
HBR Ranked	-6.106e+02	9.808e+01	**-1.690e+04***	-7.715e-01	**7.535e+00***	**6.512e+00*** **	**4.859e+00*** **
	(1.103e+03)	(8.110e+01)	(8.596e+03)	(2.595e+00)	(4.549e+00)	(2.195e+00)	(1.610e+00)
GD Ranked	5.691e+02	3.233e+01	-8.184e+03	**6.077e+00***	-4.479e+00	3.968e+00	3.013e+00
	(1.337e+03)	(9.824e+01)	(1.042e+04)	(3.145e+00)	(5.511e+00)	(2.661e+00)	(2.017e+00)
GD & HBR Ranked	**6.943e+03*** **	-2.303e+02	**5.612e+04*** **	-4.364e+00	-1.475e-01	3.839e+00	-3.105e+00
	(2.672e+03)	(1.972e+02)	(2.082e+04)	(6.286e+00)	(1.102e+01)	(5.309e+00)	(3.956e+00)
Median pay (in Thousands)	**4.731e+00****	-1.091e-02	-9.140e+00	**1.215e-02****	-1.168e-02	**-1.156e-02****	7.796e-04
	(2.357e+00)	(1.778e-01)	(1.836e+01)	(5.545e-03)	(9.719e-03)	(4.762e-03)	(3.532e-03)
Median age	2.057e+01	-1.495e+00	**7.091e+02***	**-2.475e-01****	-9.607e-02	-1.217e-01	-8.477e-02
	(5.188e+01)	(3.901e+00)	(4.042e+02)	(1.220e-01)	(2.142e-01)	(1.034e-01)	(7.793e-02)
Women on Board (%)	4.956e+01	1.886e+00	**5.306e+02****	-1.019e-01	-3.827e-02	-9.591e-02	**-8.586e-02***
	(3.261e+01)	(2.559e+00)	(2.541e+02)	(7.672e-02)	(1.346e-01)	(6.636e-02)	(5.070e-02)
No. Directors	**-3.128e+02****	-5.112e+00	-4.286e+02	-4.889e-01	2.314e-01	-3.669e-01	**-6.686e-01*** **
	(1.375e+02)	(1.033e+01)	(1.071e+03)	(3.235e-01)	(5.687e-01)	(2.751e-01)	(2.076e-01)
Board Independence (%)	2.093e+03	-2.755e+01	1.258e+04	3.562e-01	-6.630e+00	6.755e+00	-2.596e+00
	(2.610e+03)	(1.990e+02)	(2.034e+04)	(6.141e+00)	(1.078e+01)	(5.267e+00)	(4.008e+00)
R^2	0.35	0.13	0.34	0.22	0.32	0.48	0.47
AIC	7392.97	4950.59	8920.44	2890.08	3281.04	2712.62	2404.67
BIC	7553.65	5103.51	9081.11	3050.76	3441.38	2872.51	2563.20

(Standard errors) * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Regression Board of Characteristics on Various Endogenous Variables

3.1 Key Insights

The interesting finding we see with our analysis is how different board characteristics affect different company metrics. We find that different aspects of a board are significant to a company in one sense, though insignificant in another. Much of our analysis is not comprehensive, thus we are unable to ensure correlation implies causality for various of our analyses. However, we break down some of the key insights which derive from Table 3 as follows:

3.1.1 Gender Diversity Impact

Our analysis suggests that gender diversity within a company board positively affects a company's revenue, but negatively affects the 10-year return on investment. For every percentage increase in the number of women on board, we see that a company's revenues increases by approximately $530 million. This may be the case for two particular reasons:

1. Companies with larger revenues have higher percentage of Women on their Board.

2. Higher percentage of women on their board help companies increase their revenue.

Although both of these are potentially true, we believe it will be interesting to see that more mature companies with larger revenues see the value of having more women on their board to increase diversity. Higher diversity would in theory improve company culture and promote the growth of the company.

In contrast, a percentage increase in the women on board implies an approximately $8,500 dollar decrease to return to investors over a 10 year period. This is a more puzzling statistic that we lack the domain knowledge to explain. Regardless, the fact that this is a significant coefficient implies this is worth exploring in greater detail in future studies.

3.1.2 Compensation Incentives

We found that median pay positively affects a company's profits and revenue growth, but negatively affects return to investors over a five year period. We scaled median pay to be in thousands of dollars, thus for every additional $1000 dollars to the median salary, we see an approximately $4.7 million increase to profits, $12,000 dollar increase to revenue growth, and $12,000 decrease to a five-year return to investors. This makes sense on a broader scale as higher salaries could mean:

1. Higher Caliber, more experienced leaders cost more in the labor market

2. Higher Compensated Leaders are likely motivated to deliver continually improving results

3. Tenured board members with higher salaries may likely have a longer vision in mind, and thus are less conscientious of fairly recent investors in lieu of more personal, long-term goals.

While there are certainly limitations to this relationship (an overwhelmingly large salary will ultimately harm a company and likely be probed for corruption), this general pattern makes relative sense. The more a board member is rewarded for creating good strategy for the company, the more the company is likely to receive beneficial attention.

3.1.3 Aging Leadership

Median age has significant negative effect on revenue growth but a positive effect on revenue. For every additional year to the median age of the board, revenues increase by approximately $700 million, and the revenue growth is noted to decrease by approximately $250,000. There may be many reasons why this is the case. One could be the fact that older companies whose growth has stabilized may have an older board since the conception of the company. Another could be that older board members may seek to maintain a company rather than help it grow. Unfortunately, though this is a significant result, it tells us little as to why the result may hold. Nonetheless, it is an important detail that holds a nontrivial impact.

3.1.4 Superstars

In most cases, we see a ranked CEO has some effect, but not always a significant one. Being ranked on HBR seems to imply great results for investors, but little else. However, we see that for the company itself, if a CEO is found to be ranked on both Glassdoor and HBR, there are extremely significant and bring positive results for their company's revenue and profit. According to the coefficients, being found on both Glassdoor and HBR results in a nearly $700 million increase in profits and $5.6 billion increase in revenues. However, our data is extremely limited, where only 7 individuals were found to be ranked on both HBR and Glassdoor. These people were comprised of mainly household names like Tim Cook, CEO of Apple. This leads us to believe that there could possibly be a "superstar" effect, where an astounding leader does have the ability to truly transform a company. While richer and more detailed data would certainly be needed, our analysis does prompt this possibility.

3.2 Consideration of Reverse Causality

While our OLS analysis provided interesting results with reasonable suggestions of causality, it is quite possible that reverse causality is actually occurring. That is, rather than board characteristics affecting company performance, the composition of a company board may actually be a result of simply belonging to a fantastic company. Companies with as much prestige as Google or Apple may face public pressure to incorporate fairer equality practices, thus affecting who ultimately makes its major strategic decisions. Thus, we consider two models in particular below:

$$CEO\_Rank_i = \beta_0 + \sum_i \beta_I Growth\ Indicators_i + \sum_j \beta_j Board\ Diversity_i + \sum_k \beta_k \eta_k + \epsilon$$

$$Board\ Diversity_i = \beta_0 + \sum_i \beta_I Growth\ Indicators_i + \sum_j \beta_j CEO\_Rank_j + \sum_k \beta_k \eta_k + \epsilon_i$$

Where η_k represent the controls as in the original hypothesis. A logistic regression for the first model and OLS for the second quickly demonstrated no significant variables, however, thus implying that board and CEO characteristics impact a company's performance, and not the other way around.

There are limitations with this analysis. Among the 1000 companies in our data set, only around 60 of those companies have a CEO that ranks among the HBR and/or Glassdoor CEO rankings. Evidently, we do suffer from a small sample bias in this general experiment. Provided richer data, our results would be more accurate and potentially yield different insights. Nonetheless, we consider this result to be valuable given the general lack of information and research done to explore the relationship between a company's executive board members and its performance.

4 Conclusion

Through our analysis and research, we were able to reaffirm that board dynamics do have a significant impact on firm performance. Although the direct mechanism leading to greater performance is not clear, it is proven that there is indeed an impact. A strengthened understanding of board dynamics and characteristics will allow further research into the potential mechanism and potential biases in decision making contributing to growth or decline in firm economics. Our research concludes that there do exist identifiable characteristics that play a role in firm policies and growth, and that future research can use these characteristics to define growth mechanisms and change leadership structuring decisions.

5 Appendix

5.1 A: Controls

The categories that made up our control variables were as follows:

Division	Market Size	Rank	Sector	Sector (cont.)
East North Central	250M	HBR	Aerospace & Defense	Household Product
East South Central	250M $\sim 1B$	GD	Apparel	Industrial
Middle Atlantic	$1 \sim 5B$	GD & HBR	Business Service	Material
Mountain	5B $\sim 50B$		Chemical	Media
New England	50B+		Energy	Motor Vehicle & Parts
Pacific			Engineering & Construction	Retailing
South Atlantic			Financial	Technology
West North Central			Food & Drug Store	Telecommunication
West South Central			Food, Beverages & Tobacco	Transportation
			Health Care	Wholesalers

Table 4: Control Variables

One way to see why it is important to control for these variables is visually to observe the differences between the categories within the variables. As we can see the mean revenue and profits tend to differ significantly by category.

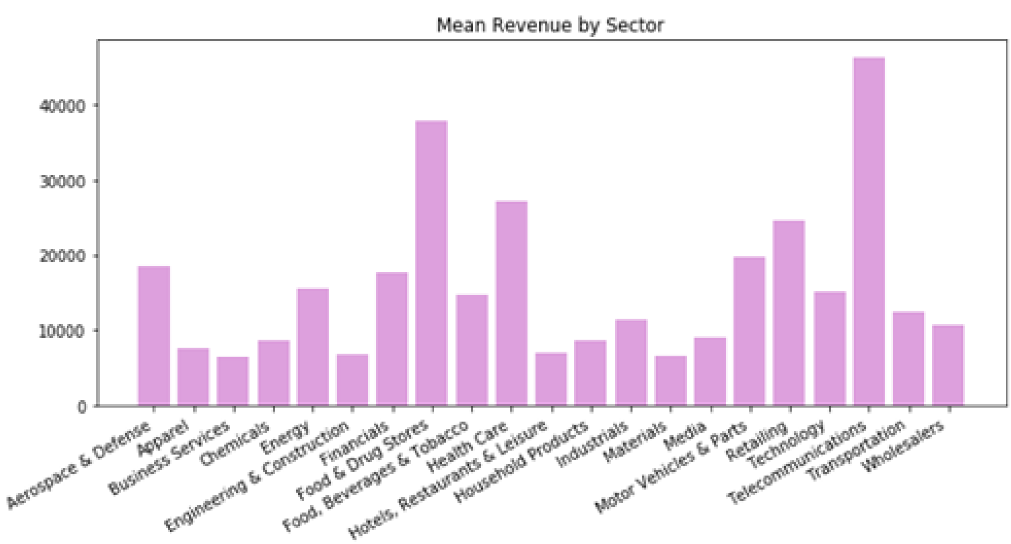

Figure 3: Distribution of Mean Revenue by Industry Sector

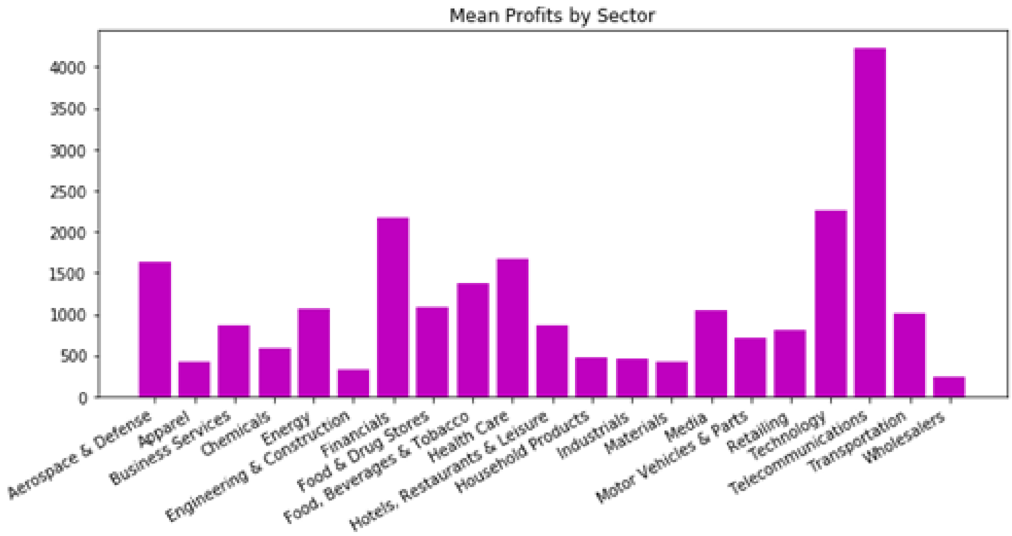

Figure 4: Distribution of Mean Profit by Industry Sector

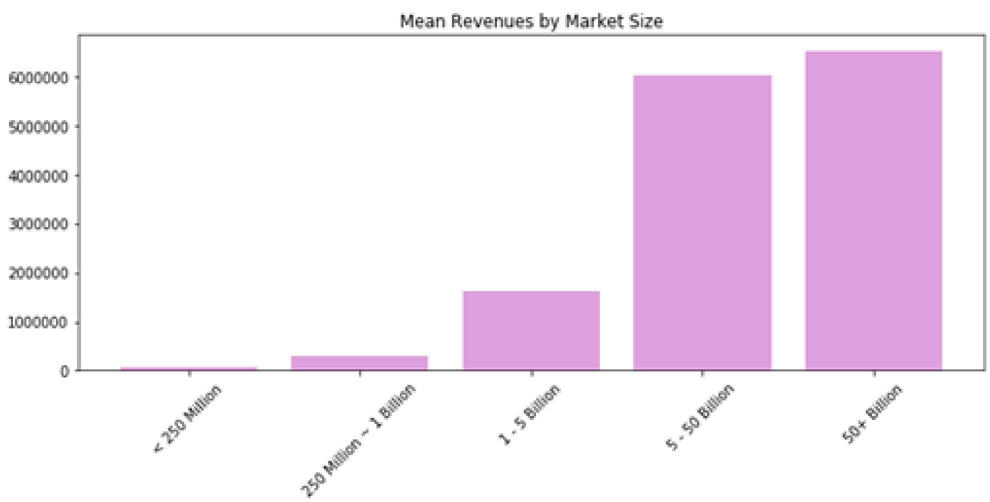

Figure 5: Distribution of Mean Revenue by Market Size

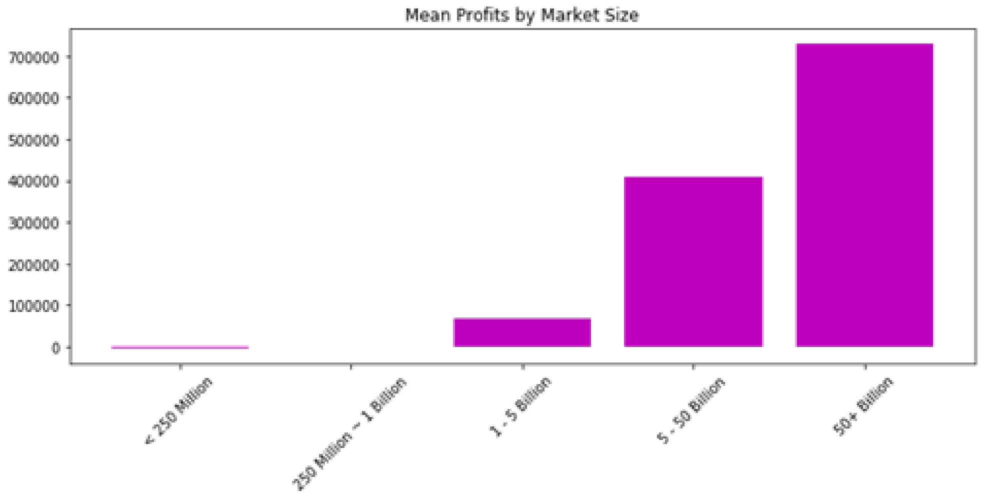

Figure 6: Distribution of Mean Profit by Market Size

Observing the stark differences between these different categories suggest the importance of one-hot encoding our control variables so that we account for these kind of exogenous variables in our analysis, and isolate the effect that board characteristics would have on a company's performance.

5.2 B: Regression Table

The following regression table, table 5, shows that the effect of adding more variables does little to affect the significance and coefficient on various variables, thus justifying our depiction of our OLS results in table 3.

	Model Regression I	Model Regression II	Model Regression III	Model Regression IV	Model Regression V	Model Regression VI
const	109.78	105.67	-667.53	-1671.49	1211.34	-645.08
	(770.61)	(2667.35)	(4180.84)	(4221.38)	(4370.62)	(4947.96)
GD & HBR Ranked	7040.05***	6810.78**	6831.74**	6809.75**	7018.69***	6942.80***
	(1562.56)	(2685.87)	(2691.07)	(2685.37)	(2669.01)	(2672.12)
GD Ranked	458.38	556.48	543.53	602.75	466.52	569.07
	(699.03)	(1337.02)	(1339.99)	(1337.68)	(1330.06)	(1336.90)
HBR Ranked	-134.88	-706.90	-709.88	-710.67	-641.65	-610.57
	(629.88)	(1109.72)	(1111.35)	(1108.98)	(1102.01)	(1103.28)
Median pay		0.00*	0.00*	0.00*	0.00**	0.00**
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Median age			12.55	13.10	20.23	20.57
			(52.22)	(52.11)	(51.85)	(51.88)
women on board				4935.56	5599.82*	4955.95
				(3167.19)	(3158.91)	(3261.01)
No Directors					-320.67**	-312.82**
					(137.07)	(137.50)
Board Independance						2093.24
						(2610.37)
R^2	0.35	0.33	0.33	0.34	0.35	0.35
AIC	18705.18	7394.54	7396.48	7395.78	7391.69	7392.97
BIC	18881.46	7539.54	7545.40	7548.61	7548.45	7553.65

(Standard errors in parentheses) * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

Table 5: Regression Models on Profits

5.3 C. Pertinent Code

In this part of the appendix we share our code to illustrate our general scraping and cleaning process. Not all code is provided, but can be upon contact of the authors. The goal of this section is to help understand how we approached the data scraping and cleaning process.

Code to Scrape Board Characteristics

```
'''
While most of the data from the WSJ was in a single paged, to access it required the use of
    XPaths. Once we understood how to use those, the process was straightforward.
'''
url = 'http://graphics.wsj.com/boards-of-directors-at-SP-500-companies/'

companies = browser.find_elements_by_xpath('//*[@companyName]')
info = []
for ii in companies:
    outerhtml = ii.get_attribute('outerHTML') #// to extract outerHTML
    tag_value=outerhtml.split("\" ") #// to extract board member info
    info.append(tag_value)



#At this point the scraped data is cleaned and organized into a pandas dataframe, the cleaning
    part of which is omitted for space

#turn into dataframe and save as csv for cleaning
wsj = pd.DataFrame({'Company':companynames,'Market_Cap':marketcap,'Industry':industry,'
    No_Directors':tot_dir,'Median_age':med_age,'Board_Independance':board_ind,'
    Median_Tenure':tenure,'Median_pay':medpay,'women_on_board':perc_woman})
wsj.to_csv('wsjS&P.csv')
```

Code to scrape Glassdoor data (scraping HBR data is near identical)

```
'''
The code to scrape the CEO rankings from both Glassdoor and HBR were similar. They also took
    advantage of the XPath classes.
'''
browser = webdriver.Chrome('chromedriver.exe')
url = 'https://www.glassdoor.com/Award/Top-CEOs-LST_KQ0,8.htm'
browser.get(url)
ids = browser.find_elements_by_xpath("//*[@class ='h2 m-0']")

#get the names from Glassdoor
```

```
glassdoor_name = []
for ii in ids:
    outerhtml = ii.get_attribute('outerHTML') #// to extract outerHTML
    glassdoor_name.append(outerhtml)

#get the corresponding companies
glassdoor_comp = []
companies = browser.find_elements_by_xpath("//*[@class='mt-xsm mr-xl mb-0']")
for c in companies:
    glassdoor_comp.append(c.get_attribute('outerHTML'))

#clean out the CEO names and their approval rating
gd_names, gd_approve = [],[]
for j in range(len(glassdoor_name)):
    n,a = re.findall(r">[\%\$\(\)\-\,\;\&\s\.a-zA-Z0-9]+<",glassdoor_name[j])
    gd_names.append(n)
    gd_approve.append(a)

gd_comps = [re.findall(r">[\''\\\%\$\(\)\-\,\;\&\s\.a-zA-Z0-9]+<",glassdoor_comp[j]) for j
    ↪ in range(len(glassdoor_comp))]
```

Code to clean WSJ data. Cleaning process followed the same process as seen here

```
#Reading in the data that will be prepped for merging
df2 = pd.read_csv("wsjS&P.csv")
df1 = pd.read_csv("Cleaned_F500data.csv")
df1 = df1.rename(columns = {"companies":"Company"})
test2 = df2.copy()
test1 = df1.copy()

#Cleaning and parsing operation to be performed on individual strings in each entry
clean_amp = lambda text: re.sub(r"&amp;","&",text)
clean_space = lambda text: text[:-1] if text[-1] == " " else text
clean_period = lambda text: re.sub(r"\."," ",text)
clean_white = lambda text: re.sub(r" ","",text)
clean_company = lambda text: re.sub(r"company$","",text)
clean_companies = lambda text: re.sub(r"companies$","",text)
clean_group = lambda text: re.sub(r"group$","",text)
clean_corp = lambda text: re.sub(r"corporation","corp",text)
lowercase = lambda text: str.lower(text)
clean_adobe = lambda text: "adobe" if text == "adobesystems" else text

#Cleaning operation loaded into a list to be iterated and applied to the DataFrame column
```

```
clean_funcs = [clean_amp,clean_space,clean_period,clean_white,clean_company,clean_companies,
    clean_group,lowercase]

#Applying the functions to the data column
for func in clean_funcs:
    test2["Company"] = test2["Company"].apply(func)
    test1["Company"] = test1["Company"].apply(func)

#Merging on prepped data
wsj_F500_merged = test1.merge(test2,on = "Company",how = "outer")
wsj_F500_merged.to_csv("wsj_F500_merged.csv")
```

Final merge of all data

```
#first data set is already merged
df1 = pd.read_csv("wsj_F500_merged.csv")

#Final 2 remaining data sets to be merged onto the first
df2 = pd.read_csv('Cleaned_ceo_rank_glassdoor_data.csv')
df3 = pd.read_csv('Cleaned_ceo_rank_hbr.csv')

#Renaming to match columns names for the joining company names
df2 = df2.rename(columns = {"GD_company": "Company","CEO_Rank": "GD_CEO_Rank"})
df2 = df2.iloc[:,1:]
df3 = df3.rename(columns = {"HBR_Company": "Company","Rank": "HBR_CEO_Rank"})
df3 = df3.iloc[:,1:]
test3 = df3.copy()
test2 = df2.copy()
test1 = df1.copy()

#Cleaning Functionalities to merge on company
clean_amp = lambda text: re.sub(r"&amp;","&",text)
clean_space = lambda text: text[:-1] if text[-1] == " " else text
clean_period = lambda text: re.sub(r"\."," ",text)
clean_white = lambda text: re.sub(r" ","",text)
clean_company = lambda text: re.sub(r"company$","",text)
clean_companies = lambda text: re.sub(r"companies$","",text)
clean_group = lambda text: re.sub(r"group$","",text)
clean_corp = lambda text: re.sub(r"corporation","corp",text)
lowercase = lambda text: str.lower(text)
clean_adobe = lambda text: "adobe" if text == "adobesystems" else text
clean_accented_e = lambda text: re.sub(r"","e",text)
clean_accented_o = lambda text: re.sub(r"","o",text)
```

```
#Operations to be performed in iterable
clean_funcs = [clean_amp,
               clean_space,
               clean_period,
               clean_white,
               clean_company,
               clean_companies,
               clean_group,
               lowercase,
               clean_adobe,
               clean_accented_e,
               clean_accented_o]

#Applying each cleaning functional on to the data frames to be merged
for func in clean_funcs:
    test2["Company"] = test2["Company"].apply(func)
    test3["Company"] = test3["Company"].apply(func)

#Final merging of all data sets
final_merge = test1.merge(test2, on = "Company", how = "outer")
final_merge = final_merge.merge(test3, on = "Company", how = "outer")
final_merge.drop(columns = ["Unnamed: 0", "Unnamed: 0_x"], inplace = True)
final_merge.to_csv("Nov_15_CheckpointFinalMerge.csv")
```