Preslav Angelov
Afsar Ali
ECON 5100

Final Project

1. Introduction

The focus of this paper is to determine what parameters influence the compensation of Chief Executive Officers (CEOs) in US publicly traded companies. There has been much discussion about how high CEOs are being compensated compared to other job levels in a company, and whether this compensation is given justifiably. Furthermore, although it would make sense that a firm's performance is highly indicative of a CEO's compensation, some organizations are used to follow a trend of over-compensating their CEOs without tying it back to performance. However, it is worth mentioning that CEOs have been under close attention by the government with legislature such as the Sarbanes-Oxley Act of 2002, which puts a lot more accountability on the CEO. With that in mind, investors and shareholders expect the CEO of a company to perform well and prove his or her worth. As a result, we would expect a positive correlation between the success of a company and the pay of its CEO. The models outlined by this paper will focus on a few key company performance measures and how they positively or negatively influence the CEO compensation. The paper will also discuss the statistical significance of the model, as well as in depth interpretation of each finding.

2. Summary of Data

The data that we worked with provided us with 114 factors about publicly traded US companies. There were 8300 data points related to a different CEO. Although most factors were related to the performance of the firm, there were a few variables such as the executive's first and last name that served for information purposes only. It is important to note that some variables did not have data for each CEO in the dataset, such as the date the CEO joined the company.

The summary statistics of our data is below. Although there are more variables that could reveal other statistics about the data, , we focused on a few such as the Male proportionality, Total Cash of the company, Net Income, Total Assets, and Age of the CEO. From prior knowledge on the subject, these variables are good candidates in linking the CEO

compensation to performance. Please note that Total Cash, Net Income and Total Assets are in units of thousand dollars.

It is interesting to note that the data presented to us is 93% from male CEOs. The value of '1' for the Male category simply means that the data point reflects a male CEO. Furthermore, there seems to be an outlier in Net Loss of $6117 thousands and Net Income of $45687 thousands. However, the two outliers seem to balance out due to the mean of Net Income being at $603 thousands. Furthermore, the median and mean age of the CEOs seem to agree with each other at 54 years.

Statistics	MALE	TOTAL CASH	Net Income	TOTAL ASSETS	AGE
Min	0.01	0.01	-6177	14	29
1st Qrt	1	59.35	18.03	1174	49
Median	1	183.5	104.34	3639	54
Mean	0.93	1003.4	603.33	26872	54.23
3rd Qrt	1	603	420	11988	59
Max	1	121711	45687	3287968	96

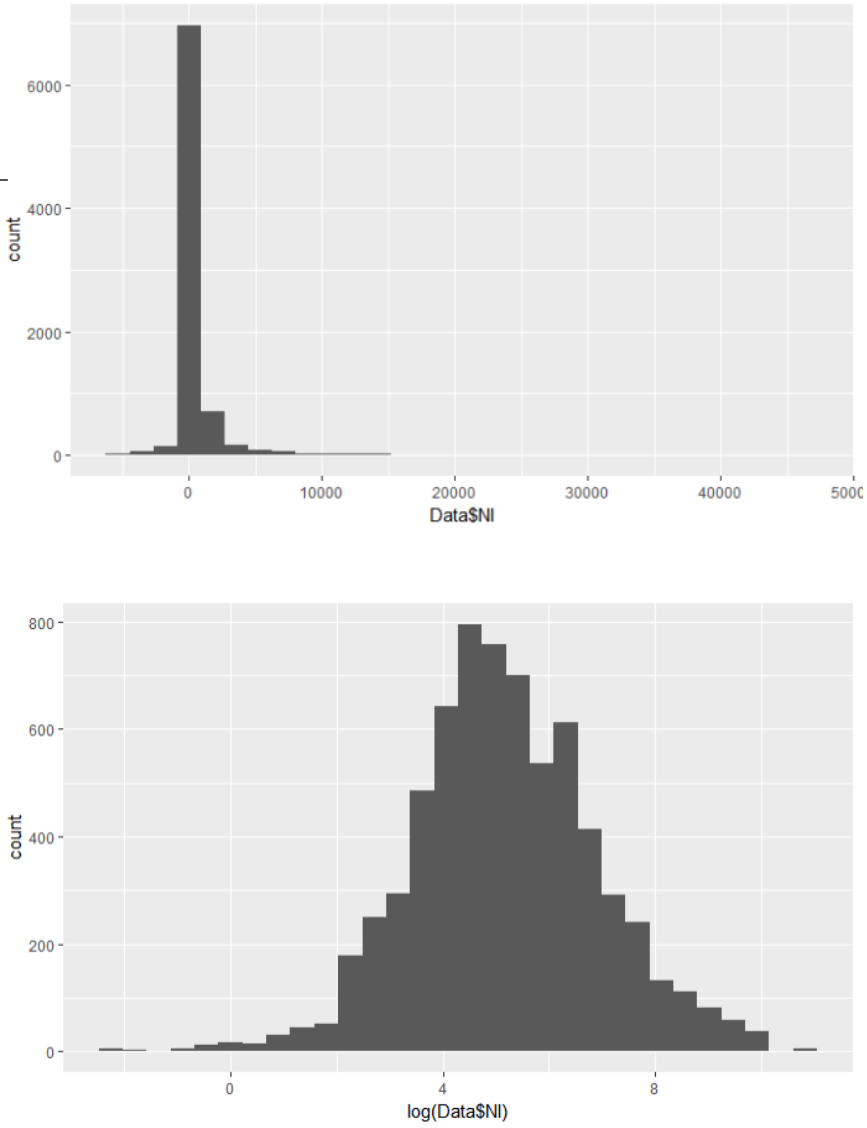
3. Motivation for equation to be estimated

The equation to our estimated model is as follows:

$$Log(TDC2) = \beta_0 + \beta_1 * AGE + \beta_2 * \log(BONUS) + \beta_3 * \log(TotalAssets) + \beta_4 * \log(TotalCash) + \beta_5 * \log(Net\ Income) + \beta_6 * \log(HighStockPrice) + \beta_7 * \log(LowStockPrice)$$

Please note that apart from Age, we took the natural log of all other variables because they were not normally distributed. Natural log allowed us to transform the data to be more normal and plug it into the linear estimating model above. Please see the graphing comparison of the histogram of Net Income versus natural log of Net Income. Without the natural log, Net Income seems very skewed and unevenly distributed toward the left. Once the natural log is introduced, the histogram showed a nice normal distribution with the characteristics of a bell curve.

Natural Log Transformatio



a. Dependent Variable

The depended variable is TDC2, which is the CEO compensation including Salary, Bonus, Non-Equity Incentive Plan Compensation, Value Realized from Stock Option Exercises, Grant-Date Fair Value of Stock Awards, Deferred Compensation Earnings Reported as Compensation, and Other Compensation. Our motivation for choosing this variable is because it encompasses the total compensation of the CEO. Therefore, it gives a wholistic view of what the CEO is really making each year, instead of just focusing on a single value, such as salary. From previous knowledge on the subject, we know that CEOs get a lot of perks and a package of benefits when they get hired,

so we wanted to explore how the performance of the company affects the overall CEO compensation.

b. Independent Variables

First, we chose the age of the CEO because it is a well-known fact that experience is proportional to age, and salary compensation increases with more experience. Therefore, age would be a good candidate to show statistical significance in CEO compensation. We chose bonus because we believe that a higher bonus would yield in a higher compensation. Next, we chose the total assets of the firm because we believe that it is one of the key measures in the company's financial statements that determine its success. The total cash was chosen with the same intention as total assets, but also because the total cash is the single important measure in companies such as Apple and Google. We picked Net Income or Loss because it is a good measure of whether the company is generating success on an annual basis. High and low stock prices were chosen because it measures the success of the CEO in returning value to his or her shareholders and investors. We believe that a CEO would not be compensated as much if the stock price is trending down, and the opposite if the stock price of the company is rising. We purposefully did not choose the Male variable due to the data being heavily skewed toward male CEOs – about 93% as mentioned in the Summary of Data section. As a result, we did not think that female CEOs had good data representation.

4. Results

The summary statistics of our estimated model is below. Intercept denotes the y-intercept of the model. Please note that the P-values equal to 0 in the table are denoted that way for easier reading due to extremely small actual p-values.

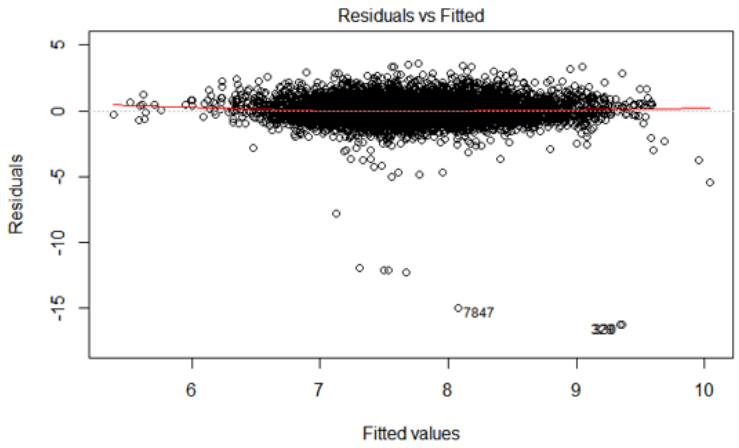
Variables	Coefficients	Std. Error	T-value	P-value
Intercept	4.611	0.124	37.181	0
AGE	0.013	0.0015	8.597	0
Log(BONUS)	-0.0015	0.002	-0.503	0.615
Log(TotalAssets)	0.0661	0.0126	5.273	0
Log(TotalCash)	0.0966	0.0091	10.645	0
Log(NetIncome)	0.174	0.0131	13.3	0
Log(HighStockPrice)	0.0442	0.0369	1.197	0.232
Log(LowStockPrice)	0.0643	0.0304	2.115	0.035

The statistics of the same model, but WITHOUT natural log, is displayed below.

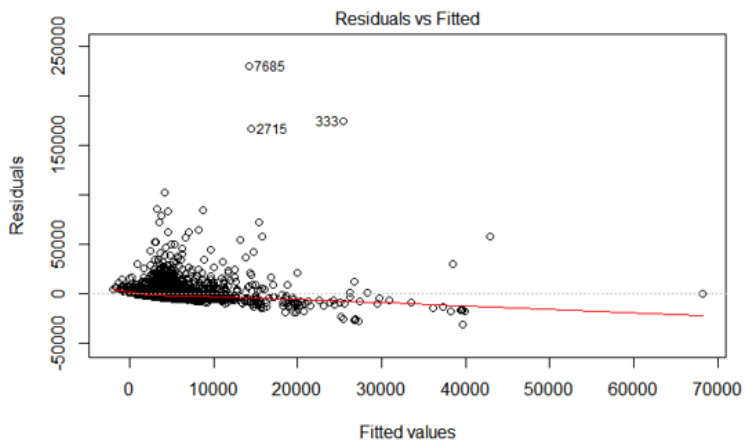
Variables	Coefficients	Std. Error	T-value	P-value
Intercept	-1751	549.1	-3.189	0.00143
AGE	79.41	9.96	7.977	0
BONUS	1.974	0.1204	16.39	0
TotalAssets	-0.0039	0.0006	-6.49	0
TotalCash	0.172	0.0221	7.737	0
NetIncome	0.738	0.0381	19.335	0
HighStockPrice	8.012	4.571	1.754	0.0796
LowStockPrice	1.749	6.45	0.271	0.7863

The main difference in the results between the two models is that the Intercept is less statistically significant and Bonus is more statistically significant in the model without natural log. To understand which model is better, we ran model diagnostic tests, including the Residuals vs Fitted values graphs shown below.

Natural Log Model



Model WITHOUT Natural Log



The plots show whether the residuals of each model have non-linear patterns. The equal spread residuals around the horizontal line in the Natural Log Model show a strong indication that the model has linear relationships. On the other hand, the model without natural log shows the residuals clumped together and not as spread out, meaning that the model is more likely to exhibit non-linear relationships between the dependent and independent variables. Knowing that Natural Log Model is a better model, we also ran the RESET test to ensure it passes. As seen in the table below, the p-value of the natural log model is 0 and the F-stat is 24.46. Because the p-value value is lower than alpha=.05 and the F-stat is bigger than the

value given in the F-distribution table for df1=2 and df2=6747, the null hypothesis is rejected. Therefore, the model is statistically significant.

F-stat	df1	df2	P-value
24.46	2	6747	0

5. Interpretation of Results

From the summary of the results of the natural log model in the previous section, we can see that all variables are statistically significant, except Bonus, High Stock Price, and Low Stock Price. Bonus might not be statistically significant due to the dependent variable (TDC2) already including the Bonus. As a result, having a bigger Bonus does not predict whether a CEO will make more money. Bonus is not a performance measure; rather, it is a reaction to performance measures. Therefore, it makes sense that Bonus is not statistically significant as an independent variable. The Stock Price measures puzzled us as to why they are not statistically significant. Our understanding of the two variables is that they measure the price per share of the company; however, the variables might be intended to measure something different as we could not find a clear explanation and background of how the two variables were calculated.

Age has a positive coefficient, meaning that as age goes up by a year, the compensation of the CEO also increases. The same goes for total assets. Mainly, as total assets increase by 1%, the CEO compensation increases by 0.06%. Furthermore, as total cash increases by 1%, the CEO compensation increases by 0.09%. Lastly, as the net income increase by 1%, the CEO compensation increases by 0.17%. The net income has the most influence on the dependent variable in our model. This makes sense because net income is the profit of the company, which for most investors is the single most important measure of how successful a company is in a given year.

One of the major problems that we ran into with our model was trying to take the natural log of variables that have values of 0. The natural log of 0 is not a real number and our initial model kept giving us errors. To fix the issue in the least disruptive way, we assigned a value of 0.001 to variables that had some values of 0.

An unaddressed problem that we faced is the lack of knowledge of how this data was collected. Without having the background knowledge of how the data was generated, we are not completely confident that our model reflects the real world.

6. Conclusion

The compensation of a CEO has long been believed to be directly related to the performance of a company. Although this belief is accepted in our society today, we built a linear model to reflect this idea backed up by data. We took the CEO's age and bonus, the company's total assets, total cash, net income, and the stock price as the performance variables that could directly alter the compensation of a CEO. Through our work with the given data, we found that the age, total assets, total cash, and net income are statistically significant and that they increase the CEO's compensation as long as the variables; values go up. Although there are more questions to be answered regarding CEO compensation, our model has been proven statistically significant through several model diagnostics.