DATA DESCRIPTION

ECON 494

Yu Xuan Deng
40173189

Research Question Summary

The purpose of this paper is to evaluate whether there is a positive correlation between the base salary of CEOs and future firm returns of public high-tech industries in the US listed on the NASDAQ. Given the high volatility of high tech industries and the speculative nature of investments in high tech, it would be interesting to see whether tech CEOs' base salary has an effect on firm returns since current literature focuses mostly on performance-based compensation. My hypothesis is that there is no relationship between base salary and future firm returns because base salary is a non-performance-based compensation, hence CEOs would not exert effort based on the salary. However, the analysis could have an unexpected result, similar to the current literature around performance-based compensation and firm performance; performance-based compensation is often given to CEO's to solve alignment problems and encourage hard work, but recent papers show a negative relationship between the two, contrary to common practice and belief.

In the following data section, I will describe the data I will be using, how some of the current literatures are related, provide some preliminary data analysis, a description of the data source, and the variables I intend to use. I will also provide some tables, figures, graphs of my data as supplementary information. Finally, I will share some preliminary linear regression results related to the data set to give some insight on how to further improve this paper in future assignments.

Data Section

Data Description

The data I will be using is from three separate sources. First, I obtained sector and industry information on the firms I will be analyzing from NASDAQ Stock Screener under Global Select with the Technology sector filter on and North America region filter on. Next, the company data including the company name, calendar year closing price, market value, and company name are obtained from Wharton Research Data Services CRSP/Compustat Merged database under Fundamentals Annual. Finally, the CEO compensation data is retrieved from Compustat Execucomp Annual Compensation. Key variables include Salary, CEO name, company name, and gender. All data is for the year 2017 to avoid pandemic effects, and the future returns are calculated with 2017 to 2018 calendar close price data.

Related Literature

Currently, the existing literature focuses on performance compensation rather than base salary. The closest and the most relevant research to my paper is by Cooper et al. (2016) where he finds performance-based compensation is negatively related to firm performance. The authors' "data consists of all NYSE, AMEX, and NASDAQ firms jointly listed on the Compustat Execucomp Database, the Compustat annual industrial files, and the CRSP files from 1994 through 2015. Specifically, CEO compensation data from Execucomp and accounting data from Compustat are collected over 1994 to 2015 and CRSP data is obtained over 1994 to 2015" (Cooper et al. 2016). In contrast to my paper, the authors use panel data over time and firms in various sectors which would result in a higher degree of confidence that the results are causal. Furthermore, his method would have a more robust result given the large set of data over time. I will be using only data from 2017 to 2018 of companies listed on the NASDAQ.

Preliminary Data Analysis

In my sample, prior to data cleaning, there are 95 technology companies and CEOs with their respective salary. Variables include coname (company name), salary (in thousands), gvkey, year, pceo (present CEO), title, exec_lname, exec_fname, gender, ticker, PCC (price close in dollars), MV (market value in millions), sector, industry, returns (in percentage). The mean salary is $680,593 with a standard deviation of $405,148. The highest and lowest salary are $3,057,692 and $1 respectively. In terms of market value, the mean is $45.8 billion with a standard deviation of $150 billion. The highest and lowest market value are $77.8 million and $790 billion respectively. The mean return is -3.7% with a standard deviation of 30.4%. The minimum and maximum returns are -63.3% and 80.8% respectively. The sample size is small, however the variation among different variables are significant. The regressor of interest will be Salary in 2017 and the dependent variable will be returns in 2018. I will attempt to use market value and market price as controls to my regression. This study will have a statement about correlation and not attempt to reach a causal statement. Below is my proposed regression model.

$$Salary = \alpha + \beta Returns + \gamma Price + \delta MarketValue + \varepsilon$$

The data was obtained from the sources mentioned above, then downloaded into CSV files, the documentation was also provided by the sources I downloaded the data from. For most of the data, I am using the variables as is. The future returns are calculated as simple returns by using the calendar year closing price of 2018 and 2017.

Currently, the data have been downloaded, I merged them using Python with Pandas on gvkey and ticker. I imported Numpy, Matplotlib, Sklearn, and Statsmodels into Python as libraries for analysis. Then, I cleaned the data by dropping NA observations and sorting them by company name in alphabetical order. I did some preliminary plotting to visualize the dataset. I have noticed there are some heteroskedasticity and outliers that are currently affecting my linear regression model, so I dropped the outlier observations. Specifically, I dropped observations with <$10 base salary and observations with >$2 million base salary.

After dropping the observations, the Salary now has the following summary statistics. Mean: 665, std: 237, min: 210, max: 1450. The tables and figures attached in the following sections will detail the data results before and after cleaning.

Tables & Figures

Table I.

	SALARY	GVKEY	YEAR	PCC	MV	returns
count	95.000000	95.000000	95.0	95.000000	95.000000	95.000000
mean	680.593042	71676.568421	2017.0	87.234947	45840.681017	-0.037420
std	405.147769	67102.532898	0.0	152.838782	149594.009840	0.304431
min	0.001000	1161.000000	2017.0	4.200000	77.846400	-0.633175
25%	460.625000	15095.000000	2017.0	23.745000	1102.177750	-0.258670
50%	600.000000	31843.000000	2017.0	48.900000	2631.105600	-0.067896
75%	846.154000	129151.000000	2017.0	99.375000	11325.945900	0.137018
max	3057.692000	264416.000000	2017.0	1053.400000	790050.098100	0.808134

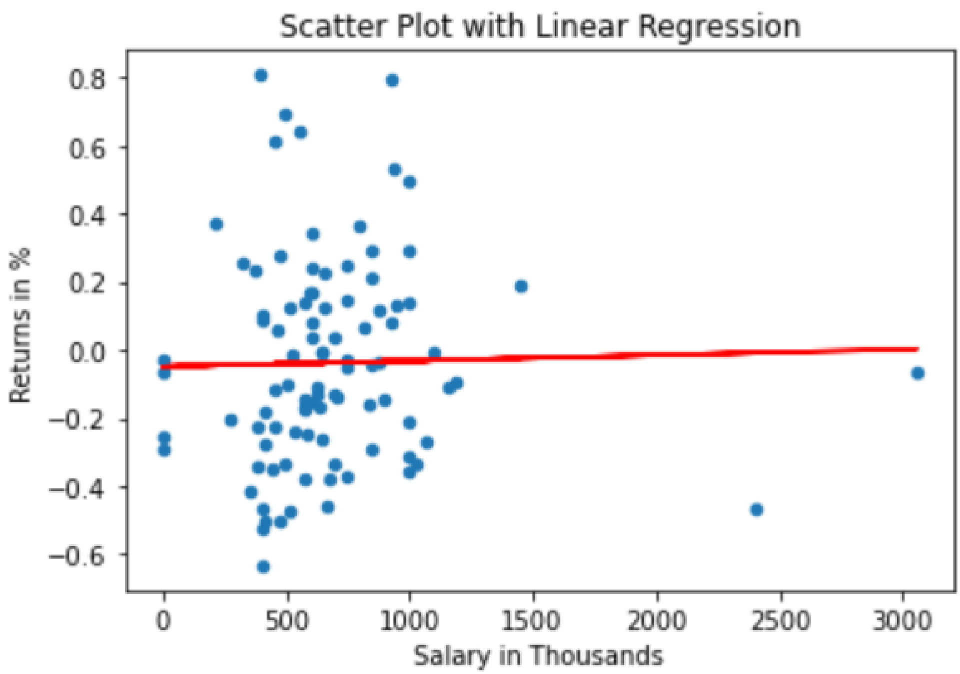
*Before data cleaning

Table II.

	SALARY	GVKEY	YEAR	PCC	MV	returns	fit
count	89.000000	89.000000	89.0	89.000000	89.000000	89.000000	89.000000
mean	665.153292	71729.280899	2017.0	85.180337	34101.435880	-0.026824	-0.026824
std	237.137403	67712.425552	0.0	157.250155	122122.606578	0.309084	0.035744
min	210.000000	1161.000000	2017.0	4.200000	77.846400	-0.633175	-0.095431
25%	475.000000	16710.000000	2017.0	22.120000	1024.909600	-0.246311	-0.055487
50%	620.000000	31607.000000	2017.0	47.160000	2631.105600	-0.053590	-0.033630
75%	842.308000	133288.000000	2017.0	89.030000	11168.314000	0.149886	-0.000121
max	1450.000000	264416.000000	2017.0	1053.400000	729439.025200	0.808134	0.091478

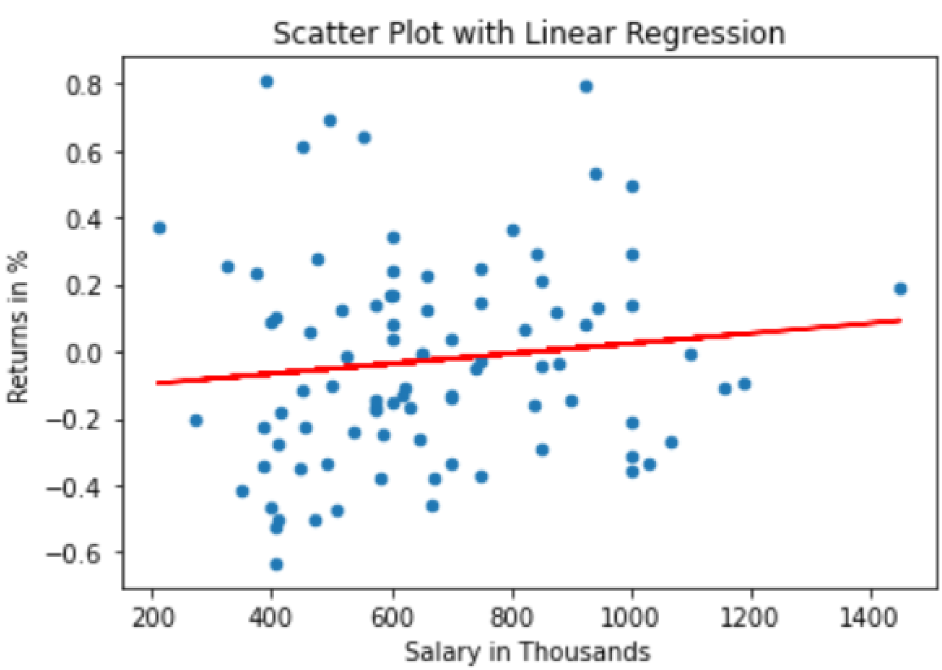
*After data cleaning

Figure I.



*Before data cleaning

Figure II.



*After data cleaning

Table III.

OLS Regression Results

Dep. Variable:	returns	R-squared:	0.001
Model:	OLS	Adj. R-squared:	-0.032
Method:	Least Squares	F-statistic:	0.02656
Date:	Mon, 20 Feb 2023	Prob (F-statistic):	0.994
Time:	13:50:17	Log-Likelihood:	-21.270
No. Observations:	95	AIC:	50.54
Df Residuals:	91	BIC:	60.76
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-0.0552	0.071	-0.774	0.441	-0.197	0.087
SALARY	2.351e-05	8.86e-05	0.265	0.791	-0.000	0.000
MV	-6.179e-08	3.48e-07	-0.178	0.859	-7.53e-07	6.29e-07
PCC	5.341e-05	0.000	0.168	0.867	-0.001	0.001

Omnibus:	6.931	Durbin-Watson:	1.767
Prob(Omnibus):	0.031	Jarque-Bera (JB):	6.571
Skew:	0.633	Prob(JB):	0.0374
Kurtosis:	3.237	Cond. No.	3.50e+05

*Before data cleaning

Table IV.

OLS Regression Results

Dep. Variable:	returns	R-squared:	0.001
Model:	OLS	Adj. R-squared:	-0.032
Method:	Least Squares	F-statistic:	0.02656
Date:	Mon, 20 Feb 2023	Prob (F-statistic):	0.994
Time:	14:00:50	Log-Likelihood:	-21.270
No. Observations:	95	AIC:	50.54
Df Residuals:	91	BIC:	60.76
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-0.0552	0.071	-0.774	0.441	-0.197	0.087
SALARY	2.351e-05	8.86e-05	0.265	0.791	-0.000	0.000
MV	-6.179e-08	3.48e-07	-0.178	0.859	-7.53e-07	6.29e-07
PCC	5.341e-05	0.000	0.168	0.867	-0.001	0.001

Omnibus:	6.931	Durbin-Watson:	1.767
Prob(Omnibus):	0.031	Jarque-Bera (JB):	6.571
Skew:	0.633	Prob(JB):	0.0374
Kurtosis:	3.237	Cond. No.	3.50e+05

*After data cleaning

References

Abowd, J. M. (1990). Does performance-based Managerial Compensation Affect Corporate Performance? *ILR Review*, *43*(3), 52S–73S. https://doi.org/10.1177/001979399004300304

Armstrong, C., Nicoletti, A., & Zhou, F. S. (2021). Executive stock options and Systemic Risk. *Journal of Financial Economics*, *146*(1), 256–276. https://doi.org/10.1016/j.jfineco.2021.09.010

Balafas, N., & Florackis, C. (2013). CEO compensation and future shareholder returns: Evidence from the London Stock Exchange. *Journal of Empirical Finance*, *27*, 97–115. https://doi.org/10.1016/j.jempfin.2013.10.011

Bebchuk, L. A., Cremers, K. J. M., & Peyer, U. C. (2011). The CEO pay slice. *Journal of Financial Economics*, *102*(1), 199–221. https://doi.org/10.1016/j.jfineco.2011.05.006

Benmelech, E., Kandel, E., & Veronesi, P. (2010). Stock-based compensation and CEO (dis)incentives. *The Quarterly Journal of Economics*, *125*(4), 1769–1820. https://doi.org/10.1162/qjec.2010.125.4.1769

Bereskin, F. L., & Cicero, D. C. (2012). CEO compensation contagion: Evidence from an exogenous shock. *Journal of Financial Economics*, *107*(2), 477–493. https://doi.org/10.1016/j.jfineco.2012.09.005

Cao, M., & Wang, R. (2013). Optimal CEO compensation with search: Theory and empirical evidence. *The Journal of Finance*, *68*(5), 2001–2058. https://doi.org/10.1111/jofi.12069

Cooper, M. J., Gulen, H., & Rau, R. (2011). The cross-section of stock returns and incentive pay. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1572085

Dittmann, I., Maug, E., & Spalt, O. (2010). Sticks or carrots? optimal CEO compensation when managers are loss averse. *The Journal of Finance*, *65*(6), 2015–2050. https://doi.org/10.1111/j.1540-6261.2010.01609.x

Frydman, C., & Jenter, D. (2010). CEO compensation. *Annual Review of Financial Economics*, *2*, 75–102. https://doi.org/10.3386/w16585

Hayes, R. M., & Schaefer, S. (2000). Implicit contracts and the explanatory power of top executive compensation for future performance. *The RAND Journal of Economics*, *31*(2), 273–293. https://doi.org/10.2307/2601041

Kaplan, S. N. (2008). Are U.S. CEOS overpaid? *Academy of Management Perspectives*, *22*(2), 5–20. https://doi.org/10.5465/amp.2008.32739755

Van Essen, M., Otten, J., & Carberry, E. J. (2015). Assessing managerial power theory. *Journal of Management*, *41*(1), 164–202. https://doi.org/10.1177/0149206311429378