

What is the most efficient ESG factors to predict financial performance results?

Introduction

There are much more companies make disclosure of ESG(Environment, Society and Governance), the investors will value more when companies have good reputation of ESG. For example, the hardware device manufacturer Apple Inc will track down all its suppliers to check if these firms follow the rules based on ESG. If the suppliers are not qualified the ESG rules set by Apple, Apple will replace the suppliers with the one with better ESG scores. Furthermore, Apple have used every means to cut down its carbon footage for becoming more sustainable. For having better ESG scores, I believe the act will enhance the financial outcomes.

Because ESG is much important in the society than before. I tried to find out if there are strong evidence that ESG scores can have positive impact to financial performance. Huang noted that the relationship between ESGP(ESG performance) and CFP(corporate financial performance) is a positive, statistically significant but economically modest; the relationship between environmental ESGP and CFP appears stronger than social or governance. The relationship between ESGP and operating CFP is stronger than with accounting CFP. (Huang D.Z.X, 2019)

I tackled the dataset which contains information about environmental, social and governance disclosure scores to find out if there is connection between the corporate financial performance such as return on common equity, return on assets and gross margin.

If there is strong relationship between ESGP factors and CFP factors, ESG factors will be useful when predicting the financial performance of companies in the future. We can forecast financial performance of a company by analyzing the ESG factors. Besides, when companies are much more aware of that ESG scores will have positive influence on their financial performance outcomes, they will take care of ESG factors, which will make environment, society and governance a huge improvement, the companies and communities will benefit from the win-win situation.

Data Collection

There is one dataset used in the research, which contains the companies' information such as "Ticker", "Name", "Year", "Environmental Disclosure Score", "Return on Assets", "Return on Common Equity", "Gross Margin", etc. during the year 2016 to 2018. I focused on which financial factors("Return on common equity", "Return on assets" and "Gross Margin") will have stronger relationship with "Environmental Disclosure Score", "Social Disclosure Score" and "Governance Disclosure Score". Thus, I picked "Return on Assets", "Return on Common Equity" and "Gross Margin" to do research in depth.

The shape of dataset is (4518, 20), which contains 20 columns with 4518 rows. When I checked the columns "Environmental Disclosure Score", there are rows with null values. There are many ways to deal with the problem. I thought of replacing the null values with zeros, interpolating the values or simply deleting the rows with null values. Sometimes, the null values may not because the values are missing, it's because of the industrial reasons. For example, the financial companies don't manufacture the physical products, so they won't have gross margin. However, I would like to remove all the null values from the columns "Environmental Disclosure Score", "Social Disclosure Score", "Governance Disclosure Score", "Return on Assets", "Return on Common Equity" and "Gross Margin" for simpler data manipulation.

At last, I checked If there are any duplicates in the dataset, and it showed that there are all unique values. After cleansing the dataset, I could manipulate the data with the factors I want. My goal is to draw scattered graphs with regression line to show the relationship between factors; to use Scikit-Learn or Statsmodels libraries to do data predictions and acquire critical statistical information like R square. From these two implementations, we can tell if there is positive or negative relationship between ESGP and CFP factors by reviewing the graphs and statistics.

R square is a vital indicator to determine if there is strong connection between factors. Ozili noted that assuming the R-squared is the only decision rule being considered, the paper argued that a low R-squared of at least 0.10 is acceptable in social empirical modelling provided that some or most of the explanatory variables are statistically significant. This means that regression models that have a low R-squared are good models if some of the explanatory variables are statistically significant. Therefore, a regression model in social science research should not be discarded solely because it has a low R-squared. (Ozili Peterson K, 2023)

Data Description and Summary

First, I compared to the columns "Environmental Disclosure Score" and "Return on

Common Equity”. I showed the information of “Environmental Disclosure Score” column, it showed there are 3179 valid rows; the mean of the “Environmental Disclosure Score” column is 19.25; the standard deviation is 18.56; the minimum is 0.78; the first quartile is 2.33; the second quartile is 11.63; the third quartile is 34.60; the maximum is 80.16. The information of “Return on Common Equity” column is that there are 3071 valid values; the mean is 16.77; the standard deviation is 44.83; the minimum is -315.62; the first quartile is 6.09; the second quartile is 11.80; the third quartile is 20.80; and the maximum is 1048.62. The range of “Return on Common Equity” is much larger than “Environmental Disclosure Score”.

I drew the graph to see the distribution of environmental disclosure score. As we can see in Figure 1, the most score point is between 0 to 5 points.

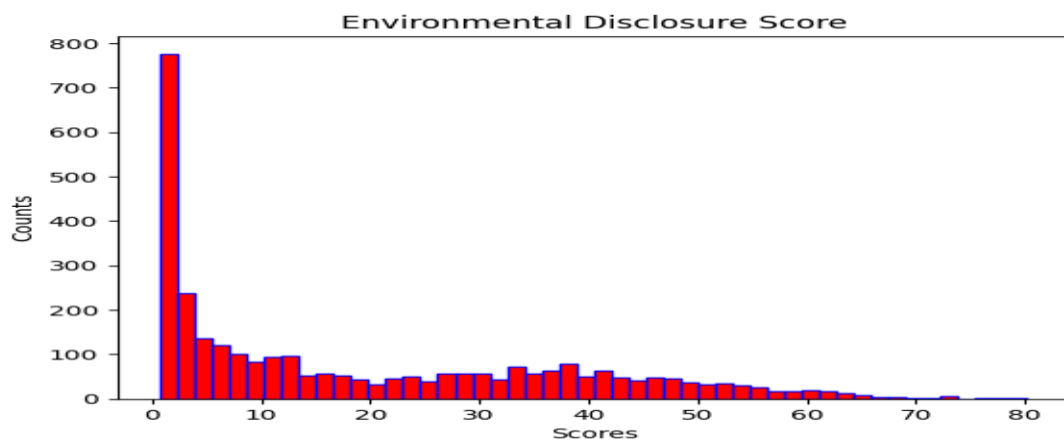


Figure 1 : The number counts of Environmental Disclosure Score

From the Figure 2, we can tell that the return on common equity for most companies is between -50 to 50.

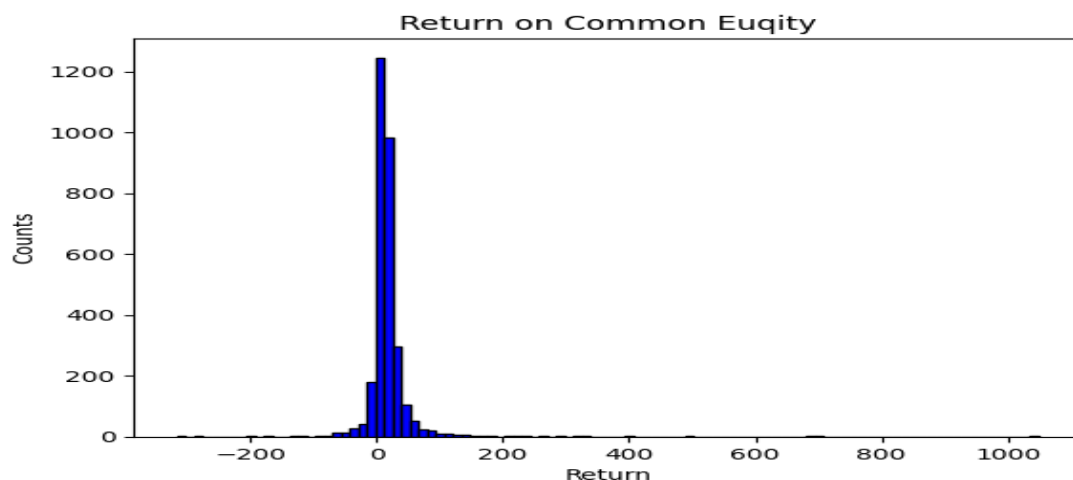


Figure 2: The number counts of Return on Common Equity

Then, I created a new data frame using the “Environmental Disclosure Score” and “Return on Common Equity” to observe the relationship. I used this dataset to draw scattered plot with linear regression line (regplot) from Seaborn library.

As we can see in Figure 3, the relationship is not so valid. So, I would use further statistical analysis to interpret the relationship.

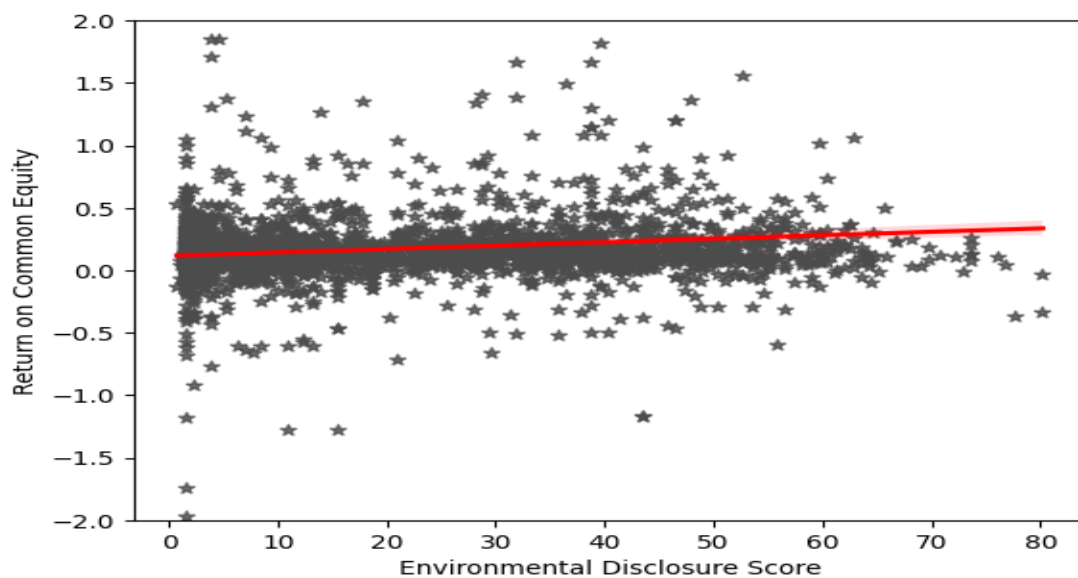


Figure 3: The relationship between Environmental Disclosure Score & Return on Common Equity

I conducted ScikitLearn library (from sklearn import linear_model; from sklearn.model_selection import train_test_split) to train a linear model. After training, I got the results that R square value is 0.0128; the coefficient is 0.286; the intercept is 11.344. Besides, I used Statsmodels library (import statsmodels.api as sm) to get the similar result, which showed R square is 0.013.

From above observation, there is few linearity between these two factors “Environmental Disclosure Score” and “Return on Common Equity”. From Figure 1 and Figure2, we can tell that “Environmental Disclosure Score” is right skewed distribution, and “Return on Common Equity” is zero skewed distribution. Besides, figure3 showed that the slope between two factors is not very significant.

Second, I tapped into the column “Return on Assets” to see if there any strong relationship with “Environmental Disclosure Score” column.

I drew a histogram graph to show the distribution of “Return on Assets” column. The result showed that it’s a zero skewed distribution.

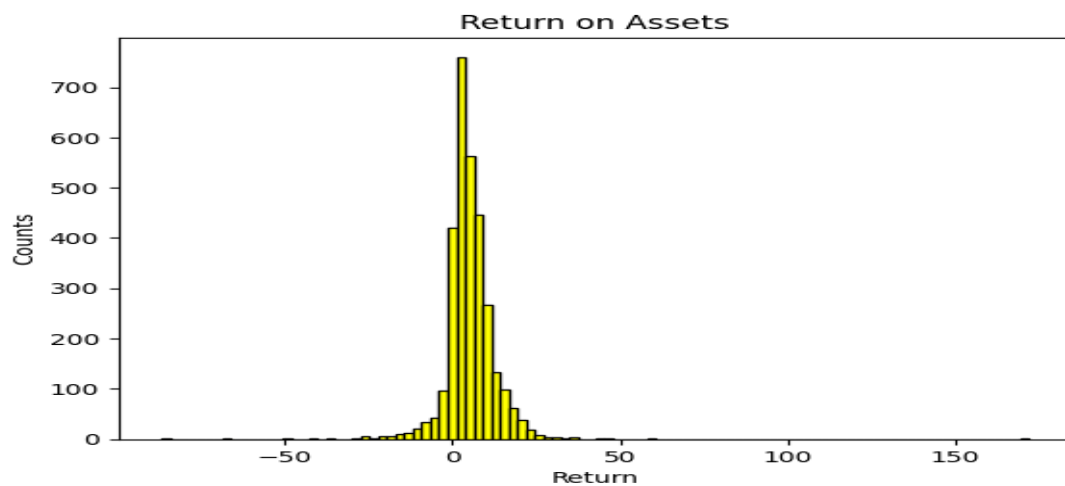


Figure 4: The number count of Return on Assets

I drew the scattered graph with regression line to see the relationship between “Environmental Disclosure Score” and “Return on Assets”. The result showed us that the relationship is as similar as “Return on Common Equity”. But, the blue line is much flatter.

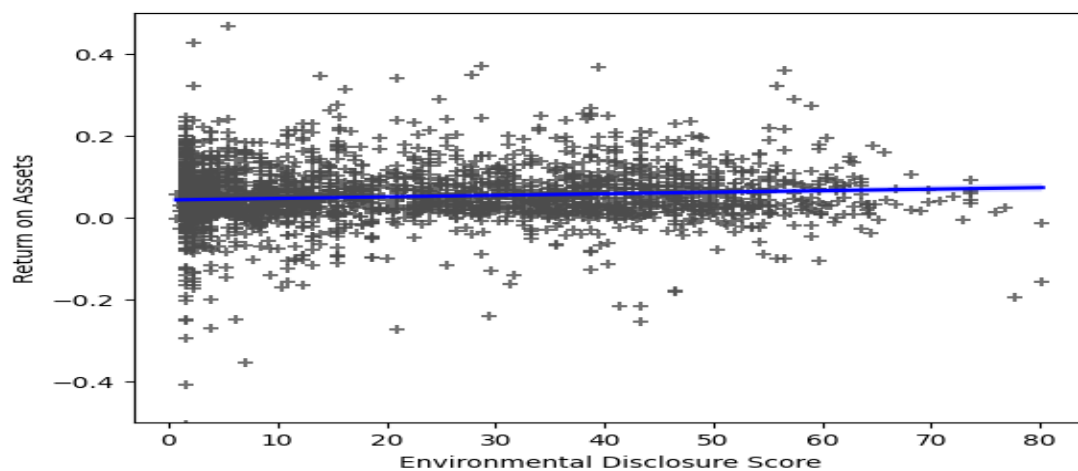


Figure 5: The relationship between Environmental Disclosure Score & Return on Assets

I used Scikit-Learn and Statsmodels libraries to get the linear regression results, which showed that R square is 0.008; the coefficient is 0.040 and the intercept is 4.272. The result showed that there is no strong linearity between “Environmental Disclosure Score” and “Return on Assets”

Third, I used the last financial performance factor “Gross Margin” in the dataset to check relationship. I counted the number of gross margin ranges and drew a histogram graph to show it. The distribution of “Gross Margin” is not so centric as the “Return on Common Equity” and “Return on Assets”.

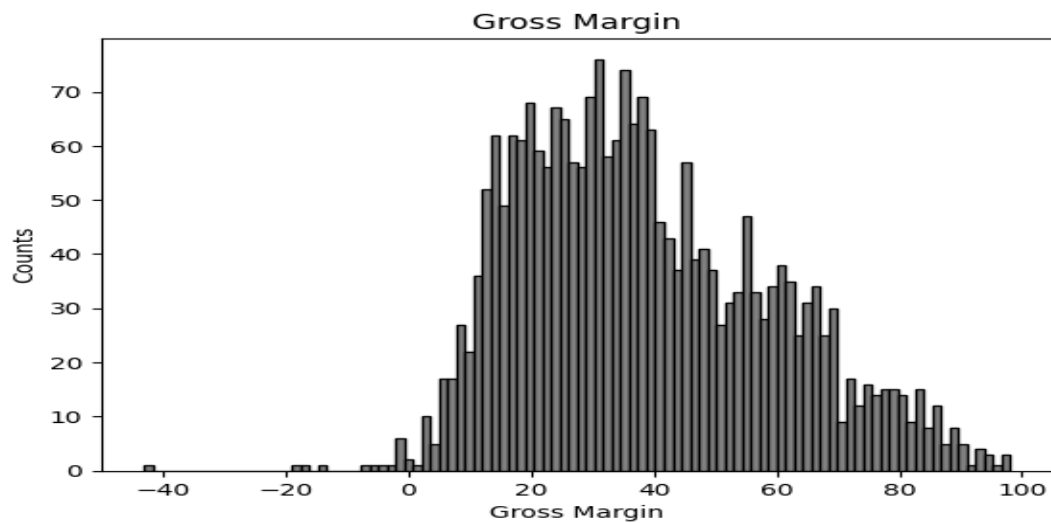


Figure 6: The number count of Gross Margin

I drew the scattered graph with linear regression line to see the relationship between “Environmental Disclosure Score” and “Gross Margin”. The result is still very alike from the previous two financial performance factors. There is no strong relationship between “Environmental Disclosure Score” and “Gross Margin”.

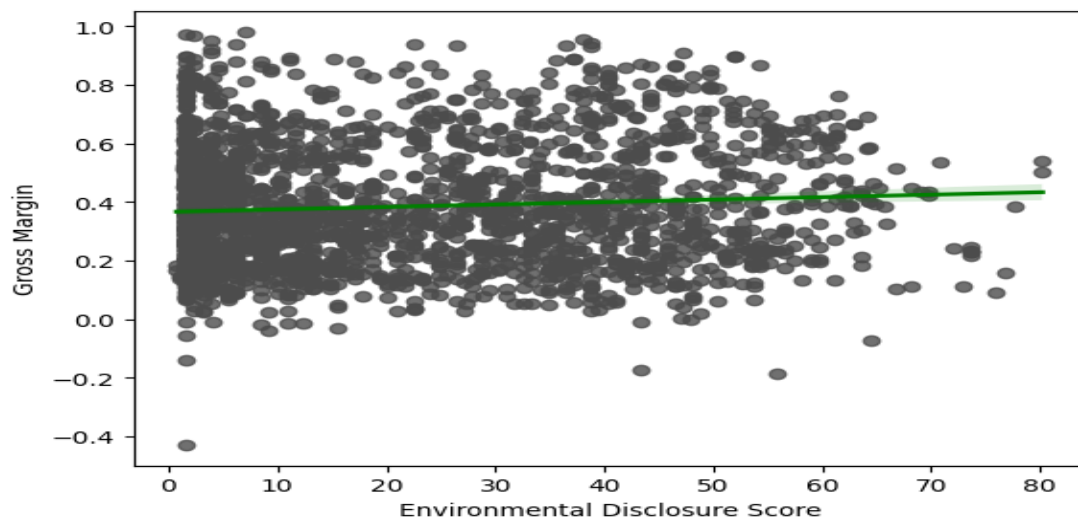


Figure 7: The relationship between Environmental Disclosure Score & Gross Margin

I used Scikit-Learn and Statsmodels libraries to get the linear regression results, which showed that R square is 0.006; the coefficient is 0.076 and the intercept is 36.983. The result showed that there is no strong linearity between “Environmental Disclosure Score” and “Gross Margin”.

At last, I took another ESGP factors like “Social Disclosure Score” and “Governance

Disclosure Score” into consideration. In the first place, I took a look at the distribution of “Social Disclosure Score” and “Governance Disclosure Score”. Below are the graphs I generated.

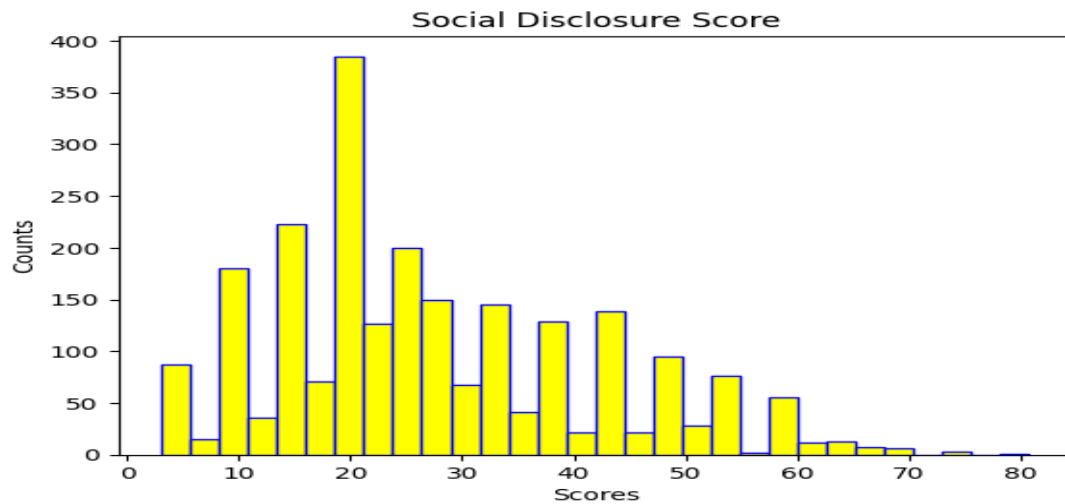


Figure 8: The number count of Social Disclosure Score

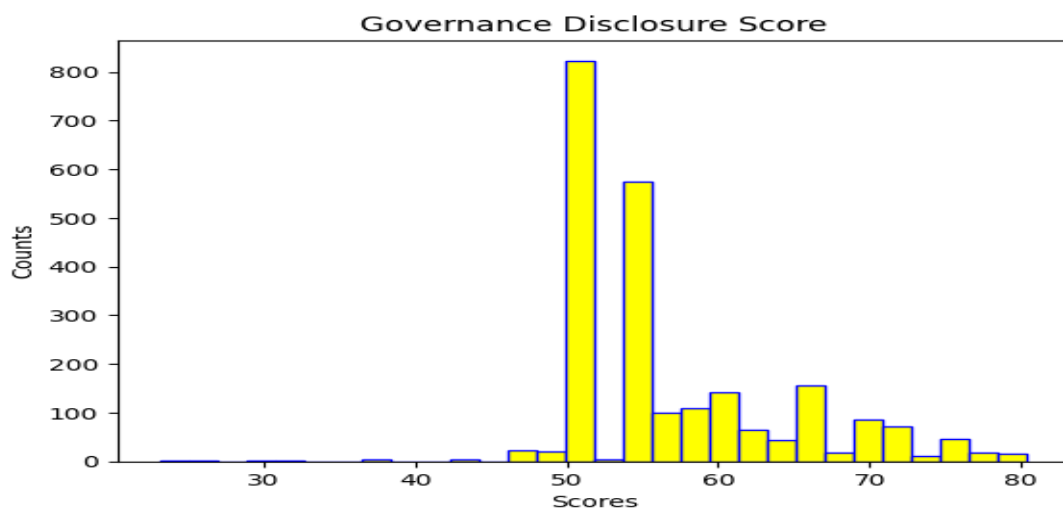


Figure 9: The number count of Governance Disclosure Score

We can tell from the Figure 8&9, the distribution is very different from the Environmental Disclosure Score. Social Disclosure Score graph shows a bit like normalized distribution; however, Governance Disclosure Score shows that many companies have high scores compared to the scores from Environmental Disclosure Score and Social Disclosure Score.

Below are three graphs (Figure10~12) that showed the relationship between social disclosure score and the financial performance factors such as return on common equity, return on assets and gross margin. The results are very similar with the relationship between environmental disclosure score and financial performance factors. We can tell there is no strong relationship in the graphs. So, The scores of social disclosure is

irrelevant for return on common equity, return on assets and gross margin.

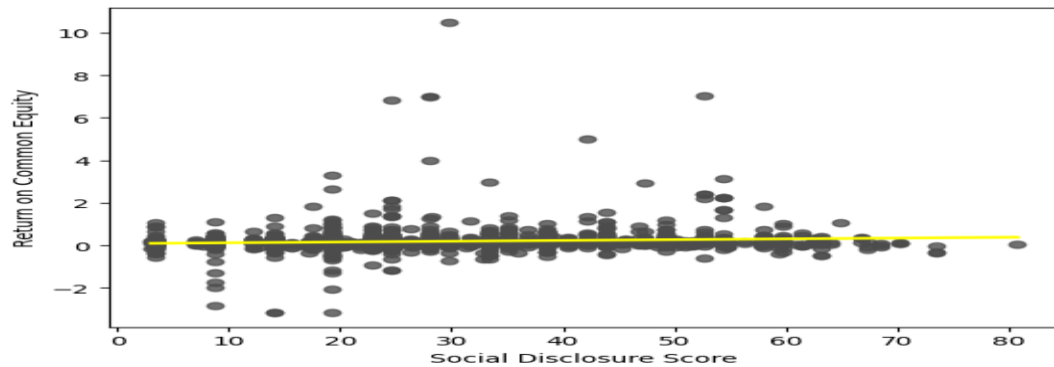


Figure 10: The relationship between Social Disclosure Score & Return on Common Equity

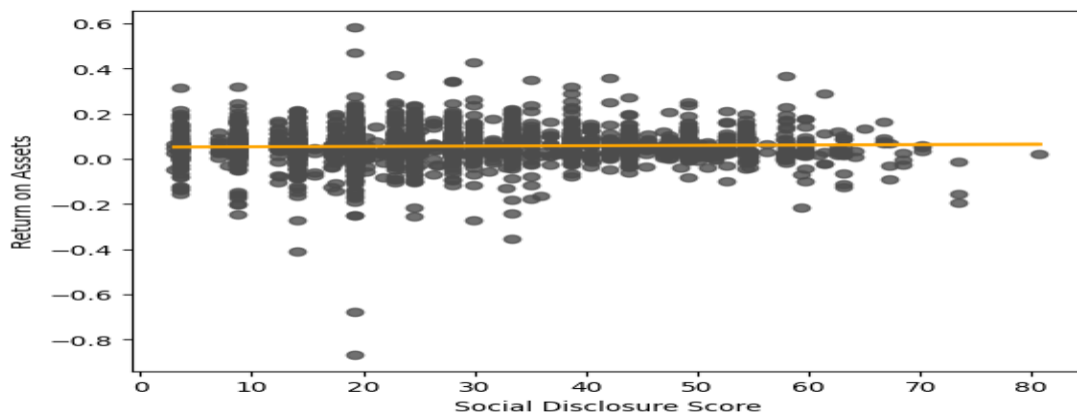


Figure 11: The relationship between Social Disclosure Score & Return on Assets

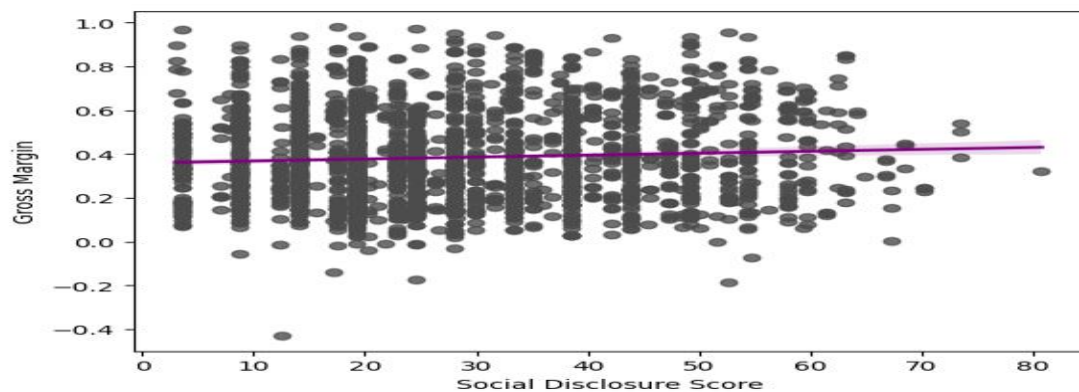


Figure 12: The relationship between Social Disclosure Score & Gross Margin

Again, I used Statsmodels library to figure out the R square value. The outcome showed that the R square of social disclosure score and return on common equity is 0.011; the R square of social disclosure score and return on assets is only 0.001; the R square of social disclosure score and gross margin is only 0.004. The R square values are too low to show that there is any linearity between social disclosure score and other three financial performance factors.

Below are three graphs (Figure13~15) that showed the relationship between governance score and the financial performance factors such as return on common equity, return on assets and gross margin. The results are very similar to the relationship between social disclosure score and financial performance factors. We can tell there is no strong relationship in the graphs. So, The scores of governance disclosure is irrelevant for return on common equity, return on assets and gross margin.

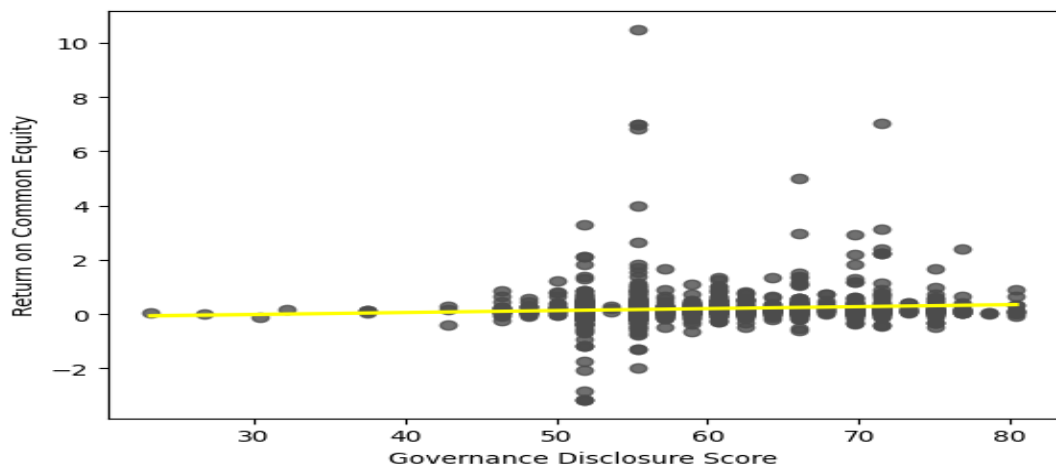


Figure 13: The relationship between Governance Disclosure Score & Return on Common Equity

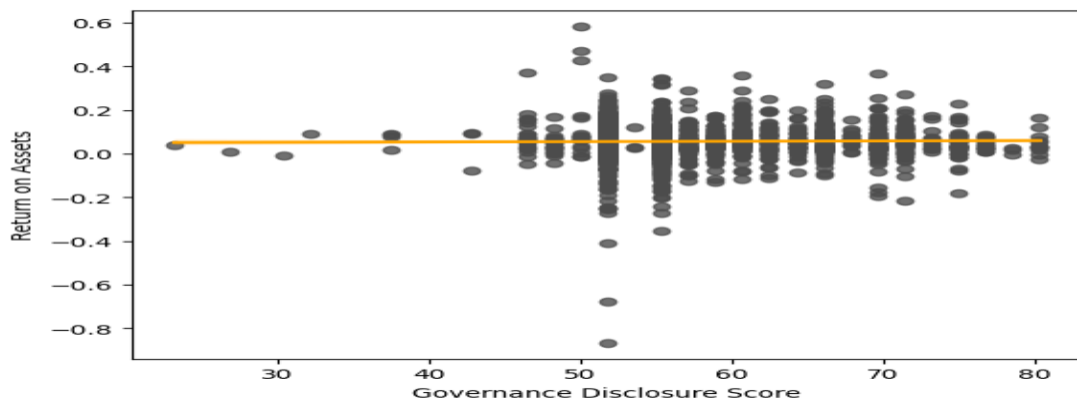


Figure 14: The relationship between Governance Disclosure Score & Return on Assets

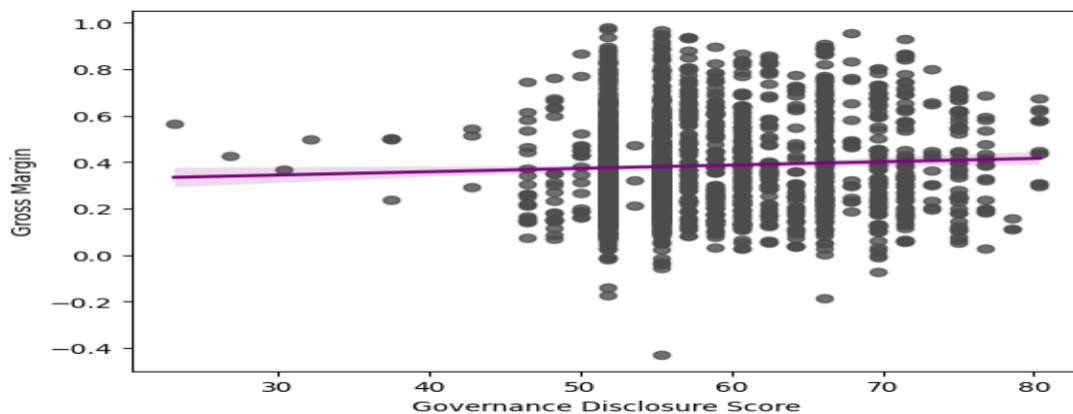


Figure 15: The relationship between Governance Disclosure Score & Gross Margin

Again, I used Statsmodels library to figure out the R square value. The outcome showed that the R square of governance disclosure score and return on common equity is only 0.01; the R square of governance disclosure score and return on assets is 0; the R square of governance disclosure score and gross margin is only 0.002. The R square values are all too low to show that there is any linearity between governance disclosure score and other three financial performance factors.

The R square between “Environmental Disclosure Score” and “Return on Common Equity” is greatest among the other relationship(See Appendix-I). The R square value is 0.013, which is better than the benchmark Ozili set(Ozili Peterson K, 2023). So, I used the “Environmental Disclosure Score ” to predict the value of “Return on Common Equity”. The formula is derived from the linear_model package from sklearn library and train_test_split from sklearn.model_selection. The result is “Return on Common Equity” = 0.29 * “Environmental Disclosure Score” + 11.39. I set the range from 0 to 100 with 100 numbers to predict the “Return on Common Equity”. The max of the “Return on Common Equity” is 40.39; the min is 11.39; and the mean is 25.89.

Below is the graph which demonstrate the relationship between “Environmental Disclosure Score” and “Return on Common Equity”.

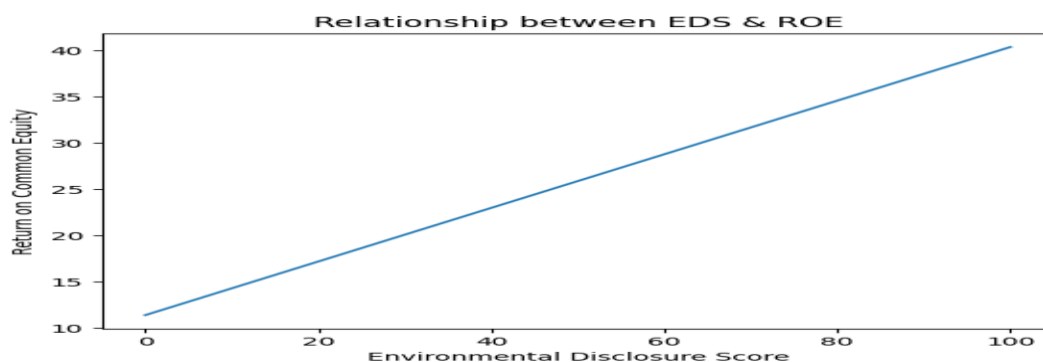


Figure16: relationship between Environmental Disclosure Score and Return on Common Equity

Results and Discussion

I took “Environmental Disclosure Score” as ESGP factor, since it was the finding from Huang (Huang D.Z.X, 2019), who stated that environmental factor appears stronger relationship than social or governance factors. However, I got the different results from the dataset I used. The “Environmental Disclosure Score” cannot be an great interpreter for financial performance. I tackled three financial performance factors “Return on Common Equity”, “Return on Assets” and “Gross Margin”. I drew some graphs to show their relationship, and used statistical library to calculate the R square, which would tell us there is strong linear relationship when R square is close to 1; but, the results from the three above financial factors showed that the relationship is not so strong. That means “Environmental Disclosure Score” cannot predict the companies’ financial outcomes well.

I used the same methodology to find out the relationship between social disclosure score and the three financial performance factors(return on common equity, return on assets, gross margin). Either the graphs or the statistical summary showed that there is still no strong relationship between social disclosure score and the CFP factors.

The last part is about considering governance disclosure score. I used the same way to discover the relationship between governance disclosure score and return on common equity, return on assets, gross margin. After plotting the graphs and showing the statistical summary, I got the result that the relationship is weakest among these factors.

The results showed that no matter what factor we used from ESG, we would not get the strong connection with the CFP factors. If a company has very good environmental, social or governance score, it would not means that the company has very good financial scores either in return on common equity, return on assets or gross margin.

Even if there is no strong connection between ESGP factors and CFP factors, we may use “Environmental Disclosure Score ” for predicting “Return on Common Equity”. Since the R square is 0.013, which is the greatest in the table(See Appendix-I). Besides, Based on research from Ozili Peterson K (2023), we can say that there is a linear relationship between factors when R square is greater than 0.1. So, ESGP factors are all not qualified.

The linear regression line between “Environmental Disclosure Score” and “Return on Common Equity” is

$$\text{Return on Common Equity} = 0.29 * \text{Environmental Disclosure Score} + 11.39.$$

The result was different from Huang D.Z.X (2019), which says that taking care of ESG scores will bring to great financial outcomes. I can only find the limited connection between environmental disclosure score and return on common equity. However, the world is still emphasizing the importance of ESG scores. So, the ESG factors may have huge impact on the other aspects of companies.

The exploration of ESG as the variable measuring the sustainability performance suggests that the choice between short-term returns and long-term value must not be made the more sustainable companies perform better with regard to the environment and society-related factors and achieve long-term shareholder value. (Zumente, I. and Bistrova, J., 2021).

Conclusion

After digging deep in relationship between ESGP and CFP factors, we got the conclusion that there are extremely weak relationship among these factors. It's hard to use ESG factors to predict the financial performance such as return on common equity, return on assets or gross margin. There is slightly connection between "Environmental Disclosure Score" and "Return on Common Equity". We may use the formula derived from the Scikit-Learn library:

$ROE = 0.29 * \text{Environmental Disclosure Score} + 11.39$ to do predictions of companies' return on common equity.

I believe that the importance of measuring ESG is on the stock price performance but financial performance. The fund would like to invest in the companies which have higher ESG scores. However, when dealing with the dataset, we could not acquire the stock price of each companies, which is limited in the analysis. The outcome may align with the observation Huang found (Huang D.Z.X, 2019), relationship with operating performance is stronger than accounting performance. The ESG factors may do a little impact on accounting performance. Serafeim found a positive market reaction to positive ESG news and a negative market reaction to negative news (Serafeim, G. and Yoon, A., 2022). So, maybe the ESG rating is useful when predicting the stock prices but the financial performance of companies. Besides, the research of Wang, H., Shen, H. and Li, S. (2023) use the stock holdings and flow data of Chinese equity funds to calculate firms' stock price fragility and further explore ESG performance's relationship with stock price fragility. They found the results indicate that stock price fragility can be reduced by improving ESG performance. The results suggest that better ESG performance reduces

stock price fragility by weakening investors' sensitivity to stock performance. (Wang, H., Shen, H. and Li, S. ,2023)

The limitation of this data analysis is that I used linear regression to find out the relationship between all these factors. However, the data may has no linearity but there is another correlation between ESGP and CFP factors, maybe the interaction is just not linear. When I choose to conduct the linear regression models into data analysis, I would have inductive bias. The bias is that linear regression may not be a good algorithm to conduct this data analysis, maybe there is non-linear algorithm to fit the models. However, we can still find out there is slightly relationship between “Environmental Disclosure Score” and “Return on Common Equity”.

References

- Huang D.Z.X (2019) “Environmental, Social and Governance (esg) Activity and Firm Performance: A Review and Consolidation,” Accounting and Finance, (2019). doi: 10.1111/acfi.12569.s
- Ozili, Peterson K (2023): The acceptable R-square in empirical modelling for social science research.
- Zumante, I. and Bistrova, J. (2021) “Esg Importance for Long-Term Shareholder Value Creation: Literature Vs. Practice,” Journal of Open Innovation: Technology, Market, and Complexity, 7(2). doi: 10.3390/joitmc7020127.
- Serafeim, G. and Yoon, A. (2022) “Stock Price Reactions to Esg News: The Role of Esg Ratings and Disagreement,” Review of Accounting Studies, 28(3), pp. 1500–1530. doi: 10.1007/s11142-022-09675-3.
- Wang, H., Shen, H. and Li, S. (2023) “Esg Performance and Stock Price Fragility,” Finance Research Letters, 56. doi: 10.1016/j.frl.2023.104101.

Appendix-I

R square(R^2)	Return on Common Equity	Return on Assets	Gross Margin
Environmental Disclosure Score	0.013	0.008	0.006
Social Disclosure Score	0.011	0.001	0.004
Governance Disclosure Score	0.01	0.00	0.002

Appendix-II