**Relationship between U.S. Banks' ESG Initiatives and Financial Performance**

**BSAD 699-101 Capstone - Master of Business Data Analytics**

# Introduction

## Research Objective:

In recent years, the significance of Environmental, Social, and Governance (ESG) factors in the business landscape has garnered increasing attention from investors, corporations, and policymakers. As stakeholders place greater emphasis on sustainable practices, U.S. banks have embraced various ESG initiatives as a means to not only fulfill their social responsibilities but also potentially enhance their financial performance. This study aims to investigate the relationship between U.S. banks' ESG initiatives and their financial outcomes, with the primary goal of analyzing and quantifying this connection.

The primary goal of this research is to predict the financial performance and volatility of US banks using ESG scores and find a link between sustainable business practices and financial success in the banking industry. We hope to get significant insights into the possible influence of ESG initiatives on banks' profitability, stability, and overall financial health by examining a comprehensive dataset containing ESG ratings and financial measurements.

Furthermore, this research aims to assess the relative importance of various granularities of ESG issues on bank financial performance. We understand that ESG comprises a wide range of subcategories, including carbon emissions, board diversity, employee wellbeing, etc. We intend to identify the most relevant characteristics of sustainable practices that banks should focus on for optimal results by examining how each individual ESG aspect affects financial outcomes.

The study's findings could determine potential implications for stakeholders, including investors looking to make informed decisions based on ESG criteria, bank executives developing sustainable plans, and lawmakers developing rules that encourage responsible corporate practices. Moreover, this research contributes to the ongoing discussion about the role of sustainability in the banking industry by shedding light on the relationship between ESG Scores and financial performance in the United States, as well as serves as a foundation for future studies and strategic decision-making in the field of sustainable finance.

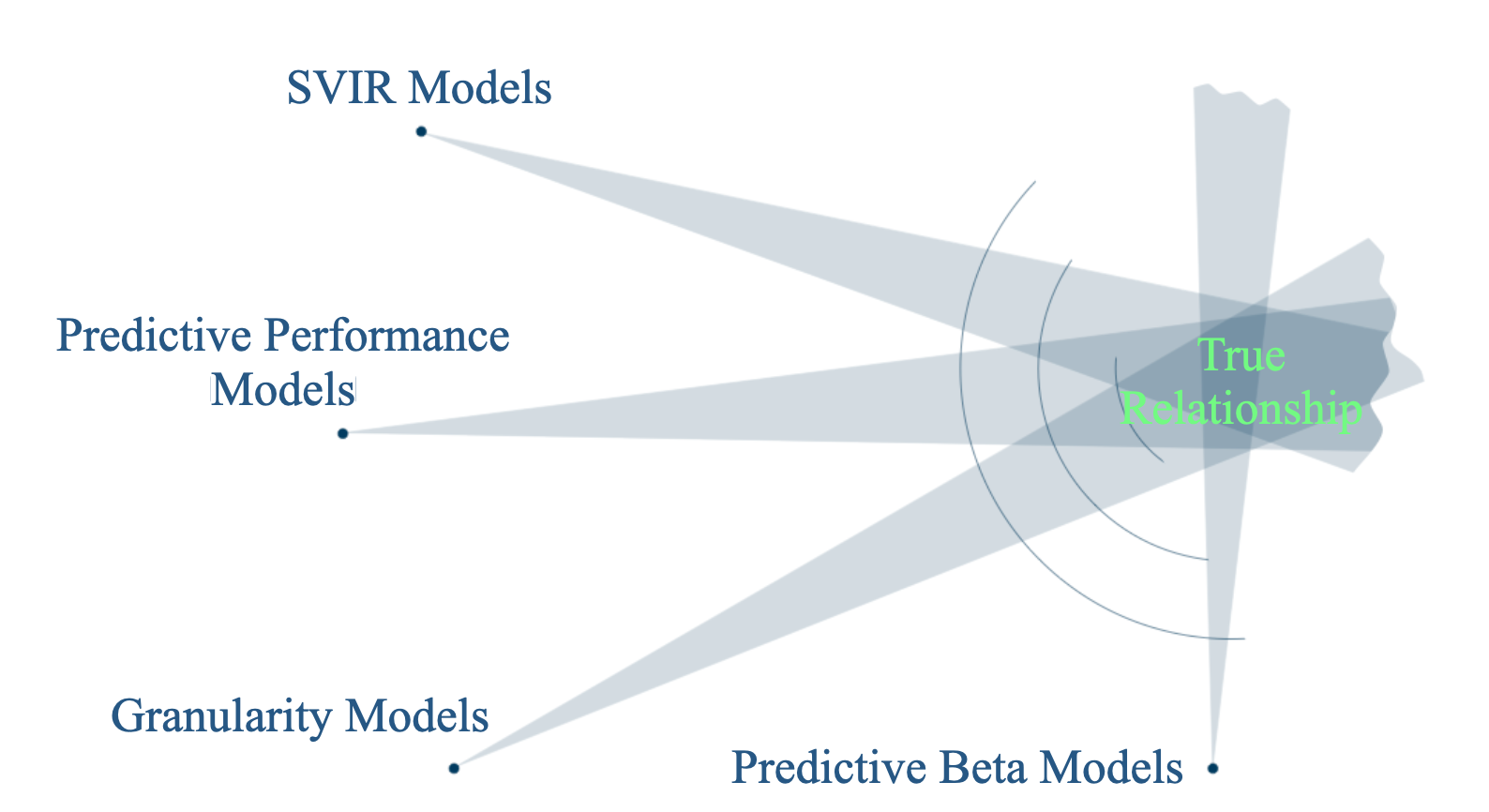
## Thesis:

There exists no relationship between current available ESG metrics and the financial performance of banks in the United States. With a goal as complex as understanding the multifold impact of ESG rankings on the banking industry, we knew that no one single method would be able to describe that relationship so without being misleading at best, and outright false at worst. With this in mind, we opted for a triangulation approach, where we would tackle the same question from a variety of different angles, hoping to arrive at a common true relationship.

For our study we employed four different methodological approaches, each making use of multiple model variations to further dissect the question at hand. The four methods are the following:

1. Predictive Performance Models
2. Predictive Beta Model
3. Single Variable Input Regression (SVIR) Models
4. Granularity Models

Each of the models devised in the study invites further analysis into their specific conclusions, yet taken as a totality they provide no unifying theory for ESG metrics. The lack of an observable effect is only valid for the US banking industry and conclusions should not be extrapolated to other industries.



# Literature Review

Sustainable investing is growing quickly and mutual funds that invest according to ESG (environmental, social, and governance) ratings experience sizable inflows (Hartzmark and Sussman, 2019). Due to these trends, more and more investors rely on ESG ratings to obtain a third-party assessment of corporations’ ESG performance. ESG ratings increasingly influence decisions, with potentially far-reaching effects on asset prices and corporate policies. However, ESG ratings from different providers disagree substantially (Berg et al. 1316).

There are a number of mediating factors that explain the relationship between ESG and CFP. For example, companies with high ESG scores are often better at managing risk, which can lead to lower volatility and a lower cost of capital. Additionally, companies with high ESG scores are often more innovative, which can lead to higher operational efficiency and higher profits.

It is important to note that the relationship between ESG and CFP is not always causal. For example, companies with high ESG scores may be in better financial shape and therefore able to invest more in measures that improve their ESG profile. However, the evidence suggests that ESG investing can be a good way to improve risk-adjusted performance.

Overall, the evidence suggests that ESG investing is a promising investment strategy. Companies with high ESG scores tend to be more financially sound and have lower risk. Additionally, ESG investing can provide downside protection, especially during social or economic crises. As investors become more aware of the importance of ESG factors, ESG investing is likely to become increasingly popular.

Although we have seen much research in recent years exploring the relationship between ESG and CFP, we have not seen any studies that are specific to the financial industry. It does make intuitive sense for ESG investing to be a promising strategy across many sectors, such as fashion, CPG, and many others. However, with financial industries not having many of the shared concerns that other industries face (i.e., manufacturing pollutants, etc.), it would be fascinating to note how ESG strategy applies in financial services.

# 

# Data Preparation and EDA

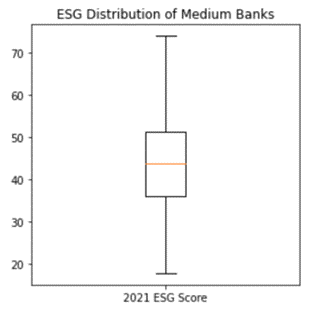
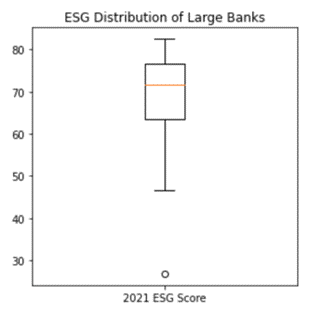
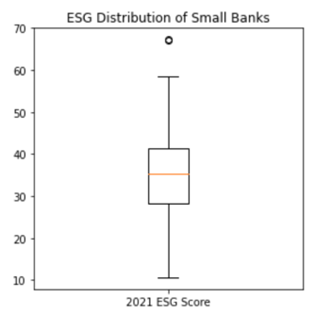
## Datafeed specs

* + Refinitiv feed: 1446 data points for the main model after removing all missing data ranging from 2014 - 2021
  + Segmentation, based on market cap:
    - Large: > $100 billion, 3 banks, 21 data points
    - Medium Large: $50 - 100 billion, 3 banks, 16 data points
    - Medium: $1 - 50 billion, 117 banks, 760 data points
    - Small: < $1 billion, 177 banks, 649 data points

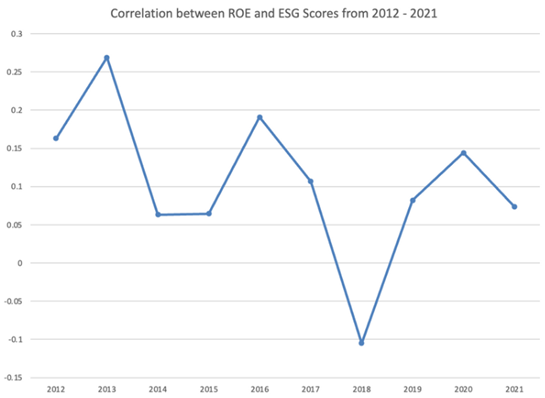
| **Name of variable** | **Description** |
| --- | --- |
| ROE | Net income divided by shareholders' equity. |
| ROA | Net income divided by total assets. |
| Beta | Market beta of banks collected from Refinitiv database.  Calculate Volatility (systematic + unsystematic risks) |
| Market Cap | Total value of outstanding common shares owned by stockholders.  Take natural logarithms. |
| P/B | Market price per share divided by book price per share.  Take natural logarithms. |
| ESG Score | ESG scores of Refinitiv (0-100) includes 34 environmental, 46 social, and 56 governance indicators. |

## EDA

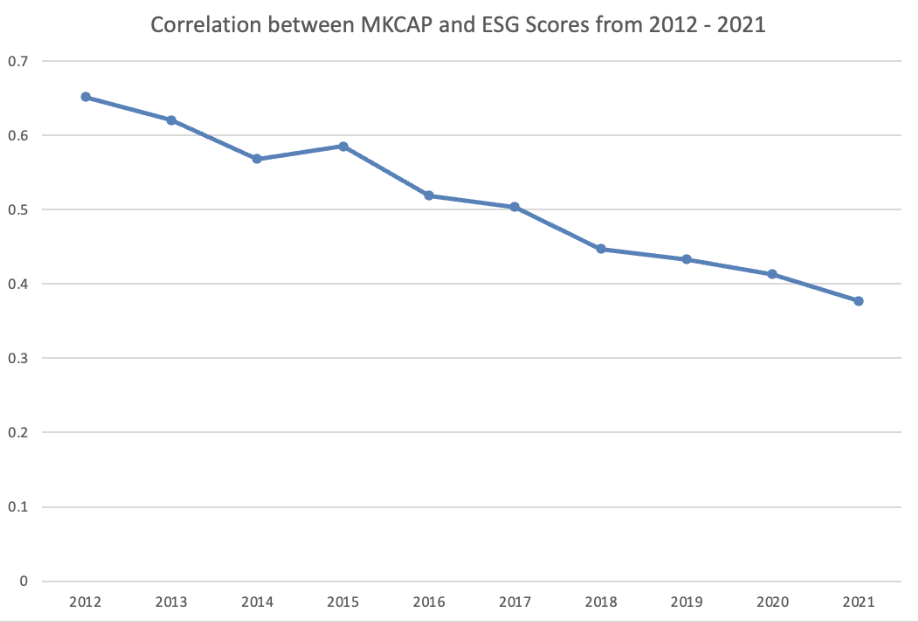
Overall Distribution of ESG



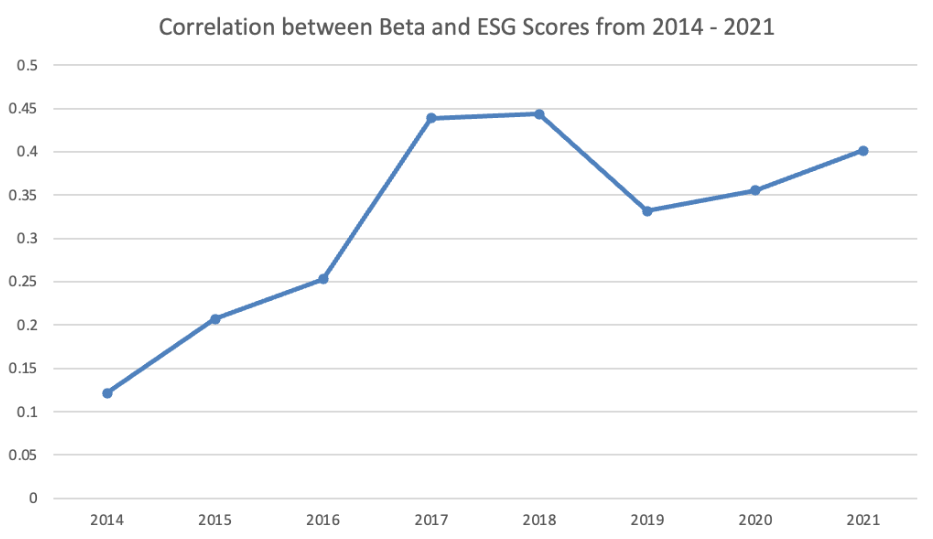
Although the range is larger among bank segments (banks scoring between 20 – 80 points) each segment has a smaller range (on average 10 interval points). Therefore, it could be important to analyze the banks by segment to provide stronger model outcomes.

Correlation between KPIs and ESG Scores

There are significant peaks and valleys in this correlation. Furthermore, the correlation only ranges between -10% and 30%. Therefore, making it challenging to find any correlation between ROE and the overall ESG Score.



There is a strong correlation between the overall ESG Score and MkCap. We suspect that this high correlation is what started the conversation theorizing the impact ESG has on total financial bank performance.



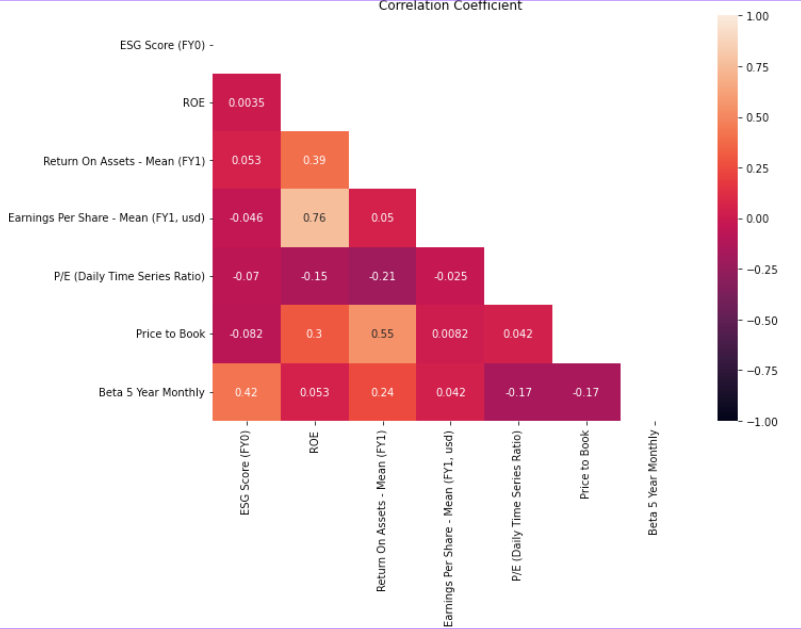
The correlation between Beta and ESG is small. However there has been YoY growth in correlation. This intrigued us to dive further into if we suspect to continue to see this correlation grow. (ie. Beta Model pg24)

10 Year History of ESG Scores for Top 10 US Banks



As the years progress, the top banks' ESG scores are increasing. There is continued discussion regarding why we are seeing a rise in scores. Could it be that banks are improving their operations? Or that there is more exposure to ESG therefore, more data points are now being provided. As time progresses and the ESG discussion stabilizes, we will have a better perception on what the average scores will be.

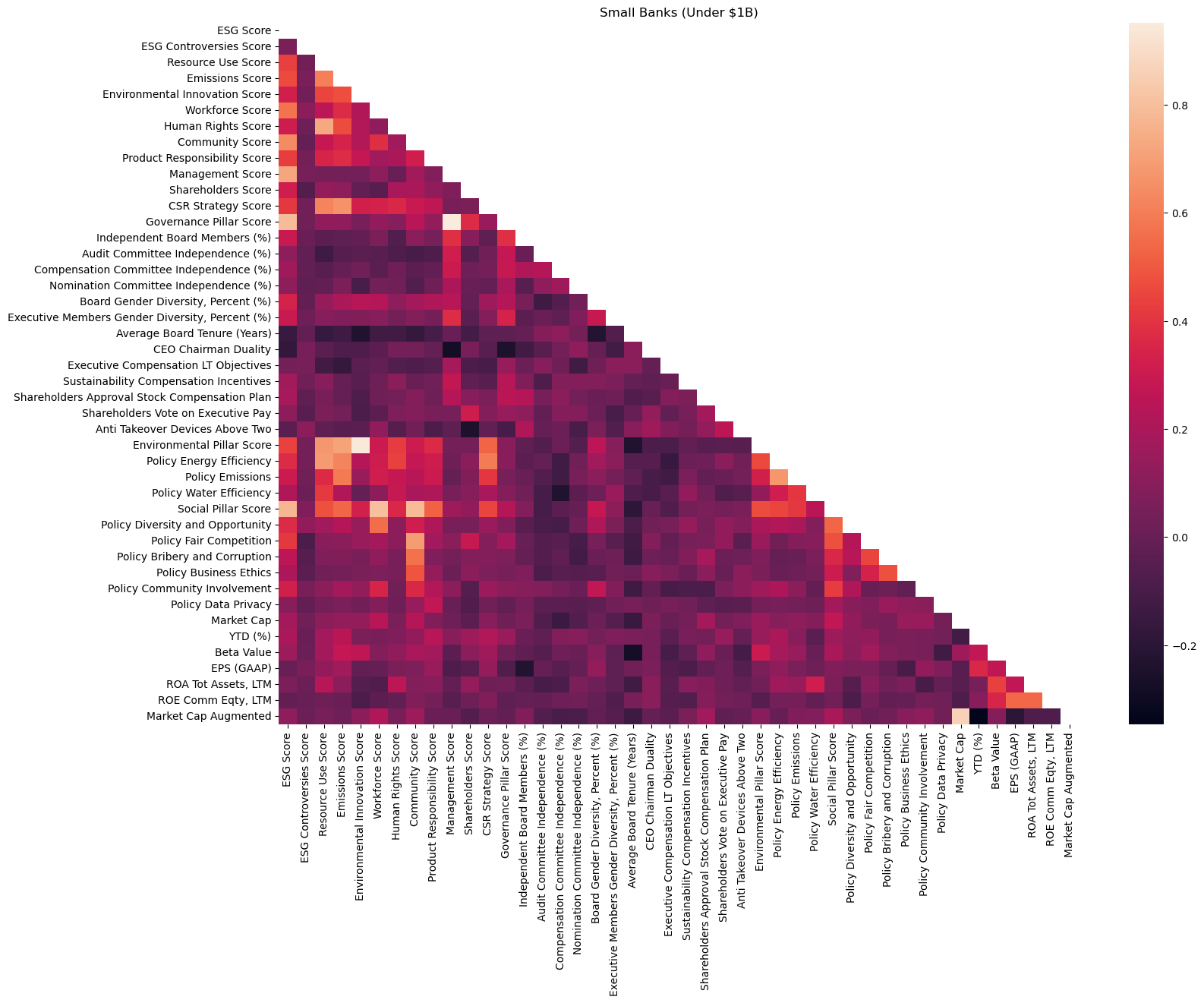
Correlation Coefficient Analysis 2022 Bank Performance



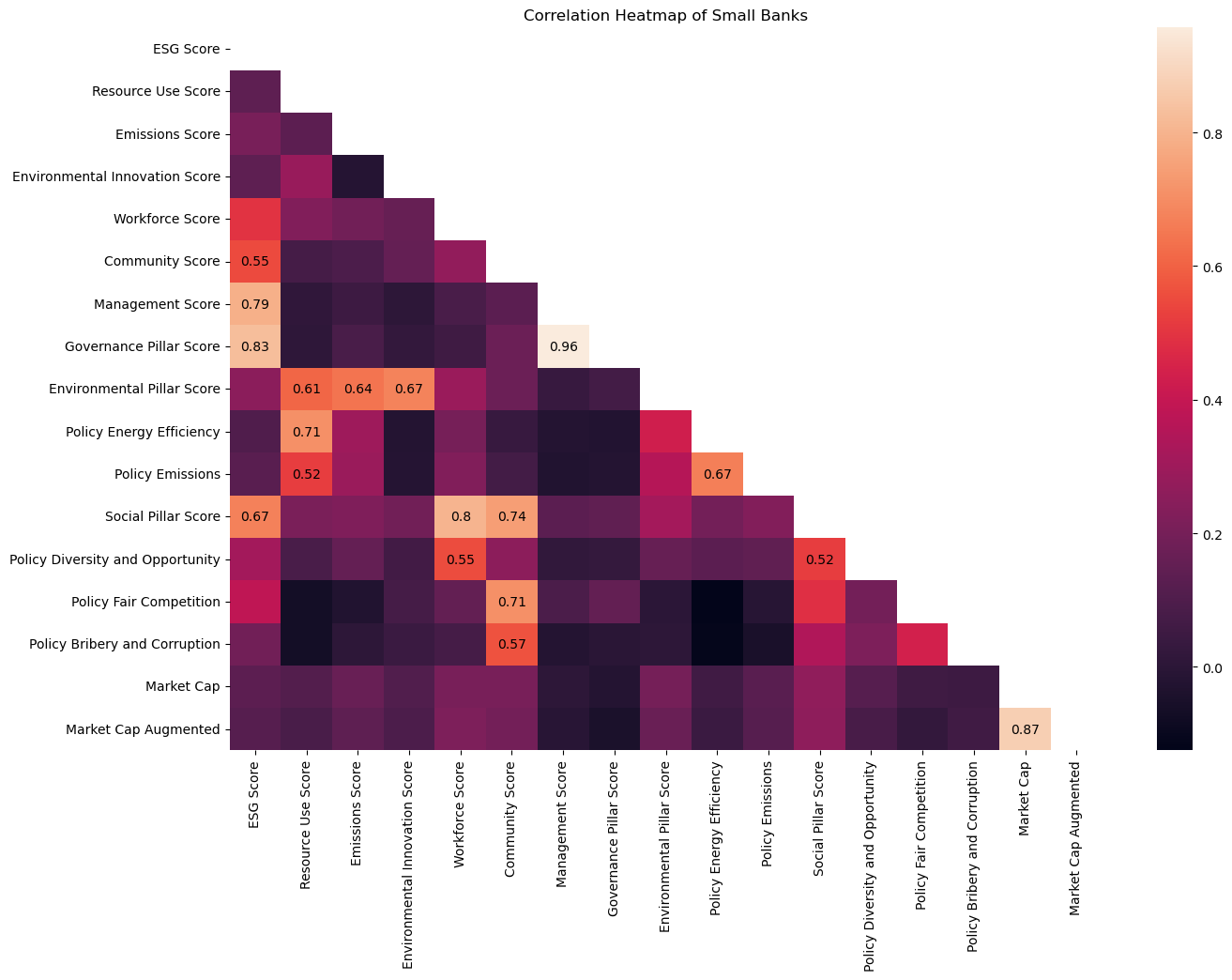
Beta has the highest correlation to ESG Scores out of all KPIs evaluated by a significant margin. Furthermore, all other KPIs had next to no correlation with the ESG score.

Further, we had considered some intra-segment correlational analysis, which revealed an interesting difference with banks that have less than $1B in Market Capitalization, and those with over $1B in Market Cap. There were far more correlations in Larger banks than in smaller ones, with regard to both financial and ESG information, in the most recent fiscal year. This is apparent with the correlation heatmaps included below.

First, the smaller banks, with lighter colors indicating more significant correlations:



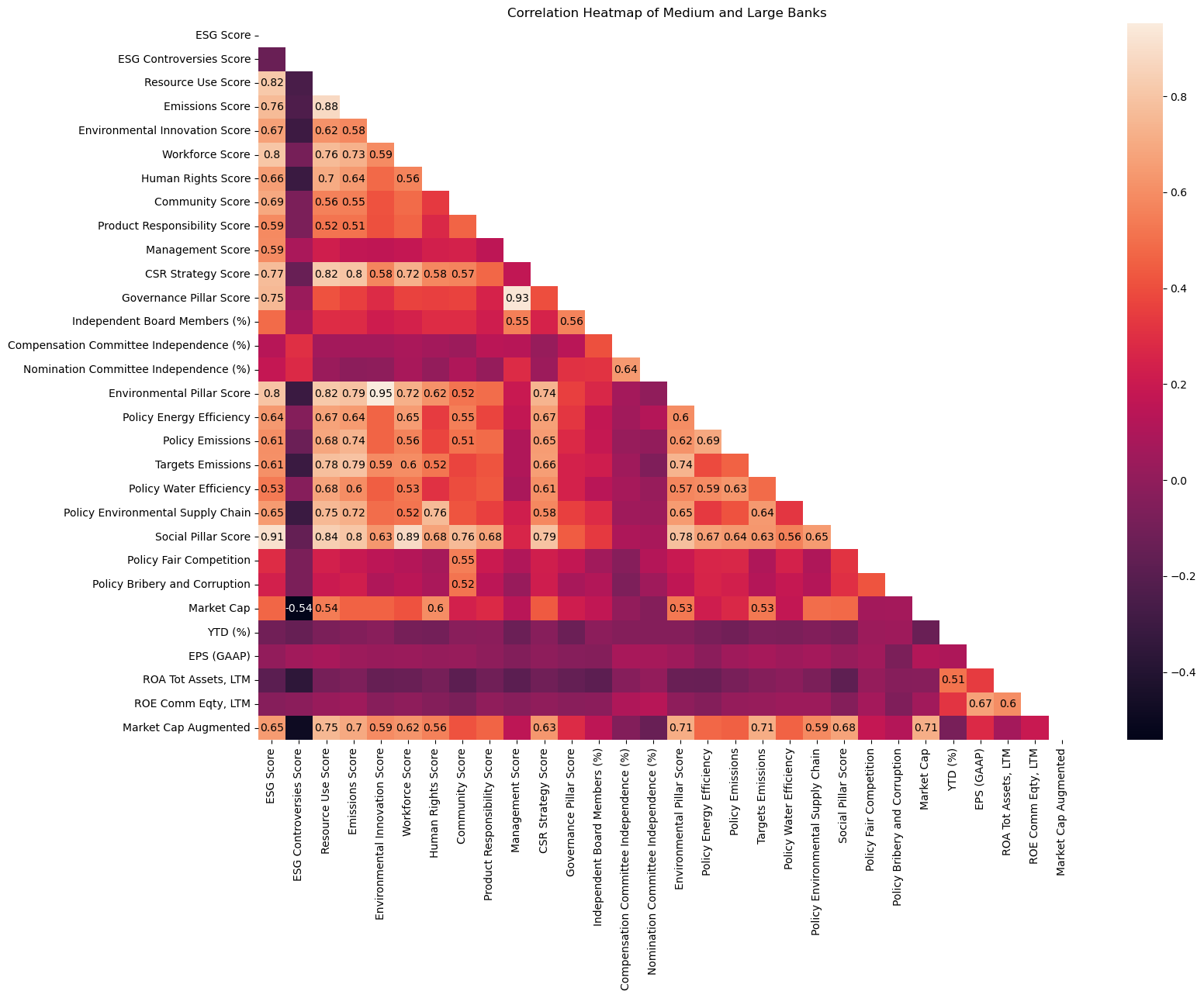
A closer look at the few significant correlations:



As contrasted with the medium and large banks as determined by Market Cap:



And a closer look at the comparably more significant correlations:



This quite stark difference between the two segments, it should be noted, is possibly due to the difference of variance in certain categories between the smaller banks and the medium to large banks. This impacts correlation due to how correlation coefficients are calculated.

# Models

1. **Predictive Performance Models**

We used a range of models, including regression, decision trees, and boosting models, to assess and quantify the relationship between the ESG activities of U.S. banks and their financial performance. The model:

The set of independent variables consists of:

* E Score as ‘e’
* S Score as ‘s’
* G Score as ‘g’
* Beta as b
* Market Cap as ‘mkc’. We then took the natural logarithm of ‘mkc’ to normalize this variable. The independent variable we used is ‘lnmkc’
* Price to Book as ‘pb’. We then took the natural logarithm of ‘pb’ to normalize this variable. The independent variable we used is ‘lnpb’

The targets include:

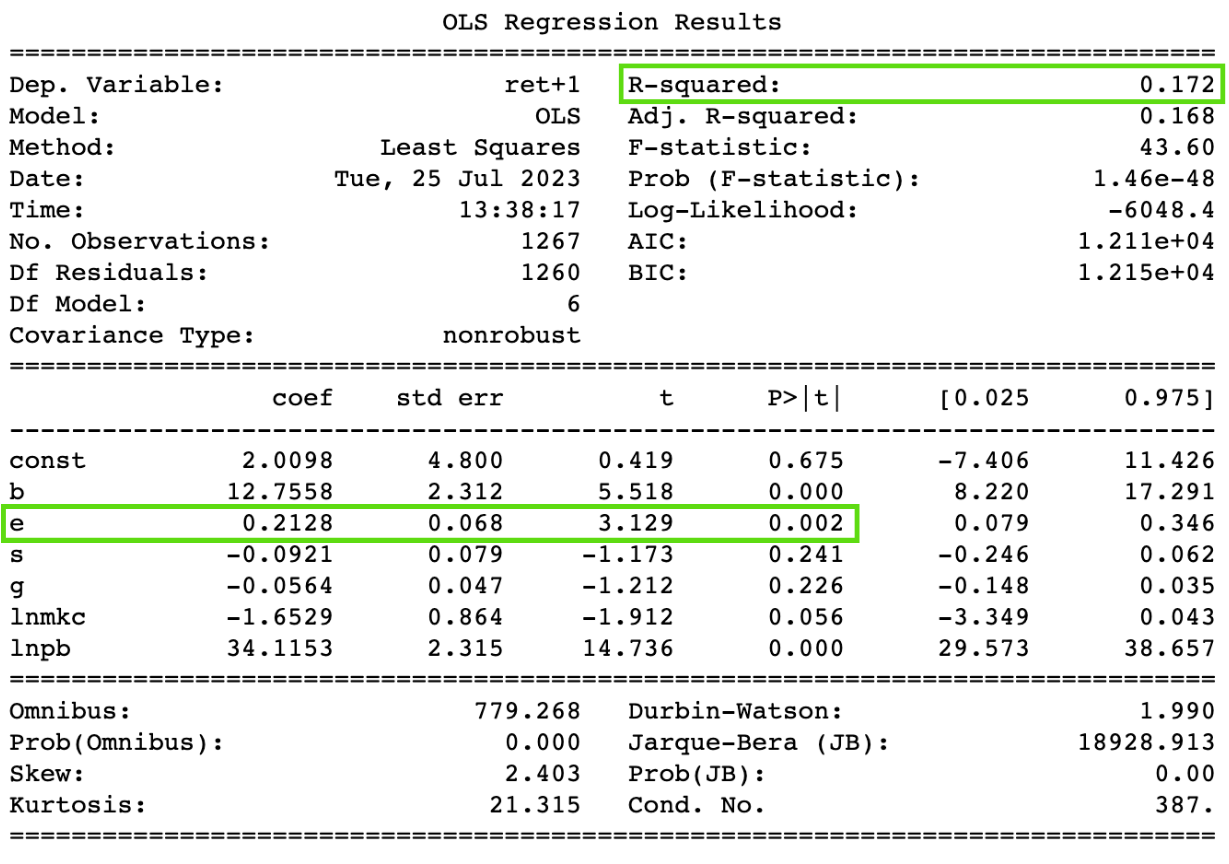


Since we were building predictive model, the target used was Annualized Stock Return Next Year as ‘ret+1’

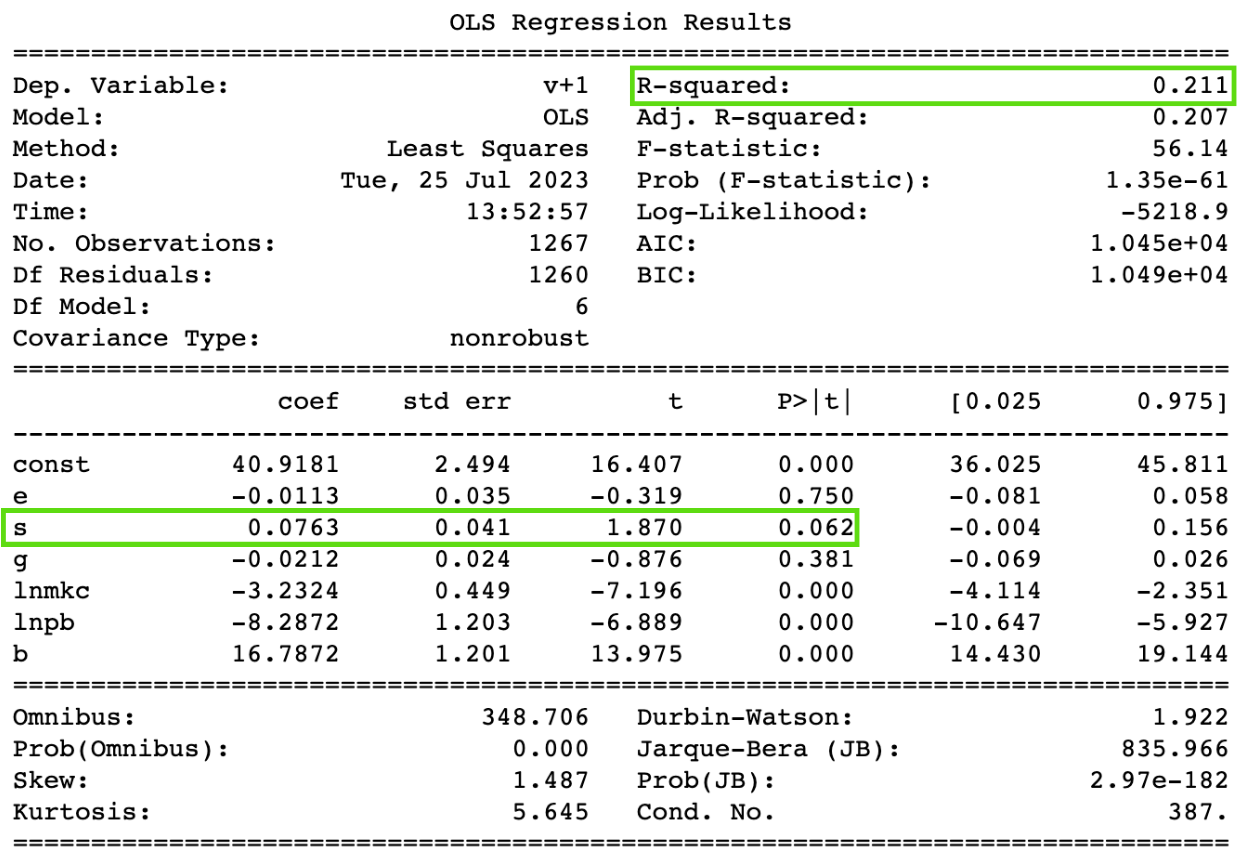
Since we were building predictive model, the target used was Annualized Volatility Next Year as ‘v+1’

We performed all models and went through the same process with both of our targets. We will walk through the process for the Stock Return Model and the results for both Stock Return and Volatility Model.

Our initial model was the ordinary least squares (OLS), performed on 1267 observations, ranging from 2014 to 2021 (targets ranged from 2015 to 2022), but for a number of reasons, OLS did not perform well. The first reason is that each data point's correlations between the predictors and the goal are unique. For instance, the beta of JPMorgan Chase & Co. is negatively connected with the stock price, but the beta of 1st Source Corp. is positively correlated with the same goal. The second reason is that for the Stock Return Model, the S Score and G Score coefficients are not significant (p-value > 0.05); for the Volatility Model, the E Score and G Score coefficients are not significant.

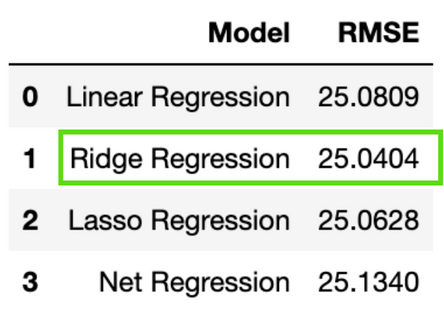
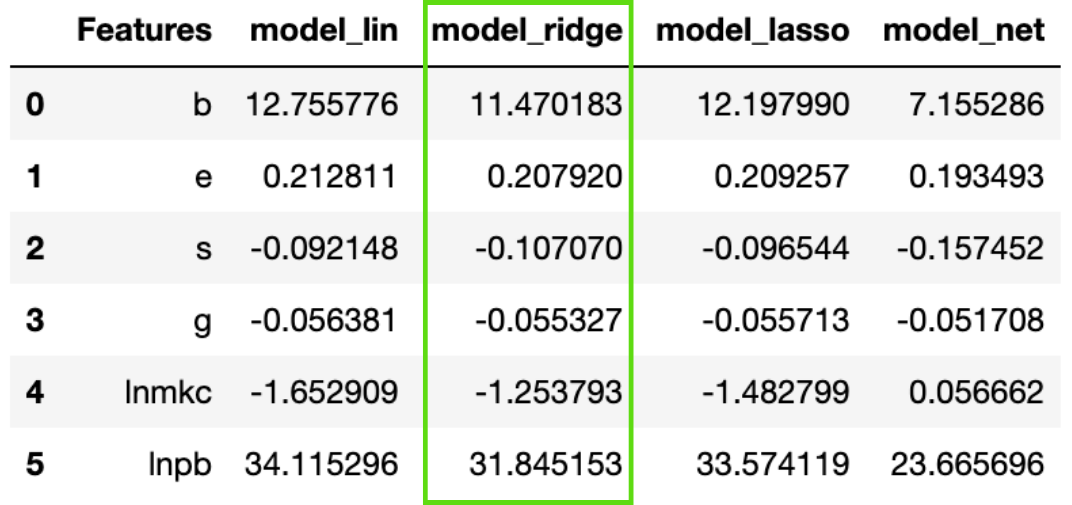


*Stock Return OLS Model Summary*

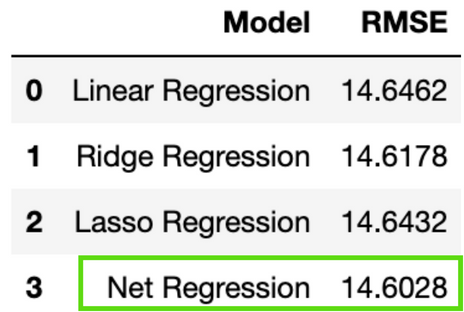
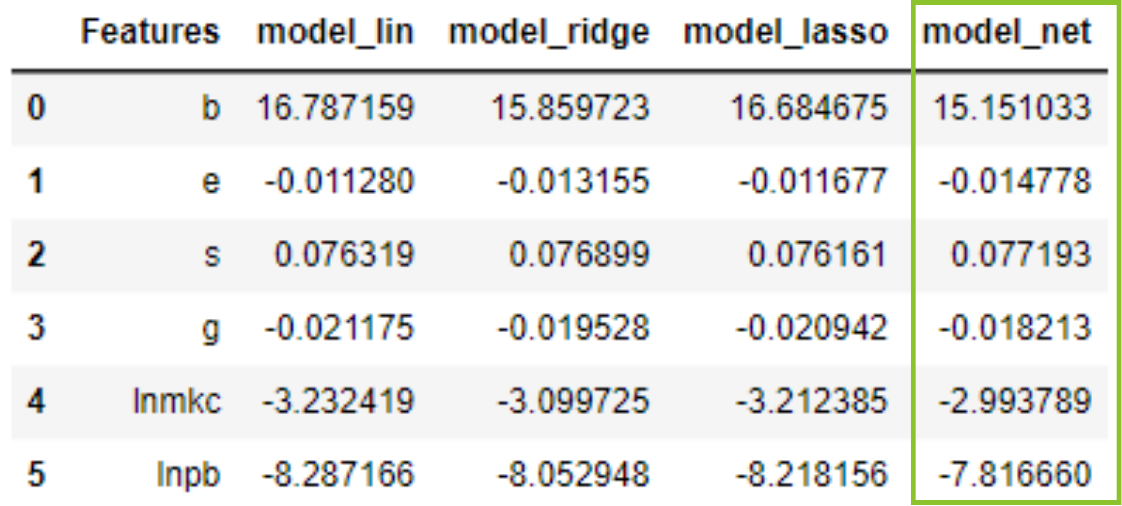


*Volatility OLS Model Summary*

We then tried different regression models like LASSO, Ridge, and ElasticNet to see if they improve performance because these models add regularization which can help with overfitting. The table below shows the model performance metrics as for Stock Return model, Ridge Regression performed the best on the dataset but its RMSE was only a small 0.02 to 0.06 different from the other models; for Volatility model, ElasticNet Regression performed the best on the dataset but its RMSE was only a small 0.01 to 0.04 different from the other models.



*Stock Return Regression Models Summary*



*Volatility Regression Models Summary*

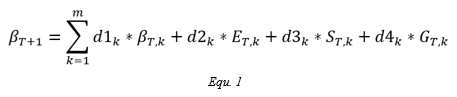
Since we didn't want to drop any of our variables, we then decided to use tree-based regression models. We tried XGBoost regression and LightGBM Regression. However, both of the tree-based models have lower model R\_squared and higher RMSE compared to linear regression. The table below shows the model performance metrics.

| Model | | R\_squared | RMSE |
| --- | --- | --- | --- |
| Stock Return model | XGBoost | -0.023 | 27.75 |
| LightGBM | 0.017 | 27.19 |
| Volatility model | XGBoost | -3.116 | 32.32 |
| LightGBM | 0.023 | 15.75 |

After trying out multiple models and boosting methods, we decided to use Ridge Regression for our Stock Return Model and ElasticNet Regression for the Volatility model. This is our conclusion about the relationship between the targets and the predictors:

* Stock Return is positively correlated with E Score (significant).
  + If E Score increases by 1, Stock Return increases by 0.21.
* Stock Return is negatively correlated with S Score and G Score (not significant).
  + If S Score increases by 1, Stock Return decreases by 0.1
  + If G Score increase by 1, Stock Return decreases by 0.06
* Volatility is correlated with E Score (not significant)
  + If E Score increases by 1, Volatility increases by 0.08
* Volatility is negatively correlated with S Score (significant)
  + If S Score increases by 1, Volatility decreases by 0.02
* Volatility is negatively correlated with G Score (not significant)
  + G Score increase by 1, Volatility decreases by 3

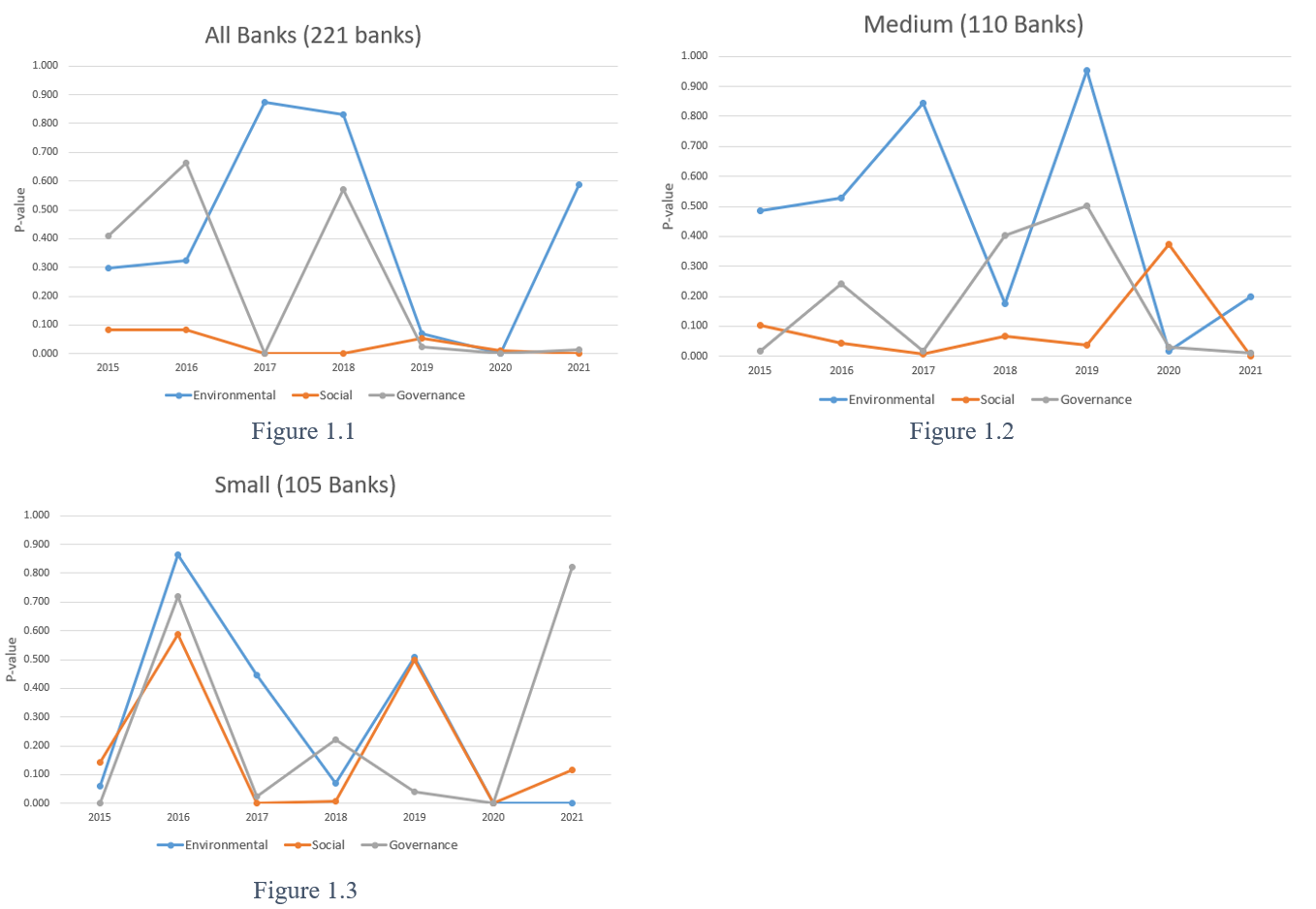
1. **Predictive Beta Model**



The Beta-ESG model focuses on looking at one year's volatility based on the prior year. Considering the data given, beta is used to measure volatility, while E, S, and G pillar scores will be used as predictors. In this model, the previous year’s beta and E, S, and G pillar scores will be the independent variable, and the current year’s beta will be the dependent variable (*Equ. 1*).

Full data used includes the beta, E, S, and G pillar score of 300 U.S banks from the year 2014 to 2021. The banks were segmented into three groups based on market cap: Large/Mid-large, Medium, and Small. Banks with over 30% missing value were dropped. The null values left in the dataset were imputed with each respective column mean. At last, data was scaled using the *normalize* function in scikit-learn (preprocessing.normalize). After the preprocessing stage, we were left with data from 221 banks containing 7 years of E, S, and G pillar scores, and 8 years of beta scores.

**Beta Model: Year-by-Year**



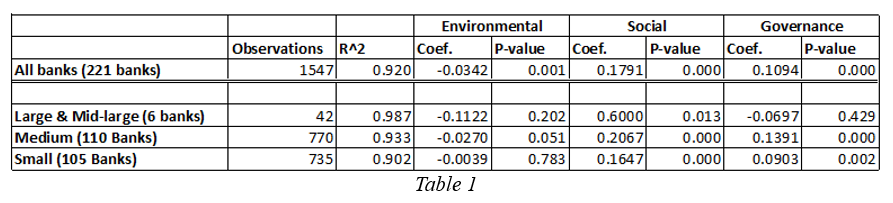
In order to visualize the trend, the Beta-ESG model was run year by year for all groups. Looking at 2015 to 2021, totaling all 7 years of data:

* R-squared has a mean value of 0.923, with a standard deviation of 0.062. The lowest R-squared is 0.726 and the max is 0.990, suggesting the model fits the data well, explaining a large proportion of the variance.
* Environmental Coefficient: The mean value is approximately 0.308, with a standard deviation of 1.61, indicating a high level of variability in the coefficients. The coefficients range from -0.458 to 7.3194. This suggests that the effect of the environmental pillar on the dependent variable varies greatly from year to year.
* Environmental P-value: The mean p-value for the environmental coefficient is approximately 0.387. Since many of these p-values are above the typical significance level of 0.05, it suggests that the environmental coefficient is not statistically significant in many years.
* Social Coefficient: The mean coefficient for the social pillar is approximately -0.0715. The maximum standard deviation of 1.16 suggests high variability. The coefficients range from -5.0923 to 0.5651, suggesting a variable impact of the social factor.
* Social P-value: The mean p-value for the social coefficient is approximately 0.106. Some of the p-values are below the typical significance level of 0.05, suggesting that the social coefficient is statistically significant in some years.
* Governance Coefficient: The mean coefficient for the governance pillar is approximately 0.00725. The max standard deviation of 0.477 indicates high variability. The coefficients range from -2.0185 to 0.3419, indicating a variable impact of the governance factor.
* Governance P-value: The mean p-value for the governance coefficient is approximately 0.225. Some of the p-values are below the typical significance level of 0.05, suggesting that the Governance coefficient is statistically significant in some years.

*Figure 1* plots the p-value of ESG pillar scores by year. The significance of social pillar scores for all three groups shows a relatively stable downward trend. This result is expected, since social pillar scores are weighted heavily in the banking industry. In addition, this downward trend may indicate that ESG scores have an increasing effect on market expectations.

However, the 2015 and 2016 small banks' models are likely to be invalid due to a plethora of missing data. Note that the year-by-year models for the large and mid-large group are missing due to the lack of data input.

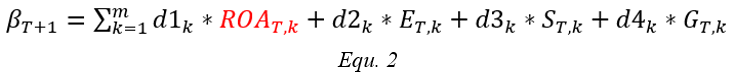
**Beta Model: Overall**



The overall model is the Beta-ESG model was run with data from all years in each group. For example, the overall model for all banks includes all 221 banks multiplied by 7 years – a total of 1547 observations.

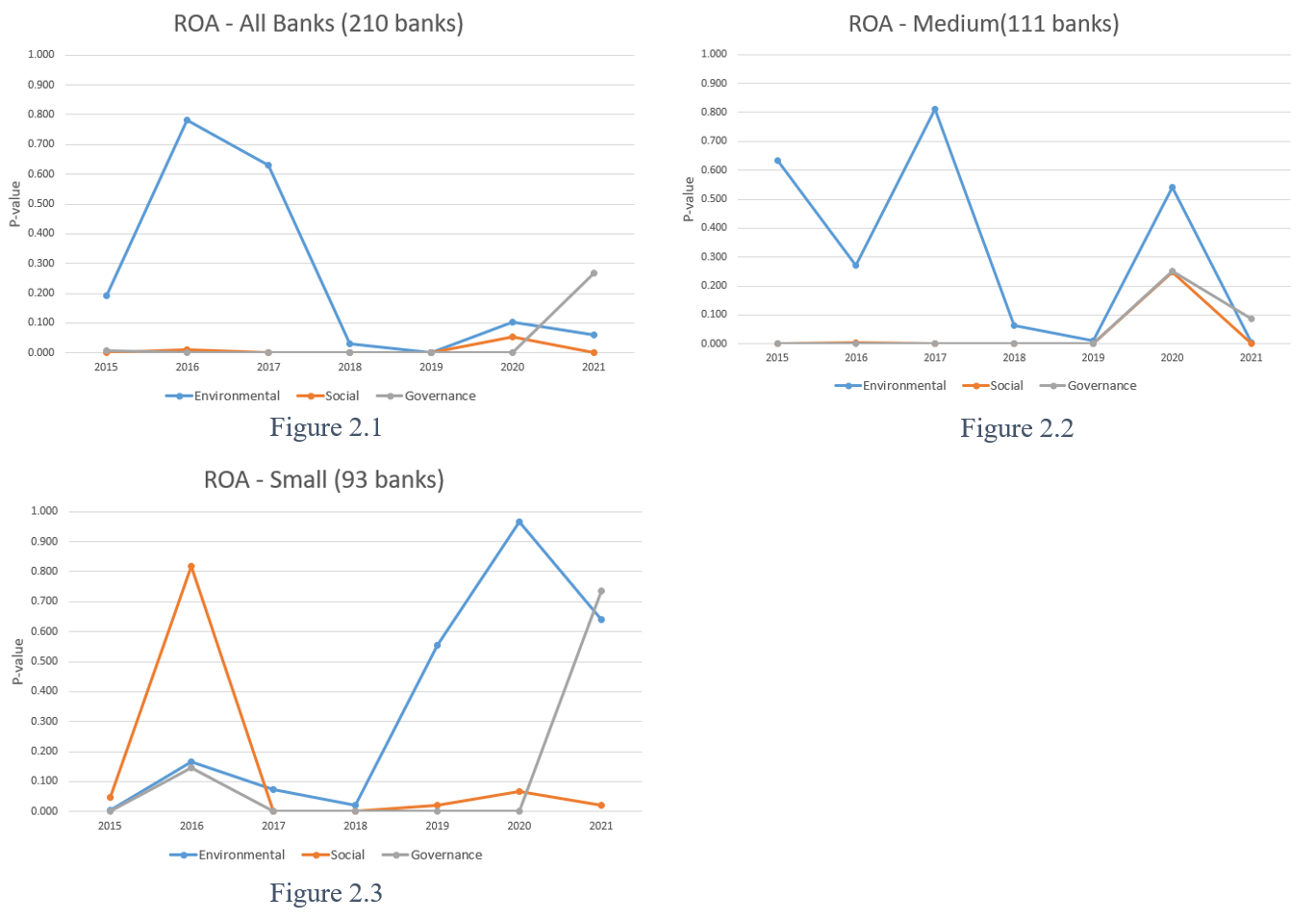
Out of the three groups based on market capital, the large & mid-large group has the worst p-value. The information overflow for larger banks could cause this. ESG scores could be easily overlooked by the market since there is so much information available for the larger banks. However, this situation could simply be caused by the difference in the number of observations, as there are only 6 banks in this group.

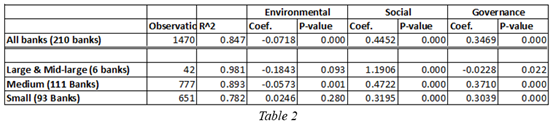
The low coefficient of the environmental pillar score for the medium and small groups could be caused by the imputation and scaling of the data. It is likely because no environmental pillar score was reported during the earlier years for many smaller banks.

**Beta Model: Modified (Return on Asset)**

Furthermore, the modified model shown in *Equ. 2* was run on the same dataset. Instead of using the current year’s beta as a predictor, this model uses return on asset in its place.

The model results are shown below:





*Figure 2* shows a low and stable p-value of social and governance pillar scores for all three groups. There is a clear downward trend in the p-value of the environmental pillar score for both the overall group and the medium group. Table 2 shows the results of the ROA overall models. All models have high R-square and low p-value for ESG pillar scores. Overall, the ROA model result supports the findings from the original model.

Note: *Figure 1 & 2, Table 1 & 2* can be find in file Beta\_model\_pvalue\_plot.xlsx

1. **SVIR Models**

Our objective for the Single Variable Input Regression (SVIR) is to determine which individual components of E, S, and G contribute the most to the overall pillar score. To do this, we first focused on the data selection process. We pulled 136 independent variables chosen from a list of ESG components that Refinitiv deemed most important to the banking industry. These 136 variables comprised 34 Environmental, 46 Social and 56 Governance components. We selected a time span of ten years (2021 to 2012) for each variable that was then averaged for each bank, with nulls being ignored in the calculation. This gave us a data frame of the average overall pillar score and all important components within each pillar.

Before we could begin modeling, we first had to do some data cleaning, since the ESG data is so unreliable. To strengthen the accuracy of our model, we decided to drop any independent variables with more than half of the banks’ averages being null. Independent variables with less than half of the banks’ data being null were imputed with the column mean, resulting in a total of 106 independent variables to be used in the model.

Below are tables showing the three datasets used in the models to predict the overall pillar score:

1. Environmental

* Independent variables (22):

| Environmental Products | Assets Under Management | Equator Principles | Renewable/Clean Energy Products |
| --- | --- | --- | --- |
| Fossil Fuel Divestment Policy | Policy Emissions | Targets Emissions | Climate Change Commercial Risks |
| Environmental Supply Chain Management | Environmental Supply Chain Partner Termination | Staff Transportation Impact Reduction | Environmental Expenditures Investments |
| eWaste Reduction | Policy Water Efficiency | Policy Energy Efficiency | Green Buildings |
| Targets Energy Efficiency | Environment Management Team | Environment Materials Sourcing | Targets Water Efficiency |
| Environmental Partnerships | Environmental Restoration Initiatives |  |  |

* Target variable: Environmental Pillar Score

1. Social

* Independent variables (31):

| Policy Freedom of Association | Policy Bribery & Corruption | Policy Community Involvement | Improvement Tools Business Ethics |
| --- | --- | --- | --- |
| Fundamental Human Rights ILO UN | Human Rights Contractor | Human Rights Breaches Contractor | Policy Fair Competition |
| Policy Child Labor | Policy Business Ethics | Policy Forced Labor | Policy Human Rights |
| Whistleblower Protection | Corporate Responsibility Awards | Policy Diversity & Opportunity | Training & Development Policy |
| Product Access Low Price | Targets Diversity & Opportunity | Employees Health & Safety Team | Employees Health & Safety OHSAS 18001 |
| Health & Safety Policy | Net Employment Creation | Trade Union Representation | Announced Layoffs to Total Employees |
| Women Employees | Flexible Working Hours | Policy Data Privacy | Quality Mgt Systems |
| Salary Gap | Day Care Services | Internal Promotion |  |

* Target variable: Social Pillar Score

1. Governance

* Independent variables (53):

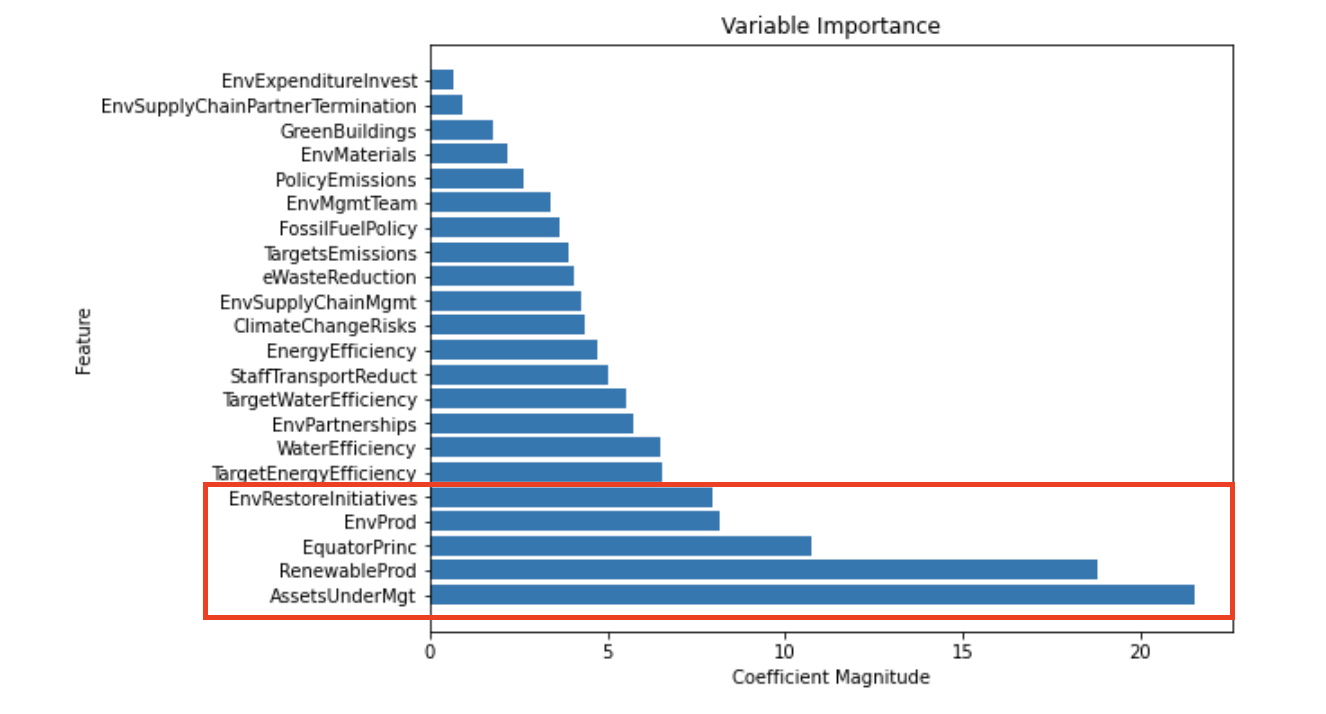
| CEO Chairman Duality | Competition Improvement Tools | Board Structure Policy | Compensation Committee Independence | Internal Audit Department Reporting |
| --- | --- | --- | --- | --- |
| Succession Plan | External Consultants | Audit Committee Independence | Audit Committee Mgt Independence | Board Functions Policy |
| Compensation Committee Mgt Independence | Nomination Committee Independence | Nomination Committee Involvement | CSR Sustainability Report Global Activities | Executive Compensation Policy |
| Board Size More Ten Less Eight | Board Background & Skills | Board Gender Diversity, Percent | Board Specific Skills, Percent | Average Board Tenure |
| Non-Executive Board Members | Independent Board Members | Board Meeting Attendance | Board Member Affiliations | Board Individual Re-election |
| Executive Individual Compensation | Total Senior Executives Compensation to Revenues | Highest Remuneration Package | CEO Compensation Link to TSR | Executive Compensation LT Objectives |
| Sustainability Compensation Incentives | Shareholders Approval Stock Compensation | Executive Members Gender Diversity, Percent | Director Election Majority Requirement | Anti Takeover Devices Above Two |
| Voting Cap Percentage | Shareholder Rights Policy | Shareholders Vote on Executive Pay | Public Availability Corporate Statutes | Veto Power or Golden Share |
| State Owned Enterprise SOE | Equal Shareholder Rights | Non-Audit to Audit Fees Ratio | Auditor Tenure | CSR Sustainability Committee |
| Integrated Strategy in MD&A | Global Compact Signatory | Stakeholder Engagement | CSR Sustainability Reporting | GRI Report Guidelines |
| Board Attendance | CSR Sustainability Internal Audit | UNPRI Signatory |  |  |

* Target variable: Governance Pillar Score

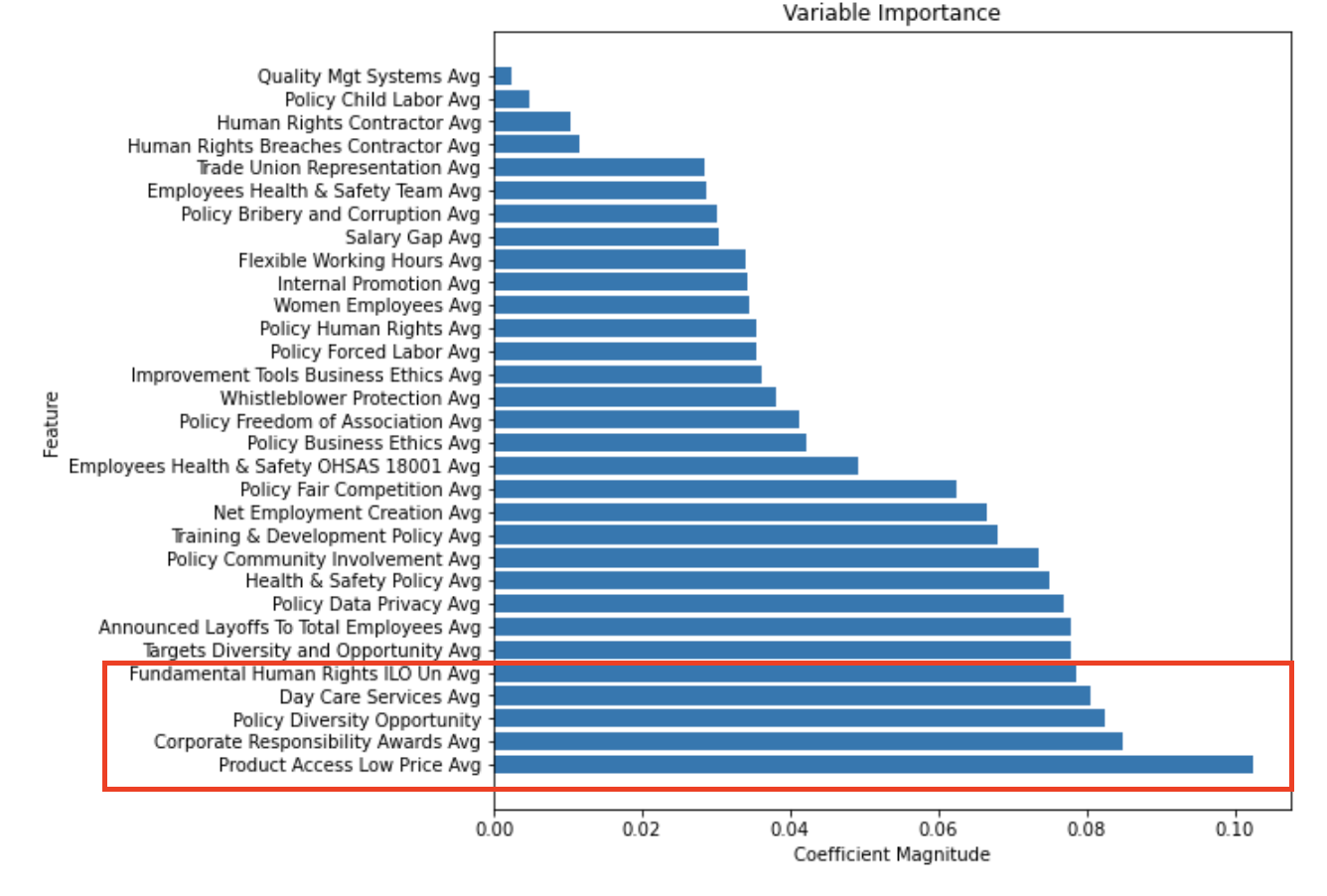
**SVIR: OLS Multiple Regression Model**

Three different regression models were selected to determine the features having the highest importance and predict overall pillar score. The first model we ran was an OLS Multiple Regression Model. The R-squared values for each pillar were very high indicating accurate results; however, they raise concern for model overfitting. Out of the 106 total variables, only 36 were deemed significant with the ranking of importance shown below in the bar charts.

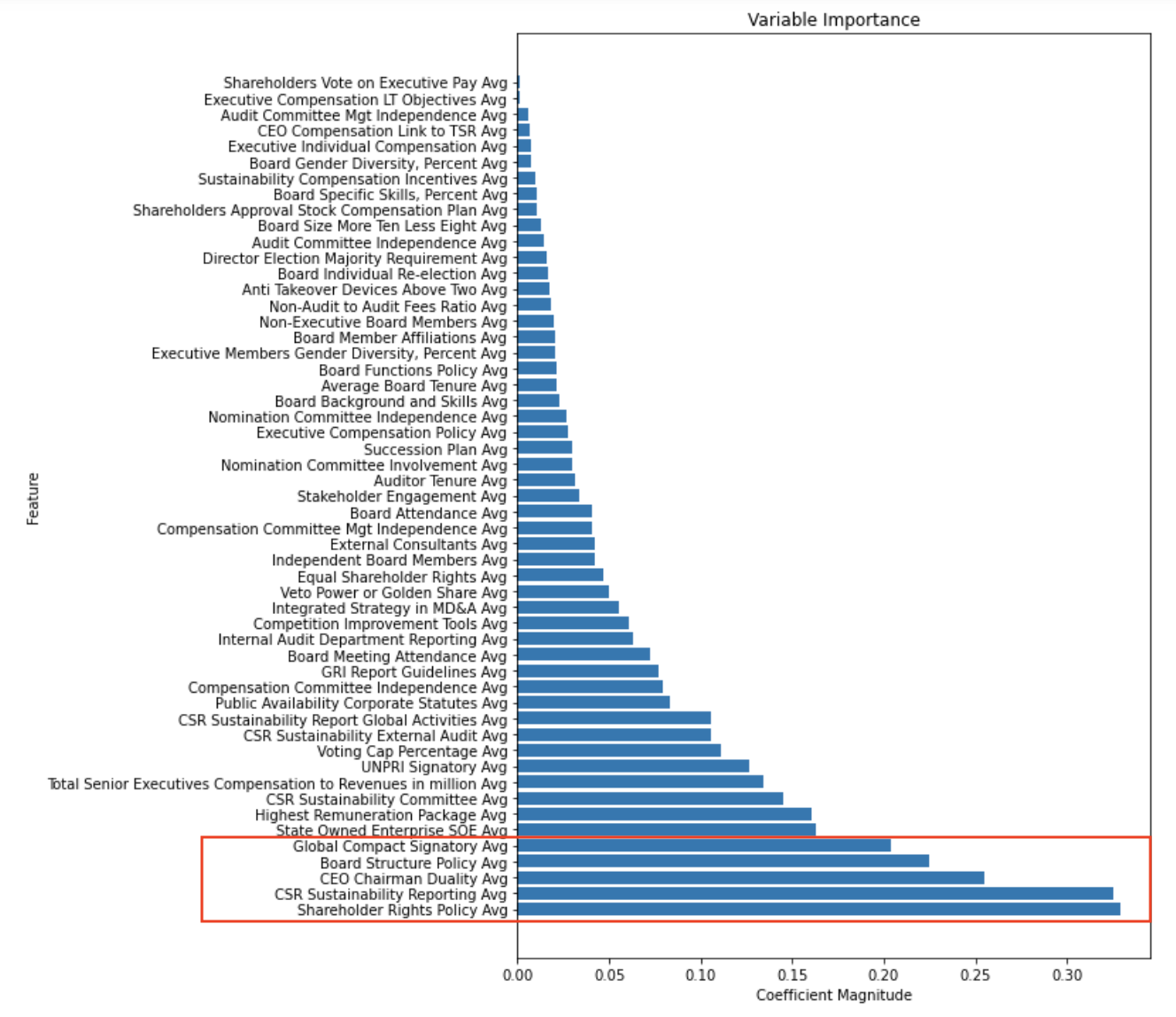
|  | R-squared values | MSE | Significant Coefficients |
| --- | --- | --- | --- |
| Environmental | 0.967 | 23.622 | 11 |
| Social | 0.993 | 8.226 | 17 |
| Governance | 0.870 | 38.908 | 8 |



The above Environmental variable importance graph details the top five most important as Assets Under Management, Renewable Products, Equator Principles, Environmental Products and Environmental Restoration Initiatives.



The above Social feature importance chart shows that the top five most important when determining overall Social score are Product Access Low Price, Corporate Responsibility Awards, Policy Diversity Opportunity, Day Care Services and Fundamental Human Rights ILO UN.



The top five most important features when determining the Governance score are Shareholder Rights Policy, CSR Sustainability Reporting, CEO Chairman Duality, Board Structure Policy and Global Compact Signatory, as shown in the chart above.

**SVIR: Classification & Regression Tree (CART) Model**

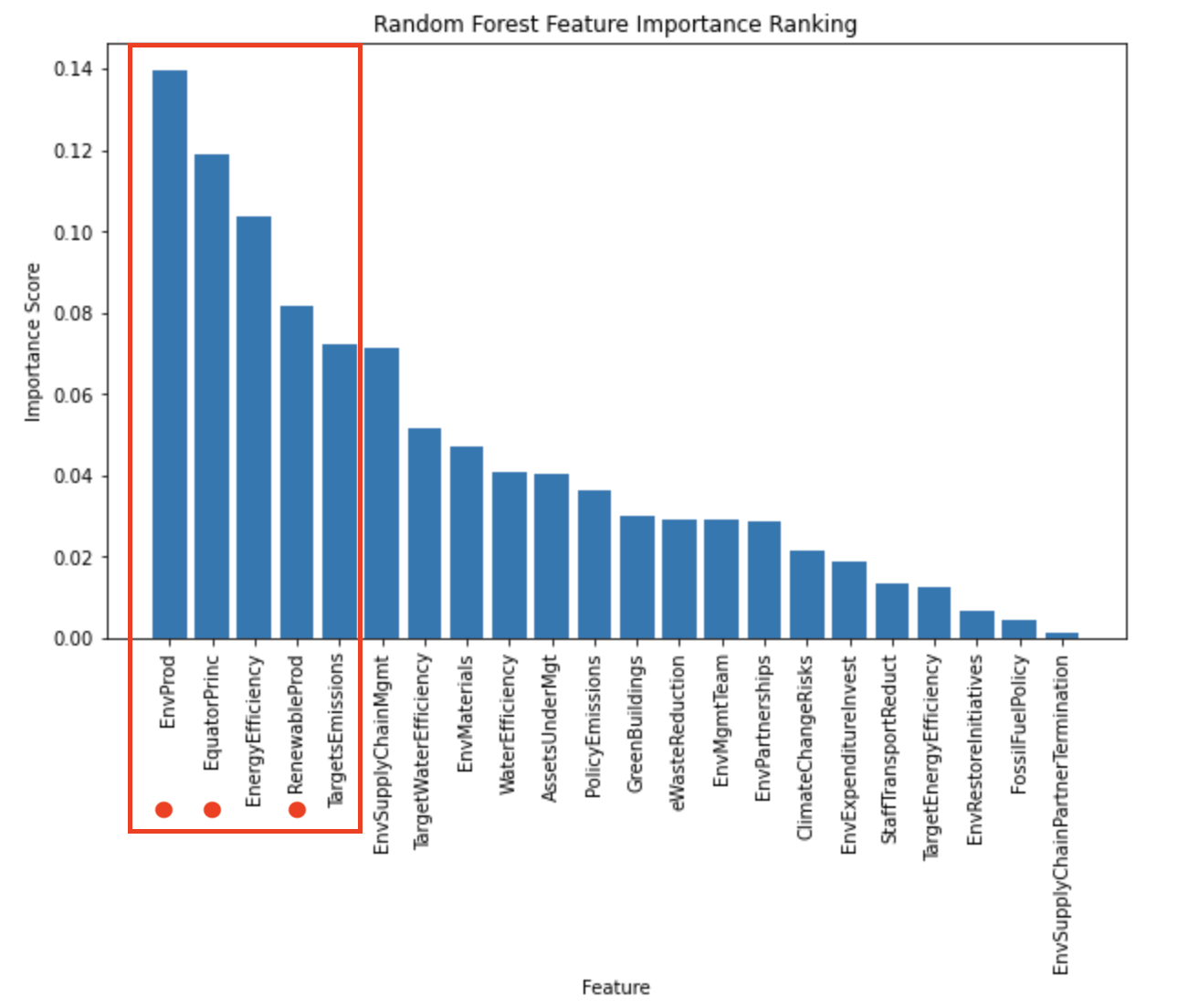
We then wanted to determine the accuracy of the OLS model’s feature importance by running the data through a Classification & Regression Tree (CART) model. The CART model took the format of a Decision Tree, and the overall results were fairly inconclusive. Each of the pillars has a much larger Mean Squared Error than the original OLS model, as shown in the below table. On top of this, the decision tree nodes also had significant squared errors and often branched out using the same feature several times. Since the CART model results were inconclusive, we decided to move on to our final SVIR model, Random Forest Regression.

|  | MSE: |
| --- | --- |
| Environmental | 71.14 |
| Social | 29.83 |
| Governance | 60.29 |

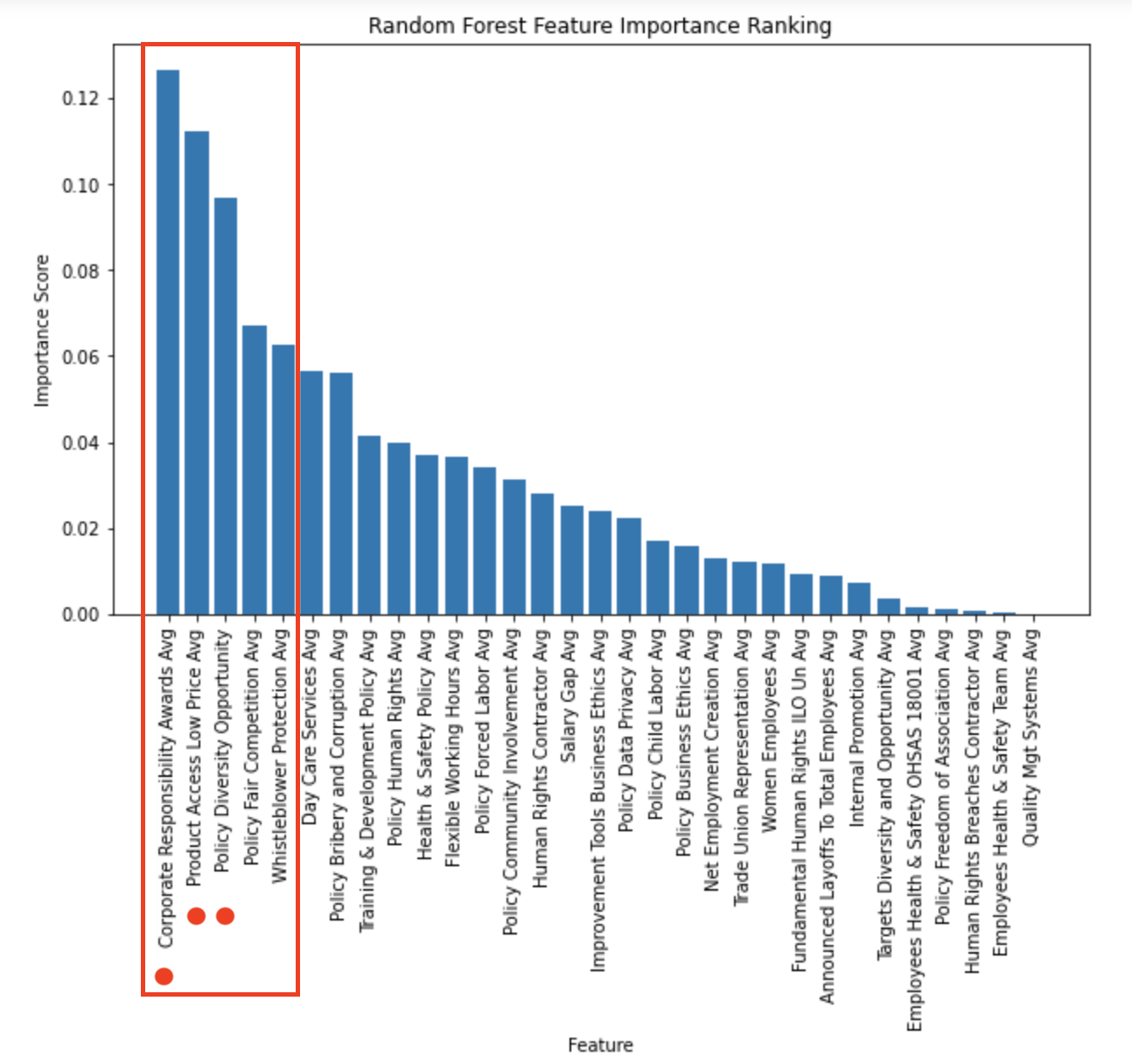
**SVIR: Random Forest Regression Model**

Since the CART model results were inconclusive, we ran a Random Forest Regression model to compare the accuracy and features of the original OLS model. We obtained lower R-squared values for this model; however, the Environmental and Governance Mean Squared Error values were lower, possibly demonstrating a more accurate model due to less overfitting than the OLS Regression.

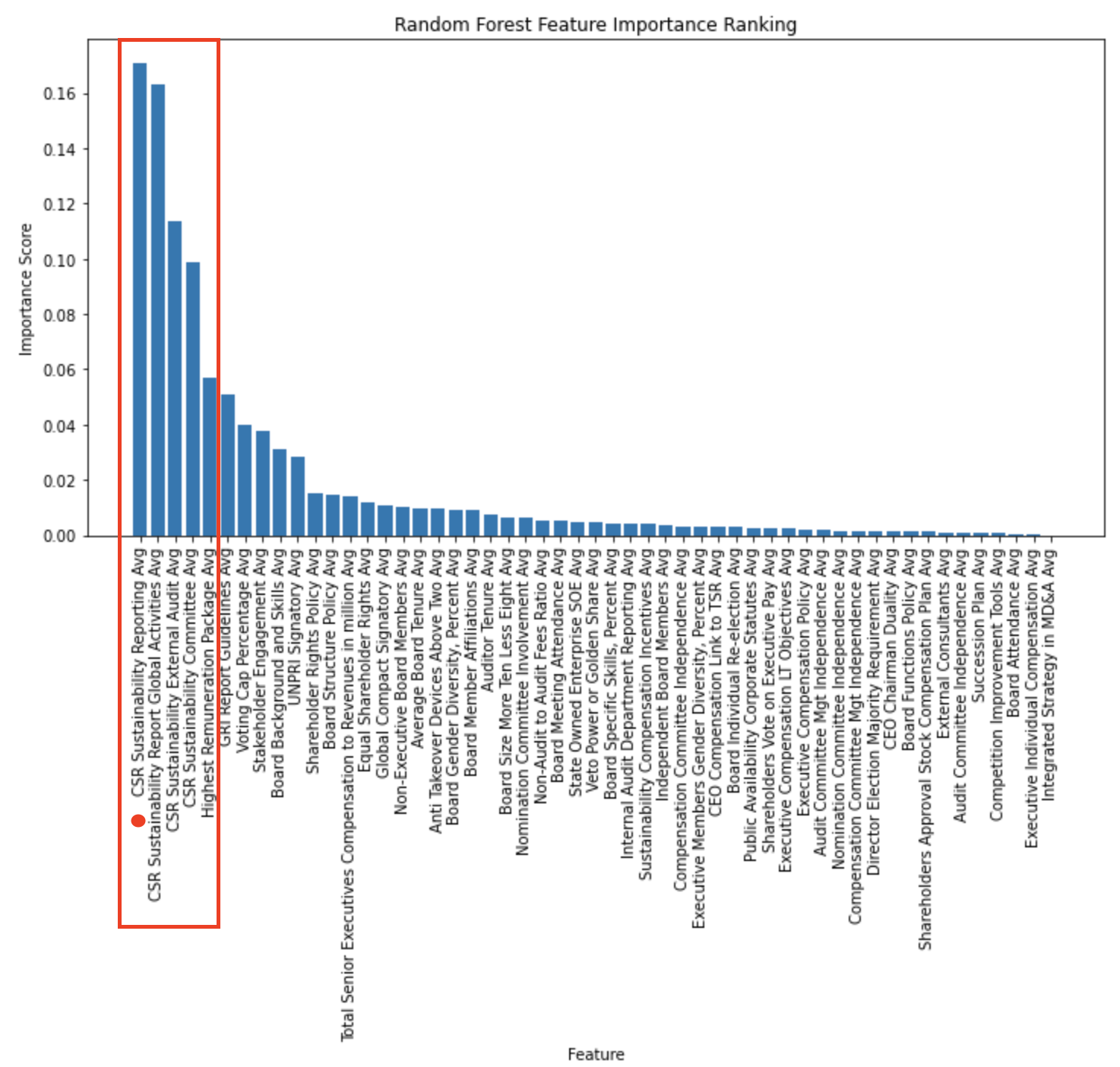
|  | R-squared | MSE |
| --- | --- | --- |
| Environmental | 0.85 | 16.21 |
| Social | 0.86 | 11.52 |
| Governance | 0.62 | 33.17 |



When analyzing the similarities between feature importance, we were able to locate several features that both the OLS and Random Forest models had given high significance to. In the above chart showing Environmental feature importance, there are 3 common features, marked with a red dot, between the two models that are within the top 5 most important. Our models indicate that the Environmental components of Environmental Products, Equator Principles and Renewable/Clean Energy Products have a significant influence on the banks’ overall Environmental pillar score.



The importance of Social components shown above also has 3 common features with OLS. This commonality indicates that the Corporate Responsibility Awards, Product Access Low Price and Policy Diversity Opportunity variables have a significant influence on banks’ overall Social pillar score.



The Governance pillar showed fewer similarities between the top five most important features, with only one appearing in both. The above graph also indicates a significant drop in importance after the tenth feature. This could demonstrate that even though Refinitiv indicates Governance as having the most features being considered, roughly only the top ten components have much influence over the score. The shared component of CSR Sustainability Reporting is shown to have a strong influence in both models, as seen in the above chart.

Note: We assumed a linear relationship between independent and target variables for the OLS and Random Forest Regression models.

1. **Granularity Models**

The final set of models that were considered is what were termed the ‘Granular’ Models, while arguably of less utility, was nevertheless illuminating. These models attempted to make use of the massive quantity of data available in the Refinitiv database. The database has hundreds of data points relating to ESG that it monitors, with a great deal of relatively complete banking data. The algorithms used in this work were the **OLS Regression, Ridge Regression, and Support Vector Regression**. Each model was used to predict three separate target variables: **Stock Return, Volatility, and Price to Book**.

To make use of this plethora of data, this began with importing the necessary ESG data for all 10 years considered. In addition to the Financial data, for each year 244 total fields were left at the end of the initial data import step. Because of the calculated nature of the Stock Return and Volatility information, there was data for both fields for FY 2022, that was not collected for Price to Book value per share. For those fields, models were constructed for FY 2022 that were not constructed for Price to Book. All in all, there were 10 years of models for Stock Return and Volatility, and 9 years for Price to Book. With the exception of Market Capitalization, the fiscal year’s ESG data was paired with the next fiscal year’s financial information. This was done with the aim of creating a predictive model using only ESG and related information. For all 10 years the number of banks was winnowed down using only banks that had ESG scores for the 10 years in question. Further a column of Market Capitalization transformed by taking the natural logarithm, and a categorical ordinal of Size determined by the untransformed Market Capitalization features. Further information such as Return on Assets, Return on Equity, and Beta was prepared and scaled but did not make it into any of the final models. This left the dataframes as follows:

| Fiscal Year | (Banks, Features) |
| --- | --- |
| 2022 | (300, 126) |
| 2021 | (297, 130) |
| 2020 | (293, 129) |
| 2019 | (286, 128) |
| 2018 | (265, 129) |
| 2017 | (238, 127) |
| 2016 | (130, 125) |
| 2015 | (66, 128) |
| 2014 | (34, 136) |
| 2013 | (34, 137) |

For each year, the reason for the number of features given is because the features had more than 70% of their data missing, due to missing data and the number of banks in question, this 70% threshold manifested differently for different years. And then of those:

If the column had boolean (True/False) values:

1. more than 80% of their values were True, OR
2. more than 80% of their values were False.

OR for the columns that were Numeric: more than 80% of their values were = 100, this was for columns with percentages for values or scoring that was out of 100.

Missing values were imputed using the mean of the intra segmentation values for that particular feature. Given that Market Capitalization changed from year to year, so too did their size designation, ensuring that imputed values were accurately reflective of their segment in that year. For the numerically valued fields, normalization was done to scale each column to between 0 and 1. Furthermore, to reduce the number of features, correlational analysis was done to reduce the number of features to only those without any significant correlations in any of the years, This reduced the number of features to 84 in the case of the Price to Book models, and 80 in the Stock Return and Volatility models. Of these features, using earlier correlational analysis done on ESG data used in scoring, the features were reduced to these 9: 'Average Board Tenure', 'Independent Board Members', 'Executive Members Gender Diversity, Percent', 'Board Gender Diversity, Percent', 'Policy Fair Competition', 'Product Responsibility Score', 'Management Score', 'Anti Takeover Devices Above Two', and 'Shareholders Score'.

Another model was used on the subscoring within the E, S, and G pillars, using Resource Use Score, Emissions Score and Environmental Innovation Score from the Environmental pillar, Workforce Score, Human Rights Score, Community Score and Product Responsibility Score from the Social pillar, and Management Score, Shareholders Score and CSR Strategy Score from the Governance pillar. Neither of these models performed well.

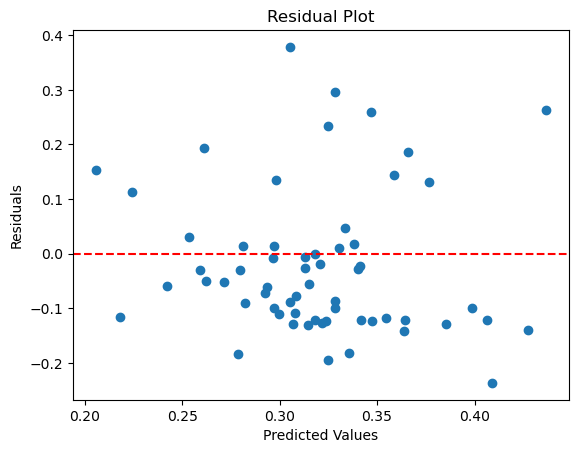
**Results from Support Vector Model**

**Target: Price to Book**

Mean Mean Absolute Error (MAE): 0.079087623080858

Mean R^2: 0.1621802346089828

This is the residual plot based upon the original test set:



The Support Vector models were typically the worst, and were mostly used to contrast and understand the difficulty in drawing meaningful natural distinctions in the data.

**Results from Ridge Model, FY 2022**

**Target: Volatility**

Mean MAE: 0.056540042619307694

Mean R^2: 0.06905844309390427

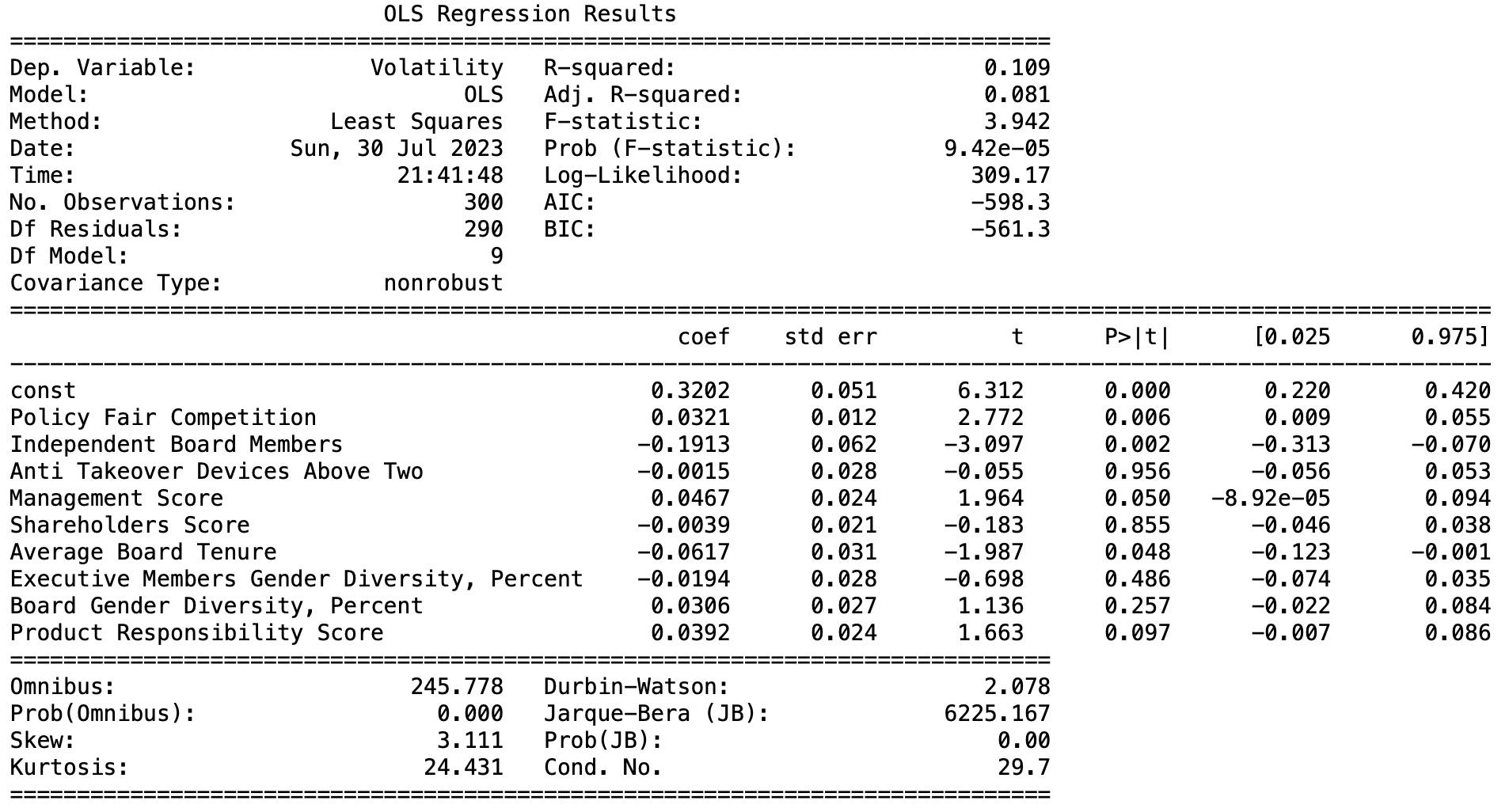
| **Feature** | **Coefficient** |
| --- | --- |
| Policy Fair Competition | 0.037475 |
| Independent Board Members | -0.141032 |
| Anti Takeover Devices Above Two | -0.007580 |
| Management Score | 0.039947 |
| Shareholders Score | -0.026904 |
| Average Board Tenure | -0.085835 |
| Executive Members Gender Diversity, Percent | 0.003654 |
| Board Gender Diversity, Percent | 0.037088 |
| Product Responsibility Score | 0.022124 |

**Results from OLS Regression, FY 2022**

**Target: Volatility**

Mean MAE: 0.05600178937483312

Mean R^2: 0.09538010398033211



Most of the best models came out of an OLS model p-value maximum reduction algorithm. In this algorithm, Statsmodels’ p-value calculation was utilized for each coefficient, and the highest p-value coefficient was removed, until the highest was at or below 0.04. The starting features were the 84 features selected at the initial correlational selection step. This is a better (relative) model developed for FY 2016 with Stock Return as the target. However the nature of the algorithm ensured that it was not stable, sometimes drastically so, from year to year. Furthermore, granting that number of variables for some models generated this way was quite large, issues with multicollinearity persist, despite the feature selection done.

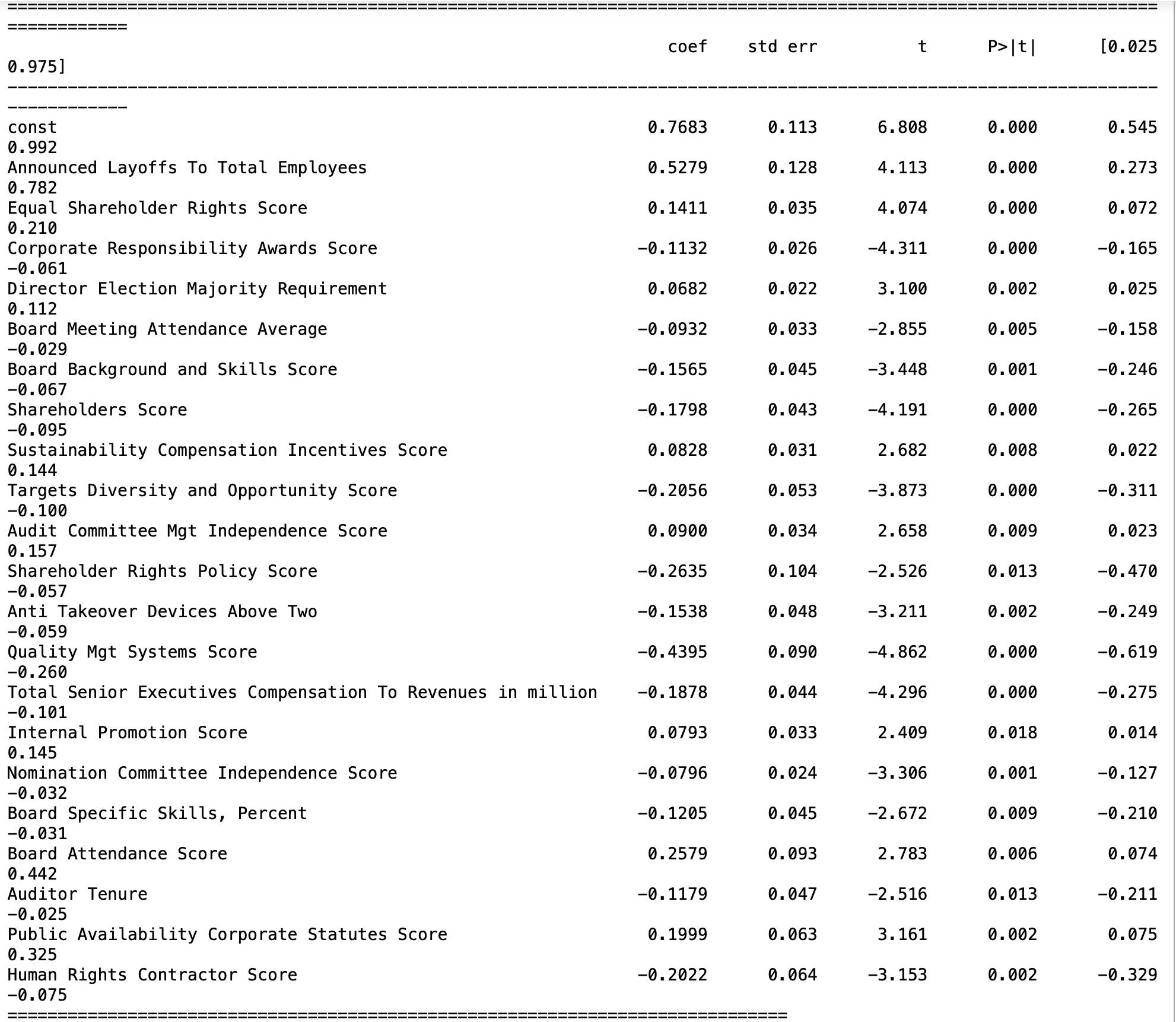
**Results from OLS p-Value Reduction Model, FY 2016**

**Target: Stock Return**

Mean MAE: 0.05963211613264909

Mean R^2: 0.48338119024015

This is the coefficients of the model and associated p-values:



In truth, at the end, most of the models generated negative R^2 values, something relatively rare in data modeling, and reserved for models that were exceptionally awful at predicting the target feature. Even those models that were predictive were not consistent across all 9 or 10 years. Another key takeaway is the target variables, the most predictive models, in general, tend to be Price to Book. Volatility and Stock Return models on the other hand tend to be very inaccurate across all models, from the perspective of mean absolute error and R-squared. This is to say that with a few exceptions for certain years and features which are not consistent from year to year, ESG data is a terrible predictor of financial data, regardless of the specific ESG features in consideration. This is meant as a summary of notable parts of these models, in truth, across the different sorts of Granular data models and years, hundreds of models were generated. A complementary user interface was constructed to facilitate experimentation by the end user.

# Limitations

There are a number of limitations to evaluating the relationship between ESG scores and financial performance that we must address. The majority of the project’s difficulties can be directly mapped to the unreliability of the data itself. One significant problem is the vast amount of missing data, particularly as a result of banks' historical underreporting of ESG scores. Since the relevance and importance of ESG metrics have been recognized only recently, this has led to a lack of comprehensive historical data to analyze. As a result, conclusions drawn from the current dataset may be incomplete or biased.

The subjective nature of ESG scores is another crucial complication. The dataset is a combination of ESG Scores provided by different auditing and rating agencies that can interpret the scoring criteria in different ways. Comparisons between banks are difficult due to the lack of a consistent approach for quantifying the criteria of ESG rankings, a problem that is exacerbated by firms that switch auditing agencies from year to year. Consequently, the conclusions drawn about the relationship between ESG performance and financial indicators may differ depending on the results derived from such subjective sources.

Additionally, the influence of large banks on the dataset presents another challenge. The top six banks with the highest market capitalization significantly skew the overall analysis. These large financial institutions may have unique characteristics and business practices, making their financial performance and ESG scores distinct from medium and small banks. The skewness can introduce biases and distort the true relationship between ESG scores and financial outcomes in the overall dataset.

Moreover, we assume a linear relationship among the variables in this research. While a linear model simplifies the analysis and interpretation of data, it may not fully capture the complex and nuanced interactions between ESG and financial performance. In reality, the relationship can be complex and nonlinear. Nonlinear correlations between ESG characteristics and financial metrics can occur as a result of a variety of circumstances, including threshold effects, diminishing returns, or interaction effects between different ESG dimensions. Neglecting potential nonlinearities may result in oversimplified findings or missing important insights into the underlying nature of the relationship between sustainable practices and financial performance.

Acknowledging these limitations, the project provides valuable insights into the current state of the relationship between ESG scores and financial performance. To acquire more precise and thorough results, future research will surely benefit from resolving these issues. Researchers may better grasp the true impact of ESG practices on financial performance throughout the banking sector by collecting more extensive historical data, standardizing ESG scoring systems, and properly accounting for the influence of major banks. Overcoming these obstacles will be essential in developing our knowledge of the connection between sustainable practices and financial success as the significance of ESG elements continues to rise.

# Conclusion

After a comprehensive analysis of ESG metrics and their impact on the financial outlook for multiple US banking institutions, we conclude that there is not sufficient enough evidence to show a significant relationship between ESG and the two financial indicators. Using a series of regression models, predictive performance models, predictive beta models and granularity models, it is difficult to claim that the financial performance is directly affected by the ESG performance, despite the growing suggestions in recent publications. However, it is important to recognize the limitations of our sample as much of the data was not available for the allotted ten-year span. Likewise, the KPIs selected for analysis could play a role in the findings. Future work could include changing the KPIs to determine other attributes not investigated by this research, as well as allowing for more data to be obtained in future ESG recordings. There is also the possibility of implementing different strategies to account for the missing data in the sampled years. While the data does show a lack of correlation between the two, it is important that companies still continue to monitor their ESG factors as sustainability procedures and responsible corporate practices by a banking institution are important to promote overall business longevity. As time goes on and the practice of measuring ESG data improves, it is possible that similar analyses in the future may yield more definitive results. In terms of deliverables, the findings of this research can be seen in the project presentation, the Jupyter notebooks, and the data files we obtained.