**Asset Return Prediction based on Decomposition of ESG Factors**

C. Ma, T. Shi, A. Yuan

School of Accounting and Finance, University of Waterloo

April 2023

DISCLAIMER

This report was prepared exclusively as an assignment in the AFM 423: Topics in Econometrics class at the University of Waterloo. This is a working paper and is meant only to be used solely as a documentation of the analysis done for the assignment. We do not accept liability for any errors in the report or if the report is used for any purpose other than that which it was created. A copy of the report can be requested from the authors.

**Asset Return Prediction based on Decomposition of ESG Factors**

C. Ma, T. Shi, A. Yuan

School of Accounting and Finance, University of Waterloo

April 2023

**Abstract**

ESG investing has grown in popularity over the last several years due to trends in the market and changing perceptions of the public. This paper researches the effect of decomposition the E, S, and G factors in factor investing. The paper aims to determine the level of significance of the contribution of each of the factors and how much the factors need to be decomposed for the models to be affected meaningfully. Other papers cited have demonstrated that the governance factor is the most significant due to the commonalities in governance risk across companies operating in all industries (Giese et. al, 2021). This paper means to further examine the effect of decomposing the E, S, and G scores and to what level they need to be decomposed to through comparing the objective loss functions of the neural net, random forest, autoencoder, and LASSO models. This report concludes that predicting stocks returns solely on ESG factors and its decomposition produce very poor results. Based on models built, in four-factor level, environment is consistently important while controversies score is least significant, and in eleven-factor models, human rights score and is consistently significant while community score, workforce score, shareholders score, and emissions score are among the least important ones. Furthermore, there is no advantage of further decomposing the standalone E, S, and G scores further one level and the best model out of the ones listed is the LASSO model. However, the above observations are all based on poorly performed models so is not as dependable.

**Introduction**

Using ESG as a factor in factor investing has gained greater traction in recent years. This paper is the second in a two-paper series in analyzing the effect of incorporating ESG scores as a factor in factor investing. In this paper, we focus on examining the following questions:

1. Which of the ESG factors is the most important? (i.e. has the largest effect on the return?)
2. What is the effect of decomposing each of the ESG factors?
3. How far do we need to decompose ESG factors? Would further decomposing each individual ESG factor improve the performance of the models?
4. What is the most appropriate model to use to evaluate these questions?

Several papers have researched the effect of treating ESG as its own factor. “Factor Investing and ESG Integration” by MSCI examines the effect of ESG scores by creating ESG constructed portfolios and comparing its results against the standard portfolio and uses this technique to evaluate the effect of using ESG constructed portfolios in conjunction with different factor investing strategies. In this paper, it was found that larger companies had higher ESG scores which was especially noticeable in their E scores (Melas et al., 2017).

Other papers have decomposed each of the E, S, and G scores to perform similar analysis. In “Deconstructing ESG Rating Performance: Risk and Return for E, S, and G by Time Horizon, Sector, and Weighting” by Giese, it was demonstrated that each of the E, S, and G factors are weakly correlated with each other. Additionally, it was found that over a short-term period of one year, G is the most significant factor (Giese et al., 2021). On the other hand, it was found that E and S resulted in greater erosion risk (Giese et al., 2021). It was also shown that the significance of each of the decomposed factors changed depending on the type of industry where the G factor was more significant to companies in the finance industry while the E factor was more significant to companies in the materials industry (Giese et al., 2021).

Another paper, “The role of ESG performance during times of financial crisis: Evidence from COVID-19 in China” by Broadstock et al. used a similar method in decomposing sub-scores for E, S, and G factors and found that cumulative returns are positively correlated to the E and the G factor, but not the S factor (Broadstock et al., 2021). The paper also found that G was the most important factor since G risk is similar across all industries, while E and S risk vary across sectors (Broadstock et al., 2021). The data used in this paper had three layers of decomposition where the individual E, S, and G factors represented one layer of decomposition and the secondary layer included 12 additional criteria related to each of the E, S, and G factors including:

* Environment: environmental management, environmental disclosure, environmental controversies
* Governance: business ethics, corporate governance, governance controversies
* Social: employee, supply chain, community, product, philanthropy, and social controversies

The third layer included over 300 related criterial formed from around 1,000 data points where all the criteria in all of the layers were given a score out of 100 (Broadstock et al., 2021).

Across these papers, we have found that G seems to be the most significant factor. The insights from these papers help us form a guideline in comparing our results and models against that of others.

In this paper, we will first introduce the data used under **Variables and Measures** to discuss the predictor and the response variables. We will then go into the **Methodology** to discuss the models chosen for our analysis and our reasoning behind selecting each model. We will then showcase the results of the models pertaining to our three research questions above in the **Analysis** section and then conclude on our findings in the **Conclusion**.

**Variables and Measures**

The source of data used in this part is Thomson Reuters. While providing aggregated numerical ESG scores ranging from 0 to 1, Thomson Reuters also provide decomposed ESG scores in two levels into the following variables:

|  |  |  |
| --- | --- | --- |
| Level | Variables/Factors | Number of Observations |
| 1 | Environment score, corporate governance score, social score, controversies score | 3773 |
| 2 | Controversies score, Resource use score, Emissions score, Innovation score, Workforce score, Human rights score, Community score, Product Responsibility Score, Management score, Shareholders score, CSR Strategy score | 3797 |

*Table 1: Decomposed variables of ESG factors*

Please notice that controversies score has replaced the role of economic score since 2018, so we use the previous one. The controversies score has been included in both levels since it is considered by Thomson Reuters as a fourth pillar in the rating and has no further decomposition.

The response variable, which is the stock return, has been calculated based on stock price obtained from Yahoo Finance through ‘tidyquant’. The predictor variables are defined as follows by Thomas Reuters:

* ESG: overall ESG score which is calculated using a percentile rank method based on how many companies are worse off than the current company, how many of the companies have the same ESG score, and how many companies have an ESG score (Thomson Reuters, 2017).
* Environment: overall E score. These can be further decomposed into:
  + Resource Use: Measures a company’s performance and ability to reduce the use of non-sustainable materials and finding sustainable alternative materials or solutions by improving supply chain management. This makes up 11% of the weight of calculating the overall ESG score.
  + Emissions: Measures a company’s commitment and effectiveness to reduce GHG emissions in production and operations. This makes up 12% of the weight of calculating the overall ESG score.
  + Innovation: Measures a company’s ability to create new market opportunities through green technologies and processes. This makes up 11% of the weight of calculating the overall ESG score.
* Governance: overall G score. These can be further decomposed into:
  + Management: Measures a company’s commitment and effectiveness towards following best practice governance standards. This makes up 19% of the weight of calculating the overall ESG score.
  + Shareholders: Measures a company’s effectiveness towards equal treatment of shareholders and the use of anti-takeover devices. This makes up 7% of the weight of calculating the overall ESG score.
  + CSR Strategy: Measures a company’s practices to communicate how it integrates economic, social, and environmental considerations in its day-to-day operations. This makes up 4.5% of the weight of calculating the overall ESG score.
* Social: overall S score. These can be further decomposed into:
  + Workforce: Measures a company’s effectiveness towards job satisfaction, a healthy and safe workplace, maintaining DE&I practices, and professional develop opportunities for employees. This makes up 16% of the weight of calculating the overall ESG score.
  + Human Rights: Measures a company’s effectiveness towards respecting the fundamental human rights conventions. This makes up 4.5% of the weight of calculating the overall ESG score.
  + Community: Measures the company’s commitment towards being a good citizen, protecting public health and respecting business ethics. This makes up 8% of the weight of calculating the overall ESG score.
  + Product Responsibility: Measures a company’s capacity to produce good quality products that protect the customer’s health and safety, integrity, and data privacy. This makes up 7% of the weight of calculating the overall ESG score.

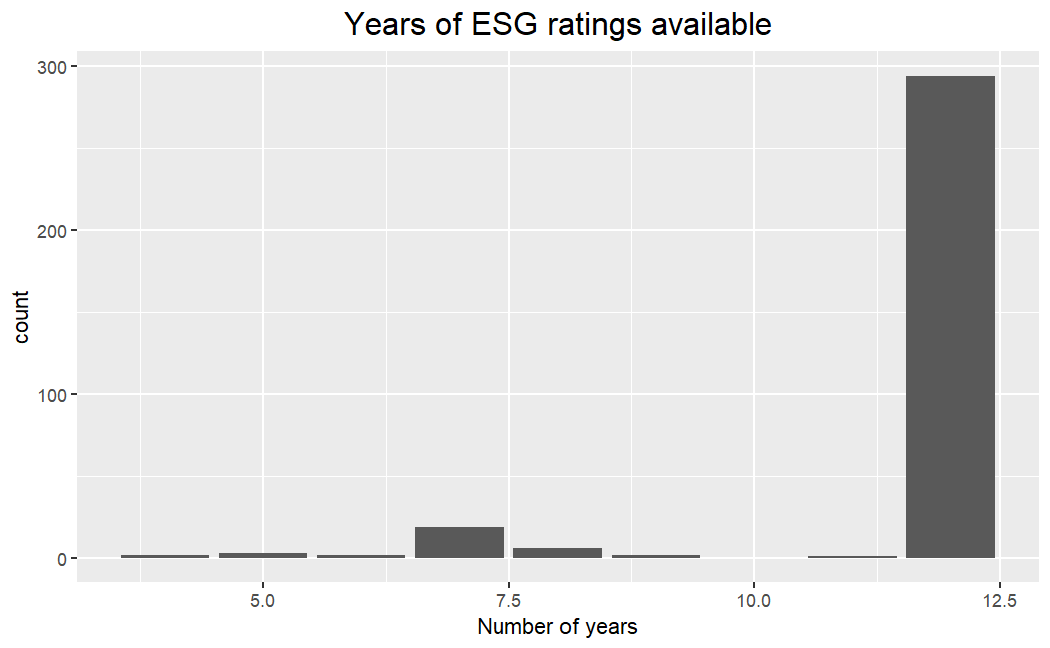
All of these scores are given on a scale between 0 and 1 with 1 being the best score and 0 being the lowest score (Thomson Reuters, 2017). The scores are given a grade based on their range from D- to A+ (Thomson Reuters, 2017).

**Data Preprocessing**

We start from manipulating the smaller dataset (with four useful decomposed scores) and then merge the larger dataset (more decomposed) to it.

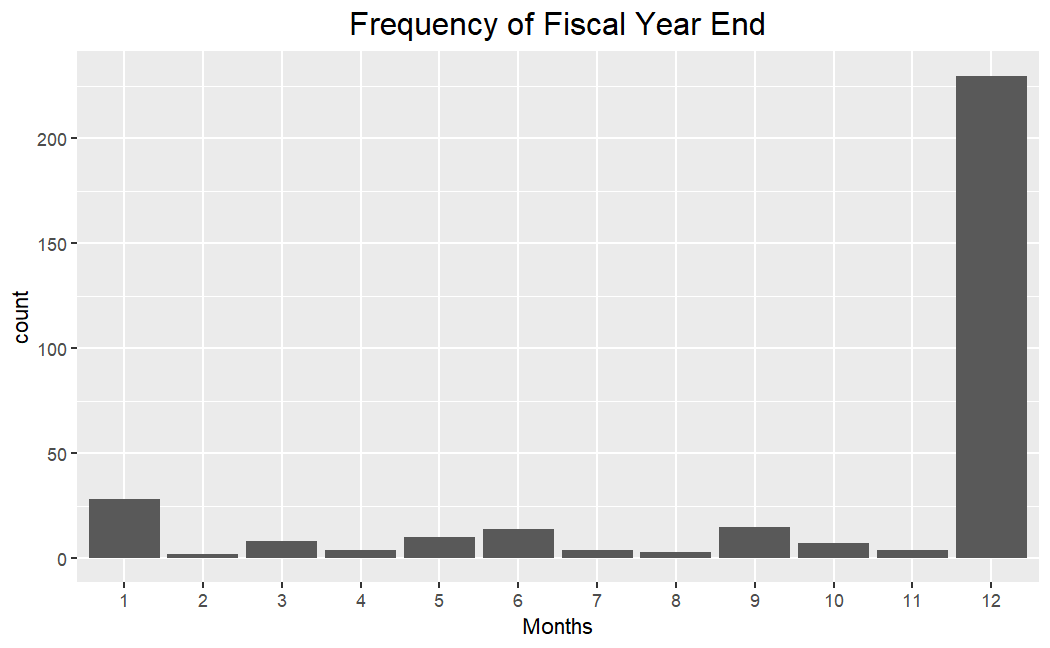
There are some missing values in the smaller set, so the following steps are processed:

1. While using the ticks to get returns from Yahoo Finance, there are several stocks that cannot be found. Thus, ticks without available return data from Yahoo Finance are excluded. There are in total 8 companies removed for the reason.
2. Though majority of companies (identified by ticks) have 12 years of data available (from 2010 to 2021), some of them has less than that (see Figure 1). For data completeness, ticks with less than 12 years of available rating are removed from the data.



*Figure 1: Distribution of years of data available*

1. There are some missing values in environment score. Since they all appear at the end of the timeframe, we filled them using values from previous years.
2. The variable representing fiscal year end date varies by companies and has some missing values. Therefore, for simplicity, we used the complete variable fiscal year instead and set December as fiscal year end while calculating returns for all companies as it is the most common one (See Figure 2).



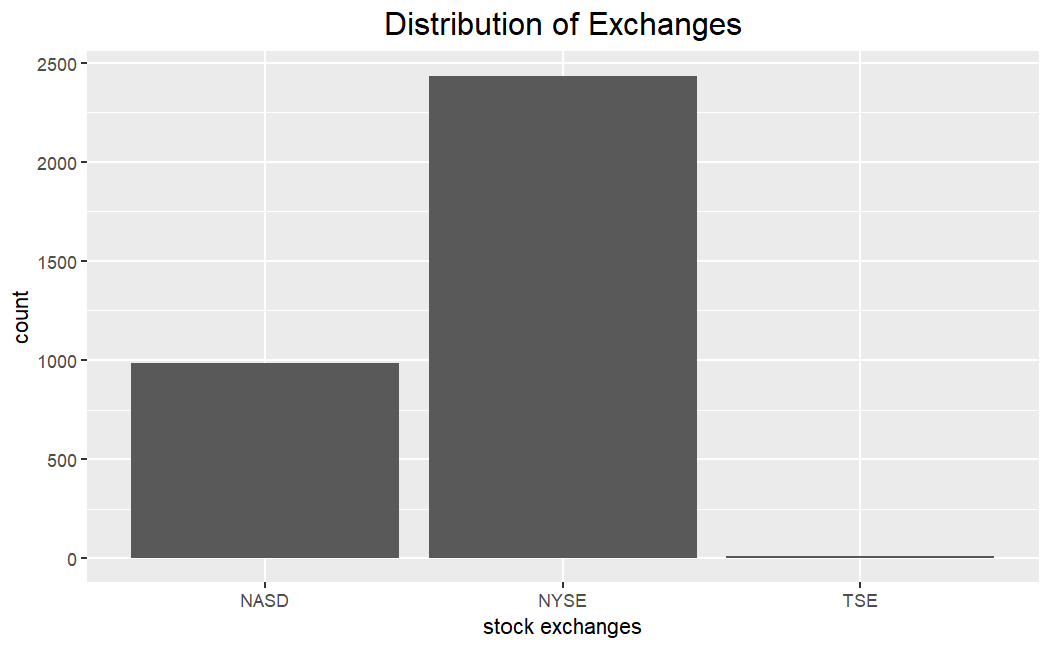
*Figure 2: Distribution of months for fiscal year end*

After data processing, there are 3432 observations from 286 companies. All corporations have 12 years of data available without missing values:



*Figure 3: Description of input and output data*

Most of the companies’ prime exchange is either NASDAQ or New York Stock (See Figure 4). Thus, the result from this project is only representative to stocks traded in U.S. market.



*Figure 4: Distribution of Exchanges of stocks involved*

In addition, some other limitations on the dataset would affect modelling:

1. There is not a huge amount of data available in the input.
2. The data in most levels are left-skewed, which may affect performance of model (Please see Appendix A). To deal with the skewness, we applied uniformization to neural network models and box-cox transformation to other types.
3. There are high correlations between some of the factors, which may lead to issues while fitting models (Appendix B).

Thus, we need to use a supervised machine learning algorithm that does not require a large input and is robust to relatively high correlation. Furthermore, we prefer regression models since it would allow us to compare and visualize the importance of input factors. We would like to implement more than one algorithm to see whether results from different models vary, and we will apply the same set of algorithms for both set of input data (level 1 and level 2) to investigate whether it is worthy to further decompose the ESG score.

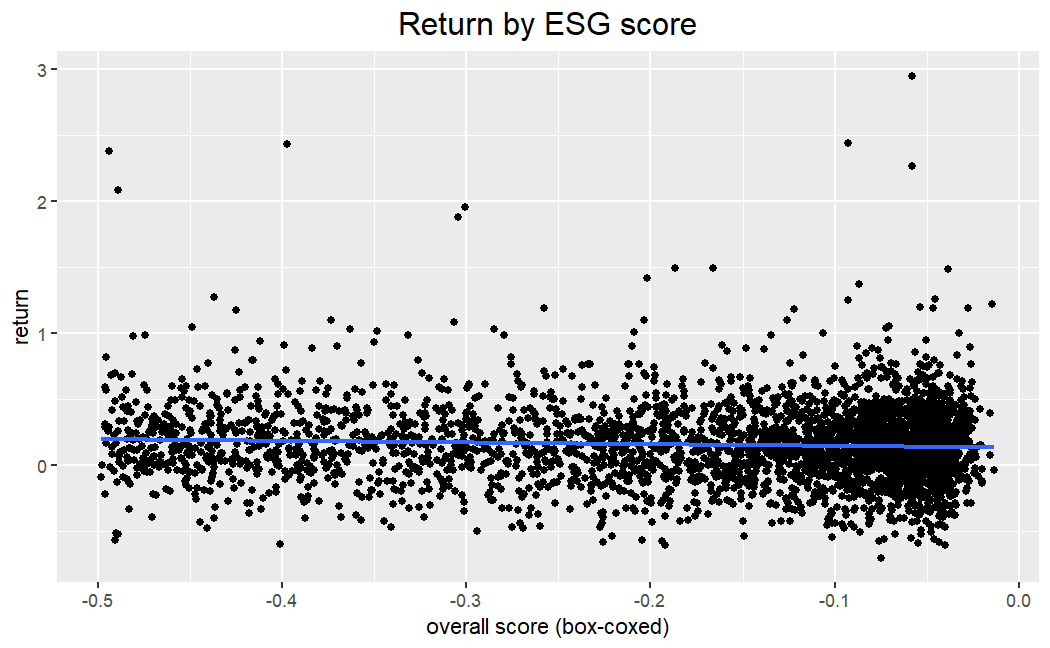
After careful consideration, we determined to use the following machine learning models, listed from simplest to most complex:

* LASSO regression: The advantage of LASSO regression is the interpretability. We can visualize the coefficient and determine importance of vectors by observing which of them get shrank first. However, the model is restricted to investigate linear relationships only.
* Random forest regression: Random Forest regression still allows us to find out the importance of factors, but we cannot extract the exact coefficients. The benefit compared to LASSO is that it can pull out more complex relationships, but since the dataset is relatively small, we need to be aware of overfitting.
* Artificial Neural Networks: To this level of model complexity, we have mostly given up interpretability in compensate for accuracy. We did not find proper way to visualize the effect of coefficients, especially when number of layers increase. However, we still cannot apply too many layers since overfitting is still likely. And due to constrain in data volume, the performance may not be satisfying, but we do want to give it a try.
* Autoencoder: Though it is likely to bring even further concern to overfitting, we would like to incorporate sequential feature into the model. We simulated the ‘Autoencoder asset pricing models’ suggested by Gu, Kelly, and Xiu (2021) and evaluated the performance. We only applied it to the eleven-factor decomposition since four is too small.

We will compare coefficients from LASSO and importance (in percentage) from random forest to discover factor significance and compare mean squared error on test set of all three models for prediction performance. Finally, we will perform back testing to build a portfolio and verify its return to determine whether a portfolio built solely based on decomposition of ESG factor will outperform or not.

**Modeling**

Before decomposing the ESG factors into pillars, we first looked at the linear relationship between yearly stock return and overall ESG score. We applied box-cox transformation to fix the left-skewness and to observe a clearer pattern. The results are demonstrated in figure 5:



*Figure 5: Stock return by box-cox transformed ESG score*

The best line of fit indicate a slight negative linear relationship between return and overall ESG return, which goes against the results from previous researches. However, the performance of linear prediction is very poor. It has an r-squared of only 0.003, with mean squared error of 0.0859 and mean absolute error of 0.214.

We would like to decompose it to see whether it would provide a better prediction.

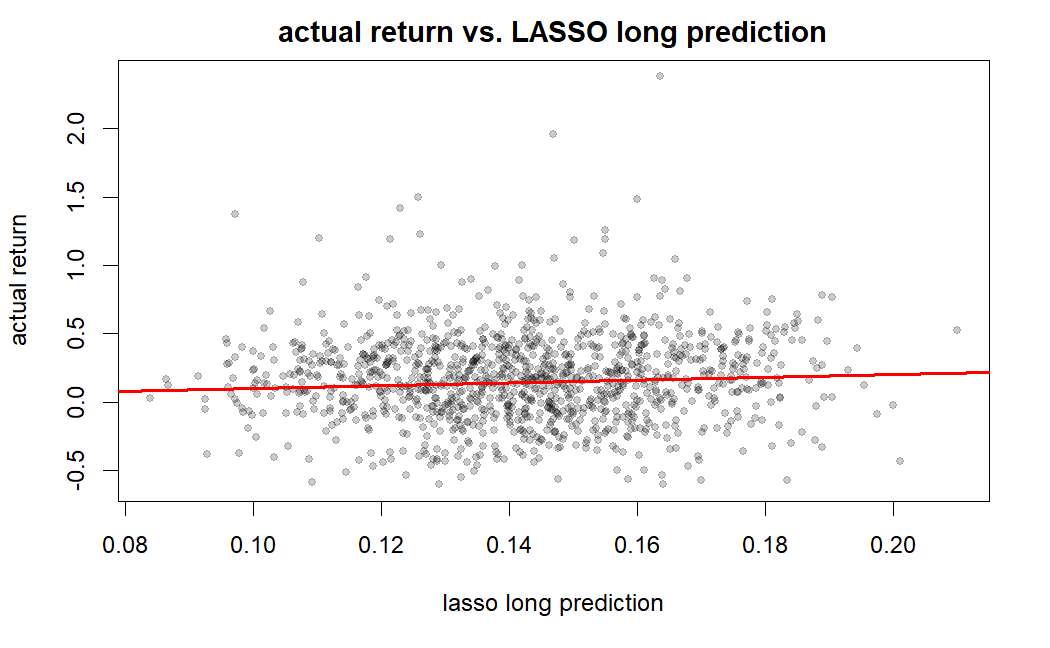
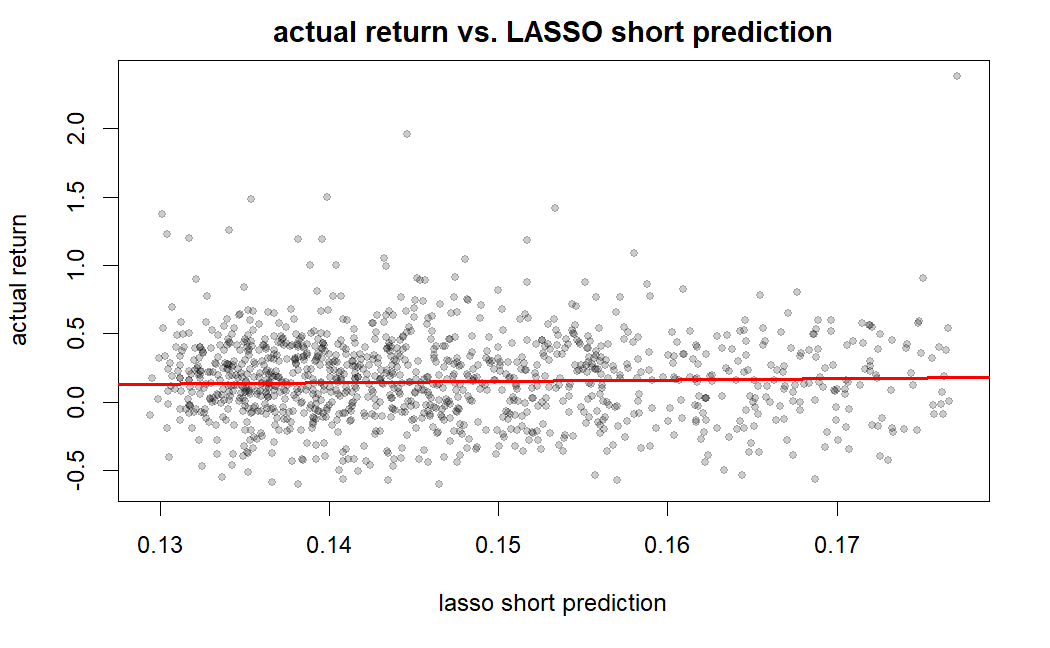
We divide the data set to training sample and testing sample using year 2018 as the separation year, so that we have 66% of the data as training set and the rest as test set. Features are also grouped into ‘short features’ (environment score, social score, etc.) and

‘long features’ (emissions score, innovation score, etc.), and the same algorithm is applied to both set to compare whether further decomposition of the model would improve prediction performance. We are also interested in observing which factor has most affect.

**LASSO Model**

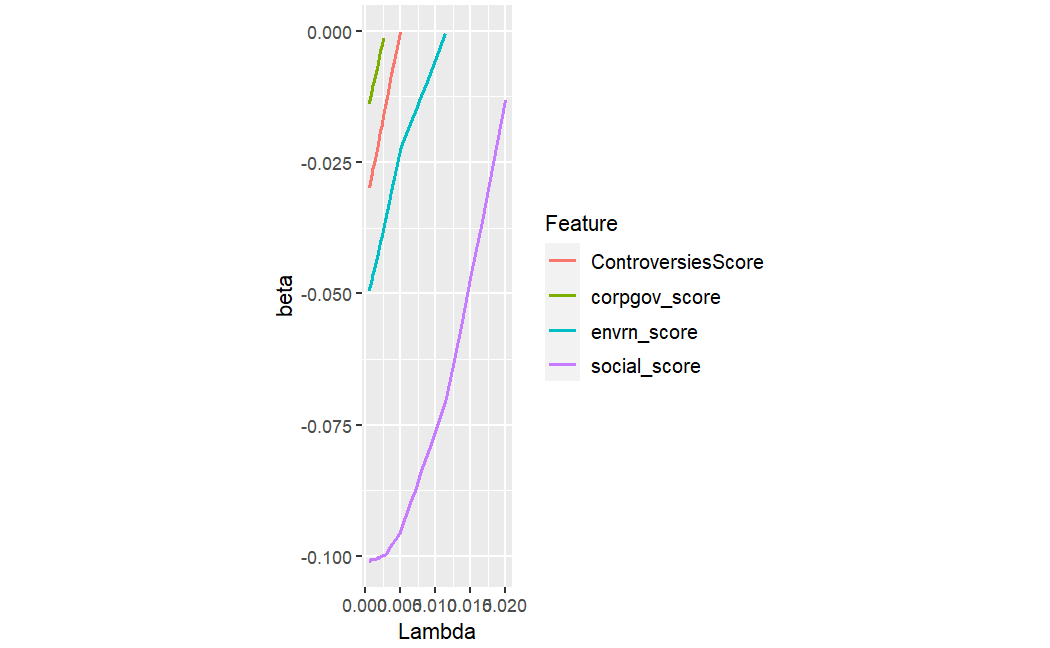
We start from the simplest model, which is LASSO regression. We apply cross validation to the training set to search for optimal lambda.

We first ran LASSO regression on ‘short features’. The mean squared error is 0.1036 while the mean absolute error is 0.2434. For the LASSO regression on ‘long features’, the mean squared error is 0.1032, while the mean absolute error is 0.2429. Further decomposing the factor does not improve the result by much, and the loss is even more than the one from simple linear regression. And we can observe from figure 6a and 6b that the actual return and predictions are not aligned with each other. We would expect the points to lie along the line of y = x, but we do not observe that pattern. What we can also discover is that, while the actual return lies in range from –0.5 to more than 2, the range of prediction is much smaller, so that model cannot predict any abnormal return.



*Figure 6a and 6b: actual return vs. LASSO predictions (short/long feature)*

In terms of factor importance, based on the result from LASSO of short feature, corporate governance score is shrunk first, followed by the controversies score, followed by environment and then social. This demonstrates how the social factor is the most important, followed by environment and then controversies and governance. All four factors have negative beta, so a higher score will result in lower expected return. For long feature model, resource use score seems to be the most important feature, followed by the human rights score, then the product responsibility score. The least important feature seems to be the emission score, shareholder score, innovation score and workforce score. Please notice that community score is least important and shrank so fast so that it is not in the graph. A difference from short features is that now we have management score and human rights score that has positive coefficients.

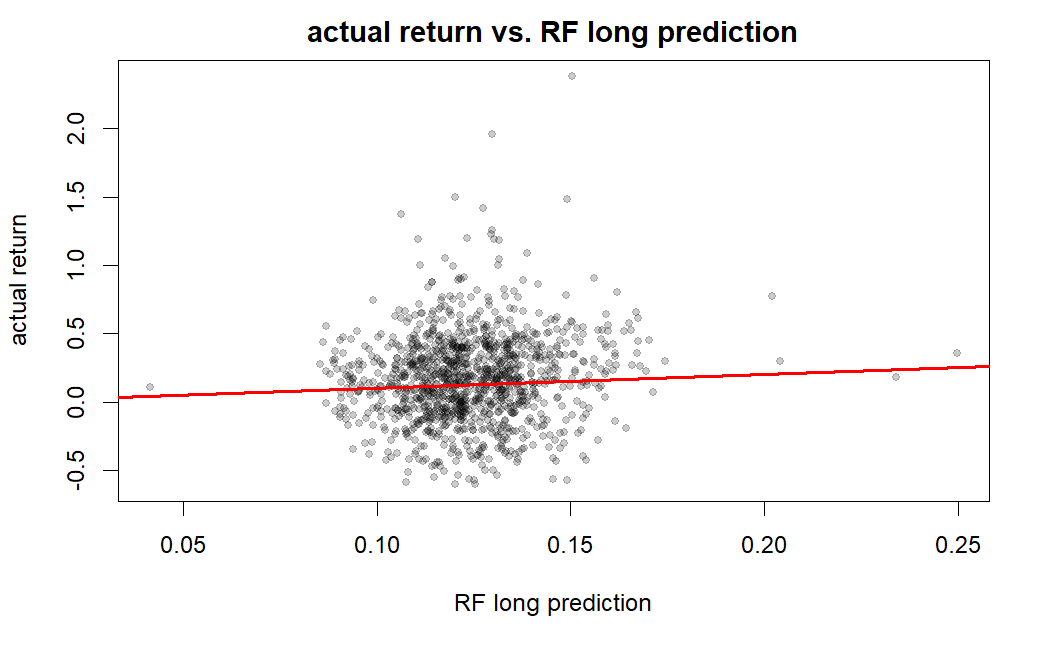
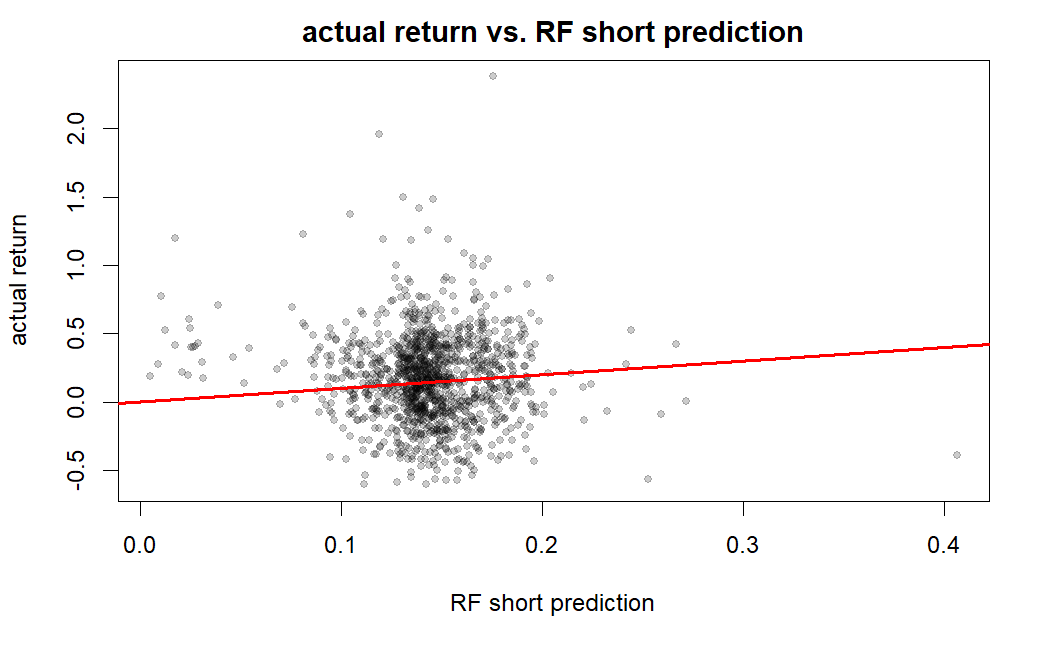


*Figure 7a and 7b: coefficient of factors over lambda (short/long feature)*

**Random Forest Regression Model**

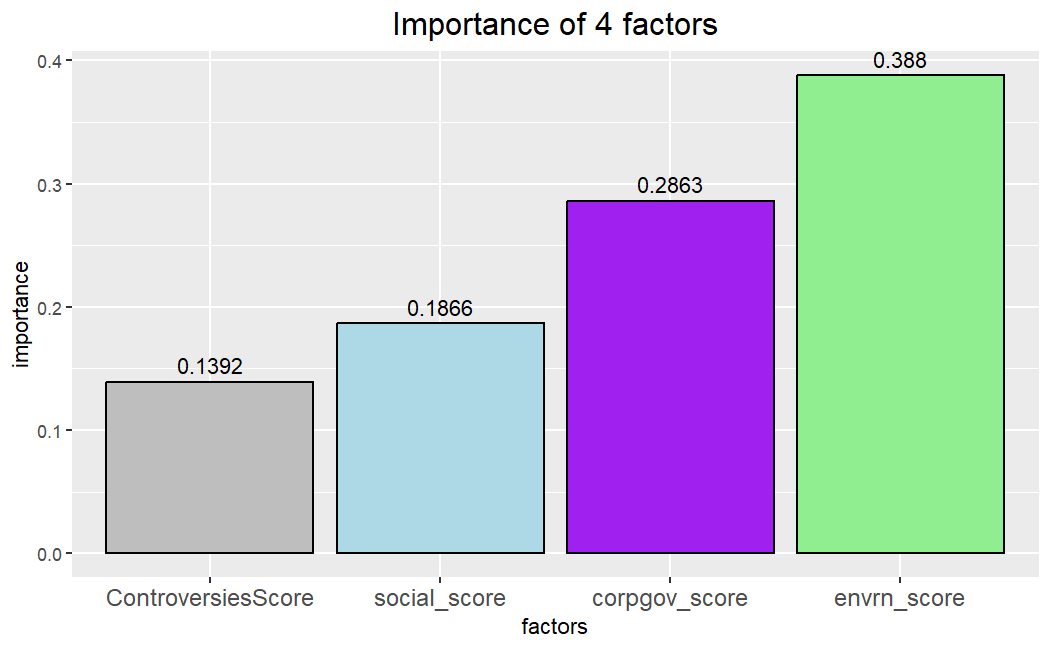
While building random forest regression models, there is problem of instability in results. The importance of factors can vary by much from runs to runs. To deal with the problem, while taking the risk of overfitting, we increase the number of trees to 50.

For random forest with short feature, the mean squared error is 0.1056 and the mean absolute error is 0.2462; for the one with long feature, the mean squared error is 0.1038 and 0.2432. Again, the performance is very poor, and further decomposition does not benefit. But one difference from LASSO is that, according to figure 8a and 8b, prediction of random forest has a wider range. That said, while most of the outputs concentrate in one area, it is possible for the output to be abnormal.

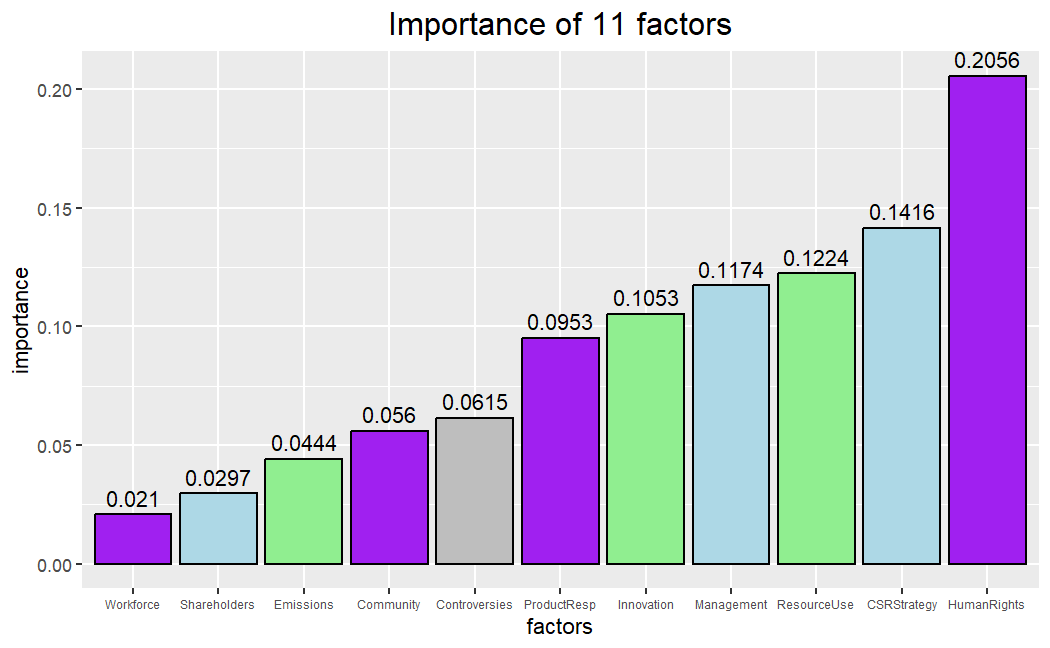


*Figure 8a and 8b: actual return vs. Random Forest predictions (short/long feature)*

In terms of significance of factor, we compare the importance of factors in percentage. The results are demonstrated in the table1 (short feature) and table 2 (long feature) below:



*Figure 9. Importance of features from RF short model*



*Figure 10. Importance of features from RF long model*

The above results are just one representative possibility from various random forest models. As mentioned before, the importance percentage is unstable. However, the above conclusion is certain based on multiple runs:

* The rank for short feature model is not changing. Environment score is the most significant, followed by corporate governance score, social score, and finally the least important one controversies score. However, the gaps in between differs. For example, the difference between environment score and corporate governance score ranges from almost to 0 to more than 10%.
* The rank for long feature model fluctuates more, but in general:
  + CSR strategy score and human rights score are the most important,
  + Emission scores, workforce score, and shareholders score are the least important,
  + Innovation score seems to fluctuate a lot

To obtain more convincing results, especially for the one with long feature, we would need to run the random forest models for hundreds or even thousands of times and record the result to compute mean and variance. Due to time constraints, we would not perform it here, but that can be a potential next step.

**Comparison in factor significance**

Since factor importance of more complicated models are not availble, we can perform comparison at this point. Due to inconsistency in ranking in long feature random forest model, we grouped the 11 factors into different level of importance. Here is the result:

|  |  |  |
| --- | --- | --- |
|  | **LASSO** | **Random forest** |
| Most important | Social score | Environmental score |
| Important | Environment score | Corporate governance score |
| Relatively important | Corporate governance score | Social score |
| Least important | Controversies score | Controversies score |

*Table 1: Importance ranking for short feature models*

|  |  |  |
| --- | --- | --- |
|  | **LASSO** | **Random forest** |
| Most important | Resource Use score  Human Rights score  Product Responsibility score | Human Rights score  CSR Strategy score |
| Relatively important | CSR Strategy score  Management score  Controversies score | Resource Use score  Management score  Innovation score  Product responsibility score |
| Least important | Workforce score  Innovation score  Shareholders score  Emissions score  Community score | Controversies score  Community score  Emissions score  Shareholders score  Workforce score |

*Table 2: Importance ranking for long feature models*

Through comparison above, we can conclude that:

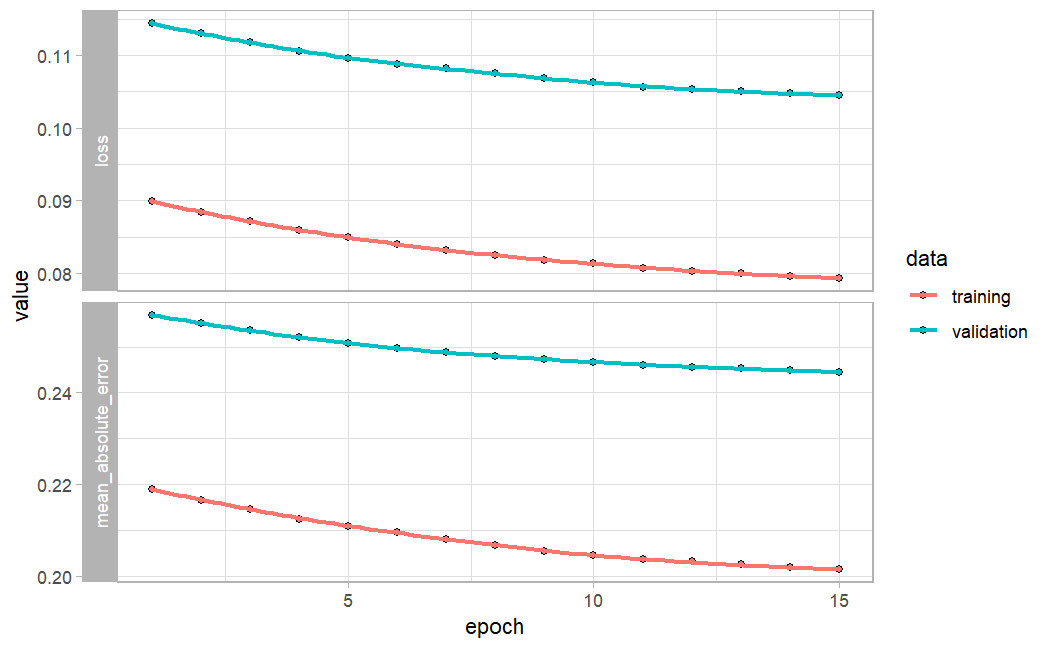
* In four-factor models, both algorithm agrees that environment score is important and controversies is the least significant, but they do not agree with the importance of corporate governance score.
* In eleven-factor models:
  + Both models agree that: human rights score is significant, managements score is relatively important, workforce score, shareholders score, emissions score, and community score does not matter much.
  + But there are differences such as:
    - CSR strategy score and innovation score are more important in random forest model;
    - Resource Use score, Product responsibility score, Controversies score are more important in LASSO regression model
* Through comparing movements of decomposed factors in E, S, G pillars, it is likely that:
  + Social factor becomes less important in random forest due to fall in importance of product responsibility score;
  + Environment becomes more important because of rise in Innovation score, though resource use score turns to be less significant;
  + Corporate governance score becomes more important because of rise in importance of CSR strategy score

If we are interested in further developing the changes from model to model, these decomposed factors are important.

**Artificial Neural Network (ANN) Model**

We uniformalised the input data before fitting them to the model. And for both datasets, to avoid overfitting, we tried to make the model simple to avoid overfitting. We built only two hidden layers, with activation ‘relu’ and ‘tanh’ for each of them and set the output layer to be linear. Since the short feature model only has four predictors, we set the units to be 1, while increasing the units to 8 and 4 for long feature as it is capable of more estimations.

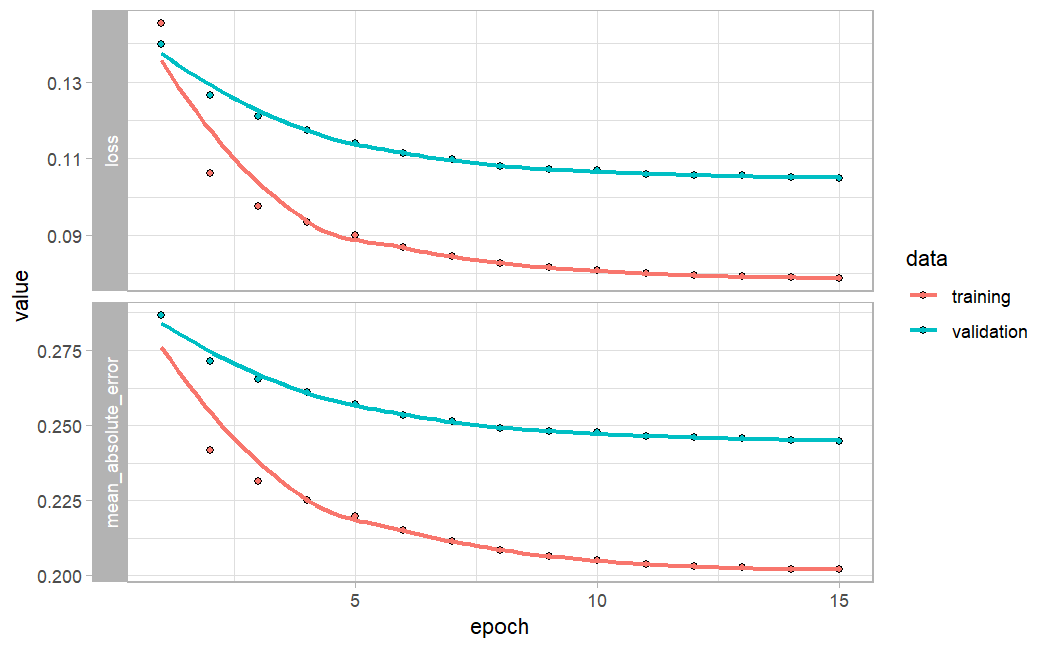
We first tried to fit an ANN model with only the short features. The loss by epoch is as below:



*Figure 11: Loss of short feature ANN Model*

Even though we have simplified the structure of the neural network, there is still clear pattern of overfitting, as validation set error is consistently greater than the one of training set. The resulting mean squared error is 0.1045, while the mean absolute error is 0.2454.

For the long neural network model, we found the same issue:



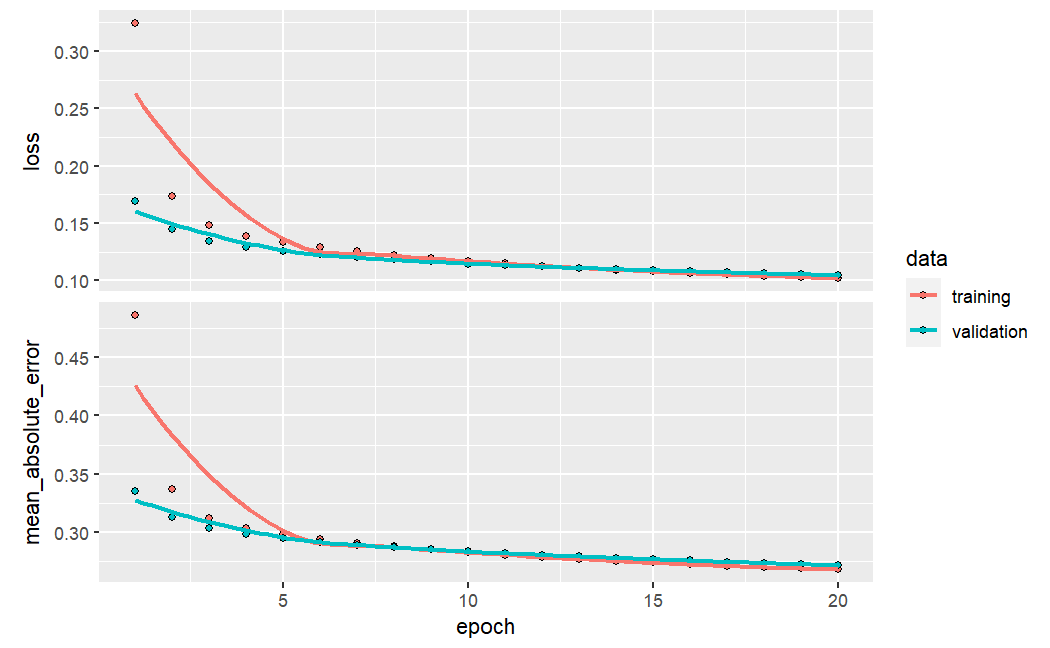
*Figure 12: Loss of long feature ANN Model*

For the long neural network model, we found the same issue:

Again, there is pattern of overfitting. The resulting mean squared error is 0.1049, while the mean absolute error is 0.2456. The performance does not change by much as well.

**Autoencoder**

We also incorporated the autoencoder model built by Gu, Kelly, and Xiu in their article in 2021 to take sequential factor into account, but I only applied it to the long feature model. Here is the performance:



*Figure 13: loss of autoencoder model by epochs*

In terms of performance, autoencoder generate a mean squared error of 0.1047 and mean absolute error of 0.2718, which is similar to other models. Meanwhile, though result varies, it has the potential of non-overfitting.

**Ensemble**

We did a simple ensemble on the LASSO, random forest, and artificial neural networks models. It turned out that the correlation between error is extremely high, so all three models actually reveal the same information. It turned out that LASSO dominates the final portfolio, so we would use it for back testing.

|  |  |  |  |
| --- | --- | --- | --- |
| Errors | LASSO | Random Forest | Neural Network |
| LASSO | 1.0000 | 0.9992 | 0.9932 |
| Random Forest | 0.9992 | 1.0000 | 0.9960 |
| Neural Network | 0.9932 | 0.9960 | 1.0000 |

*Table 3: correlation of model error for ‘short features’*

|  |  |  |  |
| --- | --- | --- | --- |
| Errors | LASSO | Random Forest | Neural Network |
| LASSO | 1.0000 | 0.9972 | 0.9864 |
| Random Forest | 0.9972 | 1.0000 | 0.9856 |
| Neural Network | 0.9864 | 0.9856 | 1.0000 |

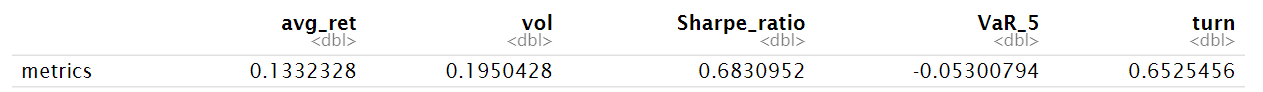
*Table 4: correlation of model error for 'long features’*

|  |  |  |  |
| --- | --- | --- | --- |
|  | LASSO | Random Forest | Neural Network |
| Short features (4-factor) | 1.1586 | 0.2282 | -0.3868 |
| Long features (11-factor) | 0.8873 | 0.0664 | 0.0463 |

*Table 5: model weights in ensemble model*

**Back testing**

For back testing, we applied long feature LASSO regression as a prediction tool to build a portfolio. The average return of the portfolio is surprisingly quite positive and Sharpe ratio is quite high, while the variance is not large. The LASSO based portfolio seems to be a good one, which goes against our expectation since the model’s performance is very poor.



*Table 6: performance of portfolio built on LASSO regression*

**Conclusion**

This report sought out to answer the following three questions:

1. Which of the ESG factors is the most important? (i.e. has the largest effect on the return?)
2. How far do we need to decompose ESG factors? Would further decomposing each individual ESG factor improve the performance of the models?
3. What is the most appropriate model to use to evaluate these questions?

First thing to point out, we have to be frank that the performance of all models built are very poor. The mean squared errors are almost equally high and are even worse than a simple linear regression model. Thus, we would try to answer questions above, but the quality is very questionable.

From our analysis, we discovered that out of E, S, and G, the factor that had the largest effect on the returns depended on the model itself. For our RF model, we noticed that governance was the most important which is consistent with the results by Giese, et. al and (2021). For the LASSO model, we discovered that the social factor was the most important. That said, there are disagreements between the models in terms of factor importance. Meanwhile, what both models suggest is that environmental scores are relatively significant, while controversies score, which is an extra component added by Thomson Reuters in their rating, is the least important.

Based on comparing the models run on the short feature compared to the long features, we noticed that decomposition of ESG factors do not really benefit model performance. This is much against common sense, and we doubt that it is due to constraint in either data size itself or ESG’s incapability to capture changes in asset returns on its own.

Lastly, we recognize that increasing model complexity through neural network does not really improve the prediction performance. Due to the small size of dataset, more complicated model structures will raise the issue of overfitting, hence is not preferrable. This helps us conclude that it is not necessary to further decompose the ESG score further past decomposing into each of the E, S, and G factors in current setting. However, if we have to evaluate the models strictly empirically, we notice that our LASSO model on the short features provides the lowest MSE score out of all the models used. So based on the MSE value, the LASSO model is the best model. In this case, simple is the best.

Some limitations of our report include the following:

* ESG data is reported on an annual basis as opposed to common predictors such as Fama’s five factors which is reported daily. Thus, there is a limitation in the amount of data that was available to analyze if solely focus on ESG factors,
* Difficulty in obtaining the decomposed dataset of each of the E, S, and G factors, as well as the breakdown one level further, which further limit the amount of data available
* High complexity of machine learning model compared to the data available, leading to overfitting and decreased interpretability
* Existence of outliers that may affect the losses
* Left skewness of the data likely resulted in the inconsistency of results between the different models
* We did not fully incorporate the sequential feature of the data other than for the autoencoder model

Our next steps will be to:

* Discover variation of factor importance in random forest model
* Investigate the effect of outlier
* Fully incorporate the sequential feature of the data for all the models through other models including Long Short-Term Memory (LSTM) Neural Network
* Incorporate ESG factors with other factors to see its significance and whether it boosts the performance of other models.

**Bibliography**

Broadstock, D. C., Chan, K., Cheng, L. T. W., & Wang, X. (2021). The role of ESG

performance during times of financial crisis: Evidence from COVID-19 in

China. *Finance research letters*, *38*, 101716. https://doi.org/10.1016/j.frl.2020.101716

Giese, G., Nagy, Z., & Lee, L. (2021). Deconstructing ESG ratings performance: Risk and return for E, S, and G by time horizon, sector, and weighting. *Journal of Portfolio*  *Management*, 47(3), 94-111. https://doi.org/10.3905/jpm.2020.1.198

Gu, S., Kelly, B., & Xiu, D. (2021). Autoencoder asset pricing models. *Journal of Econometrics*, Volume 222, 429 – 250. https://doi.org/10.1016/j.jeconom.2020.07.009.

Lioui, A. & Tarelli, A. (2022). Chasing the ESG Factor. *Journal of Banking & Finance,* Volume 139. https://doi.org/10.1016/j.jbankfin.2022.106498

Melas, D., Nagy, Z., & Kulkarni, P. (2016). Factor investing and ESG integration. *MSCI*, 389–413. <https://doi.org/10.1016/b978-1-78548-201-4.50015-5>

Thomson Reuters. (2017). Thomson Reuters ESG Scores. *Thomson Reuters*. Retrieved from

<https://www.esade.edu/itemsweb/biblioteca/bbdd/inbbdd/archivos/Thomson_Reuters_ES>

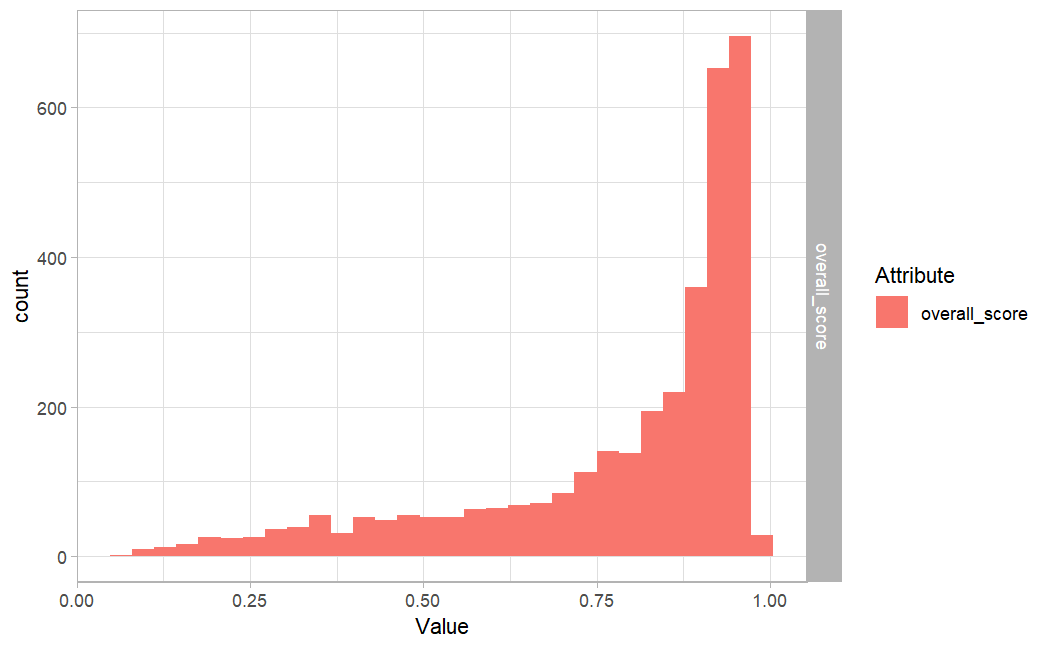
G\_Scores.pdf

Yahoo Finance. Stock Price for Components of S&P500, 2010-2022. Data retrieved from Yahoo

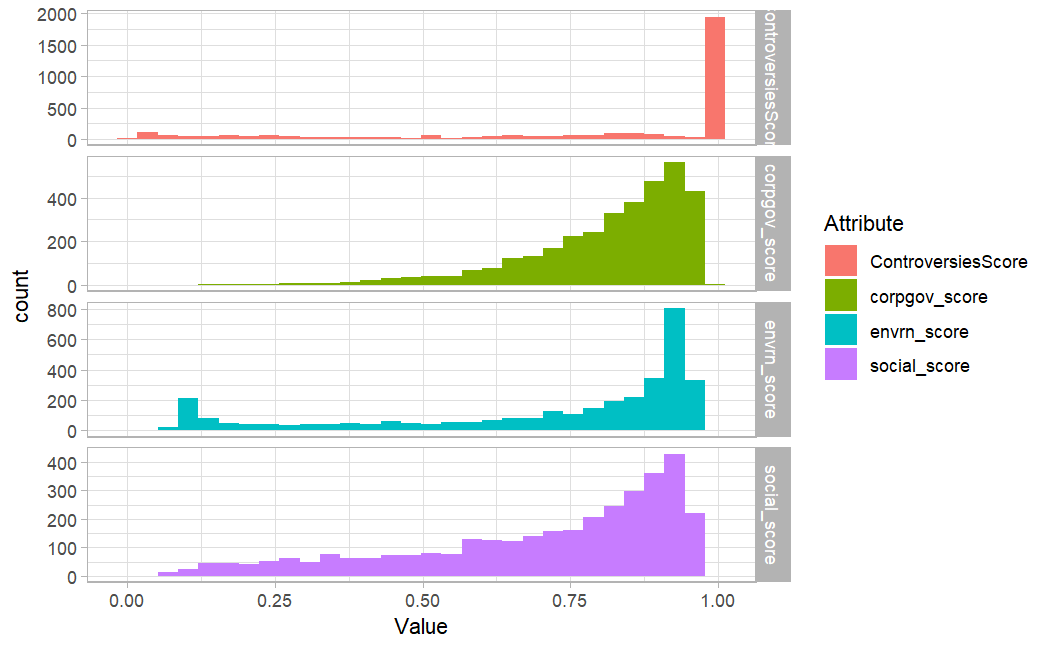
Finance by its R package

**Appendix**

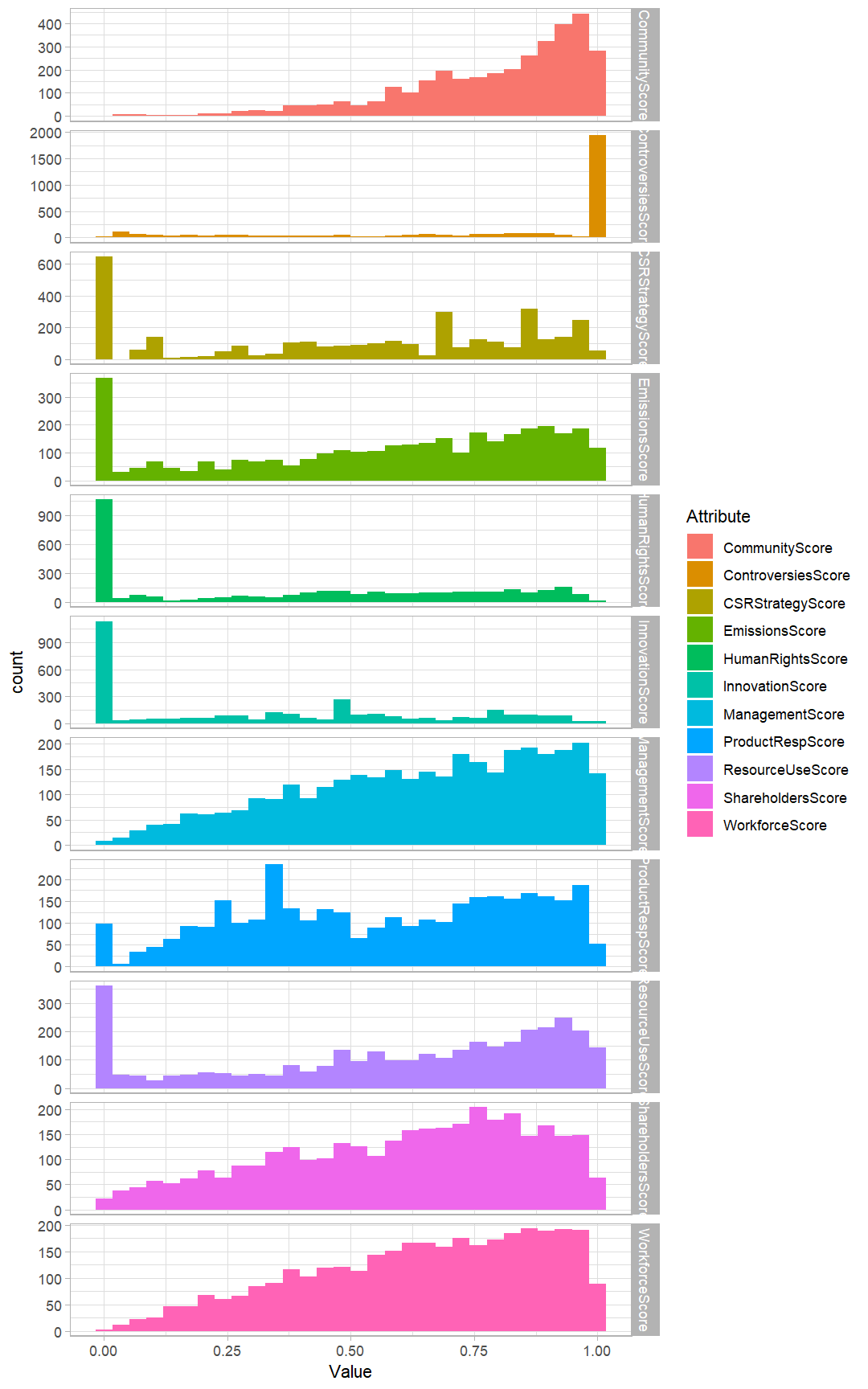
**Appendix A: Distribution of data**



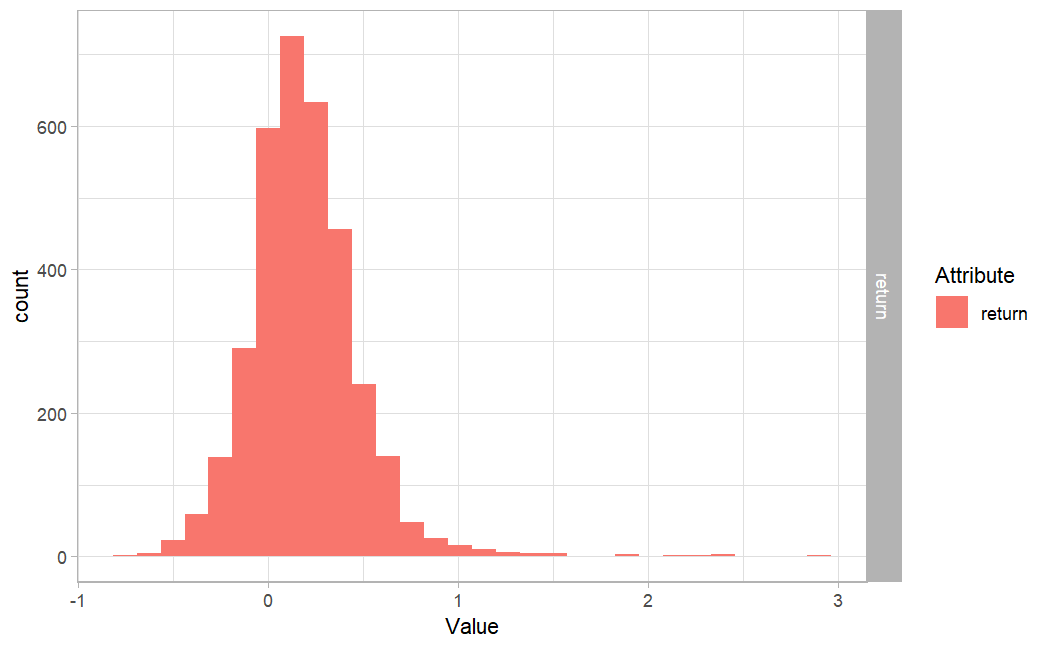
a. ESG overall score



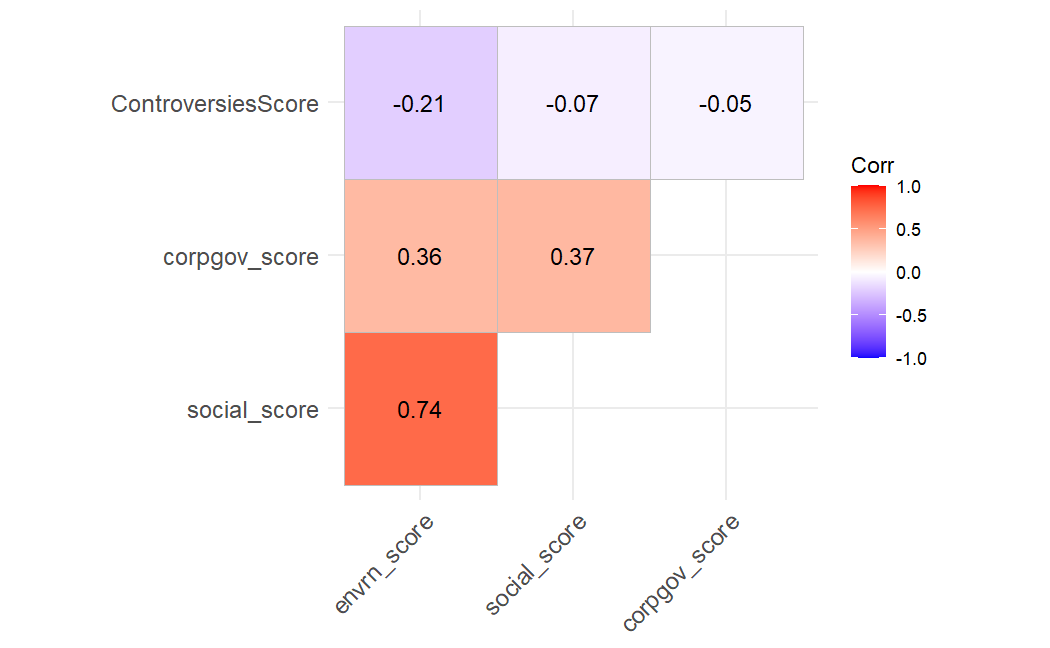
b. ESG scores decomposed into 4 factors



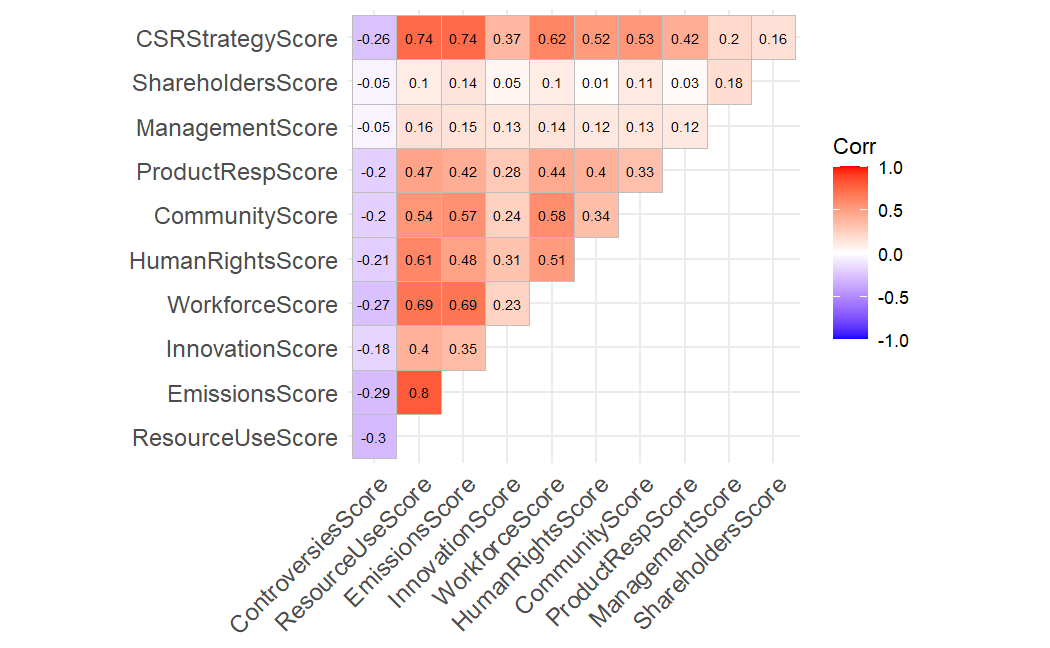
c. ESG decomposed into 11 factors

  
d. Return

**Appendix B: Correlation between input factors**



a. decomposed into 4 factors



b. decomposed into 11 factors