A Greener Future Beyond Profits: Sustainability as a Driver of Market

Jingruo Sun, Department of Management Science and Engineering, jingruo@stanford.edu

Introduction

Top 50 U.S. companies with the best sustainability scores beat the market by 8% annual returns during the economic downturns of COVID-19, which arises strong interests in sustainable investing among 83% investors[2]. ESG, standing for Environment, Social, and Governance, is a comprehensive measure of long-term sustainability. Our project explores the relationship between ESG features and stock returns via various machine-learning models, and finally derives a novel investment strategy with a 20.6% reduction of mean-squared error compared with random choices.

Data & Features

Our dataset contains 434 companies in S&P 500 with 440 ESG variables, 21 stock characters, and monthly stock returns from Jan 2004 to Dec 2020, totally 88536 examples.

- ESG features: take average of three commonly used sources: Refinitiv, MSCI, and S&P Global
- Stock characters & returns: obtained from the Center for Research in Security Prices (CRSP)
- Company: obtained from *S&P 500 Companies with Financial Information* for tickers and sectors

We use company's market capitalization for **data normalization** and robust scaling method for **data standardization** to ensure a consistent scale for all inputs. For **missing data**, we filter out examples with more than half of missing ESG information. We randomly select 30 ESG features as primary variables with 10 for each category. We apply a 80% : 20% split for training and test datasets.

Models & Performance

I. Linear Regression

Define $X \in \mathbb{R}^{70828\times 30}$ containing 30 ESG features and $y \in R^{70828}$ as stock returns. Perform the Least Mean Square algorithm, we attain a linear regression with $\theta \in \mathbb{R}^{30}$ satisfying the normal equation.

1. **Log Transformation:** solve the problem for many ESG variables of values crowded at 0
2. **Weighted Least Square:** assign higher weights for recent data and lower weights for early data

Our objective function is:

$$\sum_{i=1}^{n} w_i \cdot (y_i - \hat{y}_i)^2, \text{ where } w_i = \frac{e^{-0.1\frac{t_i}{\max t_i}}}{\sum_{i=1}^{n} e^{-0.1\frac{t_i}{\max t_i}}} \tag{1}$$

where w_i, t_i are weight and date, $y_i, \hat{y}_i$ are actual and predicted result for the i-th example.
We make a plot for stock return and operational economic efficiency, which is the ESG variable with the highest predictive power, in Figure 1 to compare the performances.

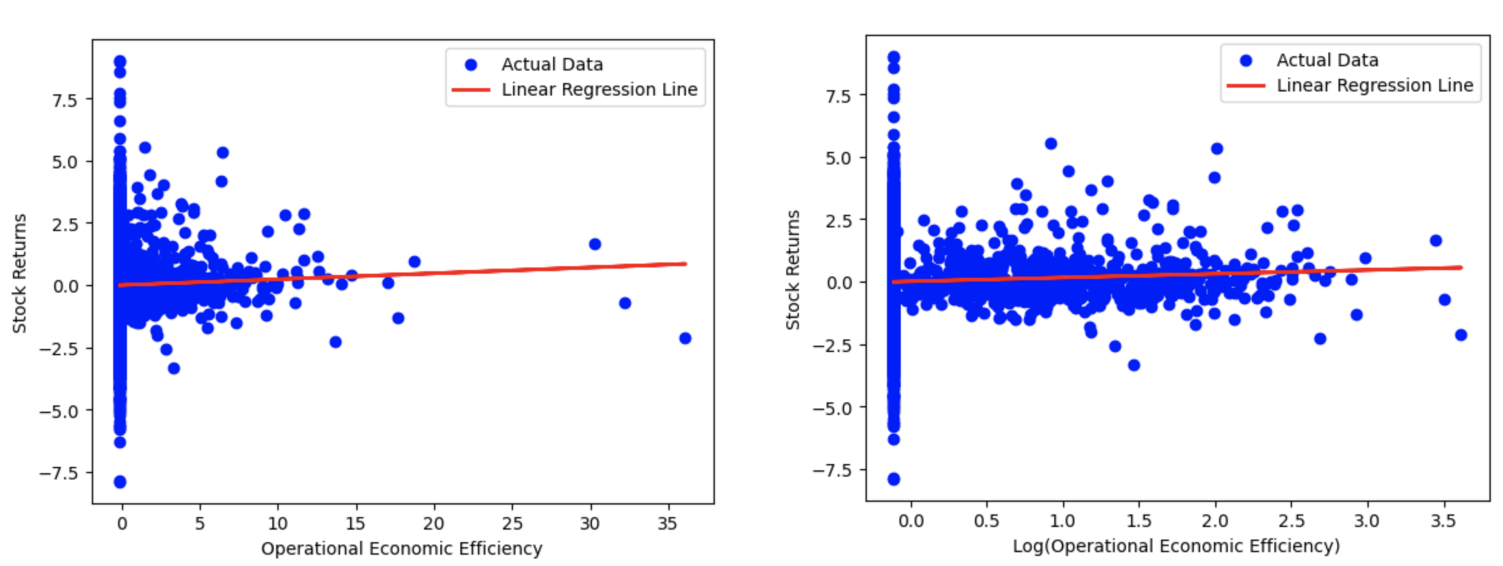

Figure 1. Linear Regression with Log Transformation

As what we expect, it shows that log transformation reduces the skewness of input variables and thus increase the accuracy of linear regression.

II. Extreme Gradient Boosting

We use the same 30 variables from linear regression and let the number of boosting rounds to be 100 for the primary Extreme Gradient Boosting (XGBoost) model.

1. **Lasso Regularization:** penalize complex structures and mitigate the risk of overfitting with

$$\text{Objective function } J(\theta) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \gamma \sum_{k=1}^{K}\sum_{t=1}^{T}|w_{j,k}| \tag{2}$$

   where $w_{j,k}$ denotes the weight of the j-th leaf in the k-th tree, γ denotes the regularization strength.
2. **Hyperparameter Selection:** use technique of k-fold cross-validation and early stopping and attain optimal number of boosting iterations of 5 and regularization strength of 0.5
3. **Variable Selection:** drop off 5 variables with the lowest importance and randomly add 5 more features from "Environment" sector, which has the highest weight among three categories. We plot features' importance in Figure 2 and we observe a more evenly-distributed weights, indicating that we find more related variables for the model.

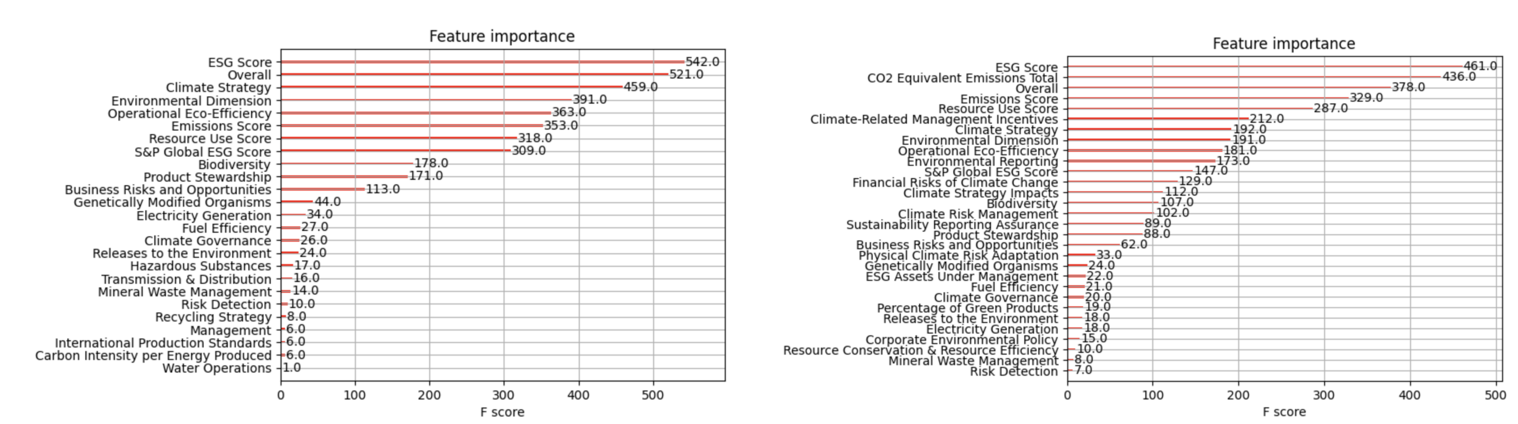

Figure 2. Features' Importance with Variable Selection

III. Deep Neural Network

We choose the model of feed-forward neural network and use the 30 variables selected by XGBoost. We apply "GridSearch" method to choose optimal hyperparameters for the neural network from:

(1)**learning rate:** test[0.1, 0.01, 0.001, 0.0001] (2)**number of layers:** test[1, 2, ..., 9, 10]
(3)**number of nodes:** test[32, 64] (4)**activation function:** test[relu, sigmoid]

where the optimal combination is (0.001, 32, 5, relu).

Results & Summary

We use the mean squared error to evaluate our model performance. The random result is generated from the uniform distribution between min and max of stock returns. Table 1 summarizes the best results of each model, with 70828 training examples and 17707 test examples for each.

Model	Training mean-squared error	Test mean-squared error
Random Choice	9.3706	1.2248
Linear Regression	1.0052	0.9743
XGBoost	**0.9988**	**0.9720**
Neural Network	1.0029	0.9731

Table 1. Summary of Prediction Results

It shows that XGBoost model has the best performance. We suppose it is because its unique structure of ensemble of trees provides its with a fair view of a broad range of features, and our improvements allow the model to have a compromise between the complexity and generalization. We then perform the error analysis based on XGBoost.

Error Analysis

1. We find that more than 75% examples have residuals < 1. Thus, to deeper understand those extreme examples, we select the top 100 stocks with the largest absolute residuals and plot their residuals and the overall ESG scores in Figure 3. We observe that for most examples, the stock with lower ESG scores tends to be under-predicted, and the stock with higher ESG scores tends to be over-predicted.

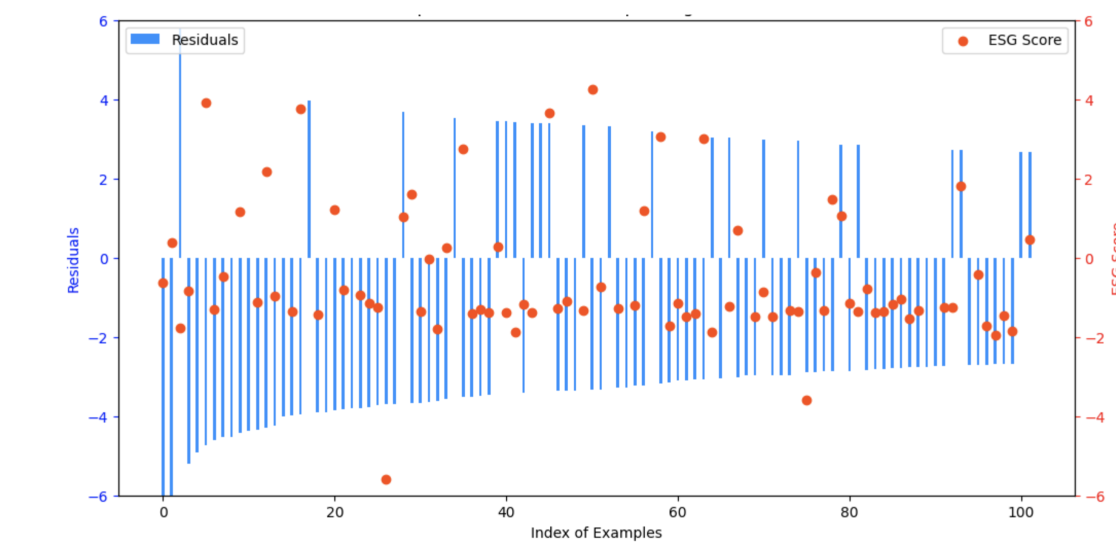

Figure 3. Highest 100 residuals & corresponding ESG scores

2. We then check the company categories of the top 100 and bottom 100 stocks sorted by absolute residuals in Figure 4. It shows that Tech companies are more likely to have higher prediction errors. We infer the reason is that ESG features may not be the most significant elements to tech companies compared with factors like innovative algorithms. Thus, the overemphasis of ESG variables is reflected more obviously in those "ESG non-sensitive" companies.

Top 100 Company Sectors	
Consumer Discretionary	22
Information Technology	22
Health Care	13
Communication Services	10
Energy	10
Materials	8
Industrials	7
Financials	6
Utilities	2
Real Estate	2

Bottom 100 Company Sectors	
Financials	15
Industrials	15
Consumer Discretionary	12
Information Technology	12
Utilities	11
Health Care	11
Consumer Staples	8
Materials	6
Energy	5
Communication Services	4
Real Estate	1

Figure 4. Different performance on company categories

3. We combine ESG features with 21 stock characters, which are proved as good predictors[1], to be new input variables. Table 2 illustrates the solid predictive power of ESG features since the combined strategy shows a better performance than "ESG" and "stock" strategies solely.

Strategy	Mean Squared Error	Mean Absolute Error
Random Choice	1.2248	0.9452
ESG features	0.9720	0.6853
Stock characters	0.9730	0.6858
ESG + Stock	**0.9694**	**0.6843**

Table 2. Strategy Comparison

Next Steps & Reference

Recognizing our results vary from company sectors, our next step is to construct different ESG-driven strategies for different categories of companies. Besides, we will include both classification and regression models to select the optimal one for specific companies. We will also consider variables from other aspects like research funding of the company to establish more comprehensive investment strategies.

[1] Green, j., j.r. hand, and x. f. zhang, The characteristics that provide independent information about average U.S. monthly stock returns, 2017.
[2] Morgan stanley's institute of sustainable investing 2022, https://www.morganstanley.com/content/dam/msdotcom/en/assets/pdfs/morgan_stanley_2022_esg_report.pdf.