Springboard—DSC Program Capstone Project 2

The Financial Performance of the ESG Funds

Proposal by Tingyin Xiao

April 2023

Introduction

In recent years, sustainability investments have raised attention in the financial sector. Although gaining traction, doubts loom over the investors whether ESG (environmental, social, and governance) funds are and will be losing money, especially in the middle of the market's struggles. Therefore, confirming the hypothesis that these assets perform better than other funds can ease the doubts and motivate ESG investments. And suppose the result differs from what we expected, we can provide more insight into how climate risk and governance issues can affect financial performance, what sustainability assets are doing well, and predictions on their prospects of gains in the future for the stakeholders' reference.

Stakeholders

The intended stakeholders are the investors, who care about the ESG topics or not, and the financial institutions that help with their investments. They will be interested in knowing the performance of these funds in terms of returns compared to alternatives to make profitable decisions.

Data

The datasets for this project are available from the "Fossil Free Funds" website: https://fossilfreefunds.org/how-it-works

They are financial data on equities and mutual funds sourced from Morningstar and have a monthly frequency covering April 2020 to the present (April 2023). The available data include 3000 funds with over 8000 share classes and only screen mutual funds with at least 50% invested in stock investments.

The datasets have rich information, containing more than ten target variables and more than 100 feature variables. Some of the feature variables are time- dependent. The data types of the feature variables include binary, categorical, and numeric.

Methods

I currently envision the use of regression algorithms to predict long-term returns as a function of features including fund information, ESG grades, and detailed ESG metrics for the funds in the

datasets. However, I will also consider time series algorithms and unsupervised approaches such as clustering.

As I will use multiple regression algorithms, their performances can be evaluated with respect to the following:

- R-squared
- MAE (Mean Absolute Error)
- MAPE (Mean Absolute Percent Error)
- Distribution of residuals with the test set
- Actual-vs-Predicted scatter plots

I will also consider Interpretability approaches to study how the variation of chosen features affects the variation of the chosen targets.

Criteria for success

The criteria for success will be generating a ranking of funds based on their predicted performance and generating analyses that can explain what factors contribute to increasing/decreasing the performance of various funds.

Scope of solution space

We can use the predicted returns of the assets to calculate the average return from the ESG vs. other types of funds and provide the ESG funds that are predicted to have the most returns. The intended clients can use these results to guide their decision-making strategies in their future investments.

Seeing that the returns from the climate-friendly funds are even higher than the alternatives, investors who care about climate change and ESG topics will know that investing in them is a noble cause and a profitable financial decision. Other investors who may not care about these topics could also be attracted by the higher returns and consider investing in them. Both can be motivated by the results to act in sustainability investments.

Constraints

The datasets are only available for the past three years. Many ESG funds are new; therefore, their longer-term return data are missing. These data limitations may lead to robustness issues.

Deliverables

To complete this project, I will submit a github repo containing all the Jupyter notebooks that I will develop for each step of the project, a slide deck, and a project report.