Springboard—DSC Program Capstone Project 2

# The Financial Performance of the ESG Funds

Final Report by Tingyin Xiao

April 2024

## Introduction

The financial performance of funds is a crucial problem for individuals or financial institutions making investment decisions. The growing interest in ESG (environmental, social, and governance) funds over the past years leads to questions about whether sustainability investments are profitable. With appropriate and comprehensive data, we built models that can explore the answers to such questions and make predictions of future financial returns of individual funds.

## Data

1. Data Description

We acquired the monthly datasets suitable for this project 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 are dated monthly from March 2020 to the present, with information on returns to the end of 2022 (November 2022). The available data include 3000 funds with over 8,000 share classes for each month and only screen mutual funds with at least 50% invested in stock investments.

The datasets we worked on contain extensive information about the funds. Over ten variables record the funds' returns, and more than one hundred describe the funds, including their ESG grades and detailed ESG metrics for various categories. To predict the future returns of the funds, we need to build models that can learn from their historical returns and characteristics. The targets of this task are the return variables, and other variables can be the feature variables for regression algorithms. This comprehensive nature of the datasets reassures our stakeholders about the thoroughness of our analysis.

2. Data Wrangling

The combined table of the raw data has 365103 rows and 146 columns. There are 4706 distinct funds and 14125 distinct fund share classes. Each row observes a specific fund share class at a particular month, with its returns, characteristics, and various ESG evaluations as the columns. The data wrangling aims to organize and clean the data to prepare it for exploratory analysis.

We did the following steps to wrangle the data: firstly, we collected and merged all the monthly data files to a single data table with all available months; secondly, we cleaned the fund and share class names that included trademark symbols to ensure the names were matchable; thirdly, we merged duplicated columns with different column names due to the merging; fourthly, we dealt with the double-reported months by keeping the most updated information for each month; fifthly, we identified the fund share classes that have target variables over the most extended period (31 months); sixthly, we extracted the time series data and data tables with target variables covering this most extended period; lastly, we defined the target variable of interest as the month-end trailing returns of one year and the feature variables as all the other variables except the return variables. The data products from the wrangling include a time series of the target of interest and a full data table with the target and all the feature variables that have less than ten percent of missing data over the 31 months of the longest available period.

# Methods

1. Exploratory Data Analysis

The cleaned data have 172949 rows and 5579 distinct share classes. The data table with features has 91 columns, including six categories of ESG feature variables. We did exploratory data analysis to discover patterns in the target variable's time series and its relationship to some key features. We also looked at the relationship between the key features to explore their impacts on the target variable.

We found that the mean time series of the target variable of interest, month-end trailing return of one year, for all the funds share classes has a clear pattern: the returns increased until March 2021, then decreased until September 2022, and since then, they have started to go upwards again (Fig. 1). Comparing the mean time series of one-, three-, and five-year returns, we can see that only the one-year return has this pattern, while the other two variables are relatively flat during this period.

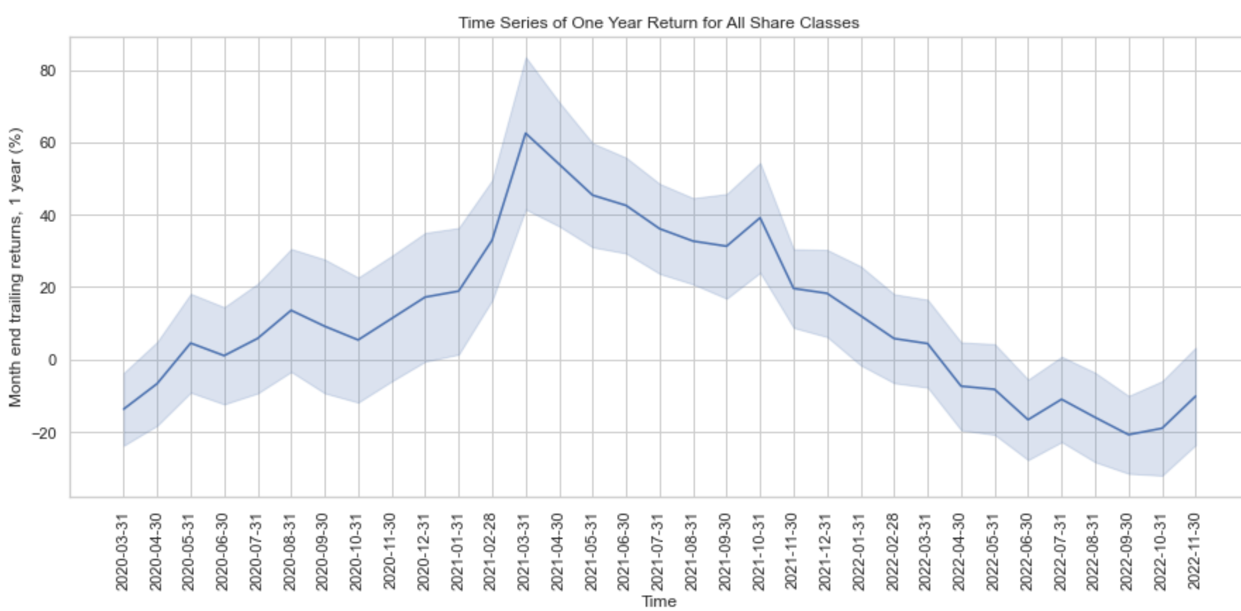


*Figure 1 Mean time series of month-end trailing one-year returns for all the funds share classes (unit: %). Shaded region around the line represents the confidence intervals around the mean using standard deviation.*

In the data table with both the target and feature variables, there are six categories of ESG features. We are most interested in the category involving fossil fuel-related variables. Using the fossil fuel grade variable, we can see from the mean time series plot for each grade that before the tipping point of March 2021, the clean funds (the ones with better grades such as A) performed better than the fossil funds (the ones with worse grads such as F) (Fig. 2). However, the fossil funds performed better when the mean values went down after that point than the clean funds. From the time series plot for the counts of share classes under each grade, we can see that since January 2021, the number of share classes of the most fossil-rich funds (grade F) increased, while the number of the clean funds generally decreased (Fig. 3).

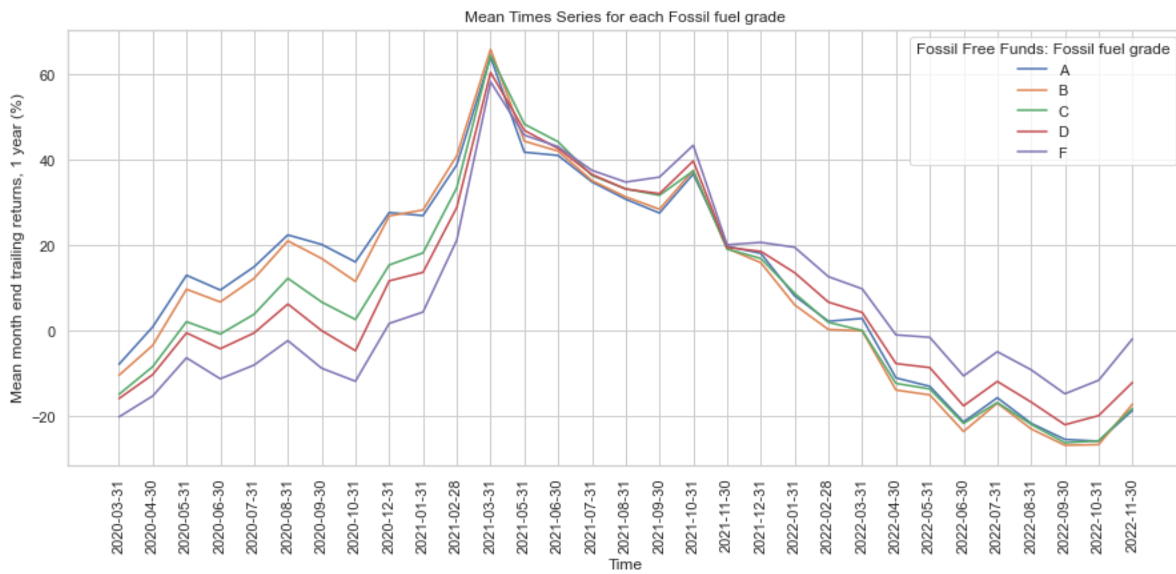


*Figure 2 Mean time series of month-end trailing one-year returns for the funds share classes grouped by their fossil fuel grades (unit: %).*

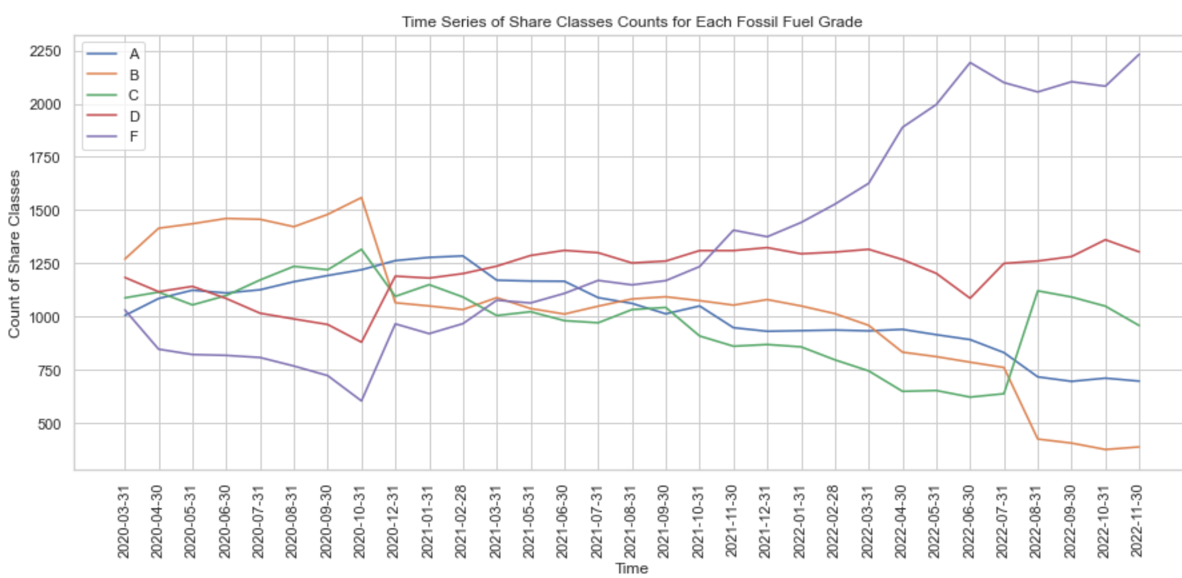


*Figure 3 Time series of the total number of the funds share classes grouped by their fossil fuel grades.*

We also examined the relationship between the fossil fuel-related features. Several feature pairs have high correlations. Some are redundant, as they are largely explained by other variables or are similar to another variable by definition. This pattern was further investigated and discussed during data preprocessing and model exploration.

2. Data Preprocessing

The feature variables in the data table include categorical and numeric data types, and some of them are time dependent. Before the final modeling step, we prepared the data using feature engineering techniques. Specifically, we transformed the categorical features into dummy features and scaled the numeric features to standardize their magnitude. We also handled the missing data in the categorical and numeric features. We created lagging and moving average features based on the target and feature variables to examine the influence of time dependency in the data.

3. Model Description and Application

We chose two methods, ARIMA (Autoregressive Integrated Moving Average) and Random Forest models, to model the time series and the regression problem concerning the time series, respectively. To explore the model application and performance, we first used them to predict a single time series. After exploring the modeling and cross-validation approaches, we built functions for any time series predictions based on these models. We applied them to a randomly selected subset of the whole dataset as examples. The model performances are evaluated by splitting the data to training and test sets. The models are fitted on the training sets and evaluated on the test sets. In the end, we compared the performance of ARIMA and Random Forest models in predicting each time series on the test sets.

The ARIMA model is a popular time series forecasting tool that aims to capture the temporal dependencies and patterns presented in sequential data. For each time series of a fund share class, we selected the best ARIMA model with the optimal parameters, including the order of differencing, autoregression, and moving average, as well as seasonal orders. Then, we used the best model to predict the returns and compare them with the test sets.

The Random Forest model is a widely used supervised machine learning algorithm. Here, we use it on a regression problem, but it can also be applied to classification tasks. It is an ensemble-based decision tree learning algorithm considered versatile, robust, and scalable. We used it to predict the funds' returns and explore the relationship between the ESG features and the returns.

# Results

For the time series of a single fund share class (1919 Socially Responsive Balanced A), using the ARIMA model, we find that in the best model, the order of the autoregressive term is 2, and the order of the moving average term is 2. The data is stationary and does not have evident seasonality. The best prediction horizon is 3 months, with mean absolute error (MAE) of 2.29, root means squared error (RMSE) of 2.72, and mean absolute percentage error (MAPE) of 12.74 using the best model. Using the Random Forest model, we find that the model prediction results have an

MAE of 6.39, RMSE of 6.74, and MAPE of 36.70 for the 3-month prediction horizon. The model performance is lower than the ARIMA model, which could contribute to the limit of observations and the noise introduced by too many features.

We performed the feature selection using Recursive Feature Elimination (RFE) coupled with cross-validation to improve the model performance by eliminating low-importance features. The results showed that the optimal number of features for this subset of data is four. The four most important features are the 3-month moving average of the relative carbon footprint (tonnes CO2 per 1M USD invested), the relative carbon footprint (tonnes CO2 per 1M USD invested), the 3-month moving average of the total financed emissions scope 1 + 2 + 3 (tCO2e), and the target variable with 1-month lag. If we only use these most important features, the model performance is improved, with an MAE of 5.54, RMSE of 5.80, and MAPE of 31.90. However, it still underperforms on a 3-month horizon compared to ARIMA.

Then, we built functions for these two models, applied them to ten randomly selected time series from thousands of fund share classes in the dataset, and compared their performances. The results indicate that the ARIMA models outperform the Random Forest models in terms of MAPE in 8 out of 10 cases (Table 1). Only for Fidelity Growth & Income K and First Eagle Global I do the random forest models outperform the ARIMA models. The performance of these models also differs for different funds. The ARIMA model performs worst for First Eagle Global I, and the Random Forest model performs worst for Vanguard Value Index I. They both perform best for Columbia Emerging Markets Adv.

*Table 1 MAPE of ARIMA and Random Forest Models Comparison for the ten randomly selected fund share classes.*

| Fund Share Class Name | ARIMA MAPE | Random Forest MAPE | ARIMA MAPE - Random Forest MAPE |
|---|---|---|---|
| American Century Equity Income Inv | 65.51 | 84.95 | -19.44 |
| BlackRock Global Dividend Inv A | 44.88 | 94.07 | -49.19 |
| Brown Advisory Sustainable Growth Adv | 23.31 | 55.26 | -31.95 |
| Columbia Emerging Markets Adv | 15.04 | 26.06 | -11.01 |
| Fidelity Growth & Income K | 82.9 | 80.18 | 2.72 |
| First Eagle Global I | 253.78 | 173.3 | 80.48 |
| Franklin Utilities Adv | 119.06 | 176.97 | -57.91 |
| SEI Tax-Managed Large Cap F (SIMT) | 52.25 | 68.49 | -16.24 |
| Vanguard Value Index I | 107.69 | 312.11 | -204.42 |
| iShares Russell Top 200 ETF | 26.5 | 57.1 | -30.6 |

We selected the ARIMA model as our final model for the prediction task based on the comparison. Here, we displayed one example of comparing the forecast value to the actual value in the test set for BlackRock Global Dividend Inv A (Fig. 4). The plot for each of the ten selected fund share classes can be found in section 4 of the Jupyter Notebook of the Modeling section. Figure 4 shows that the ARIMA model successfully captured the upward trend in the forecast horizon, though with a smaller variation.

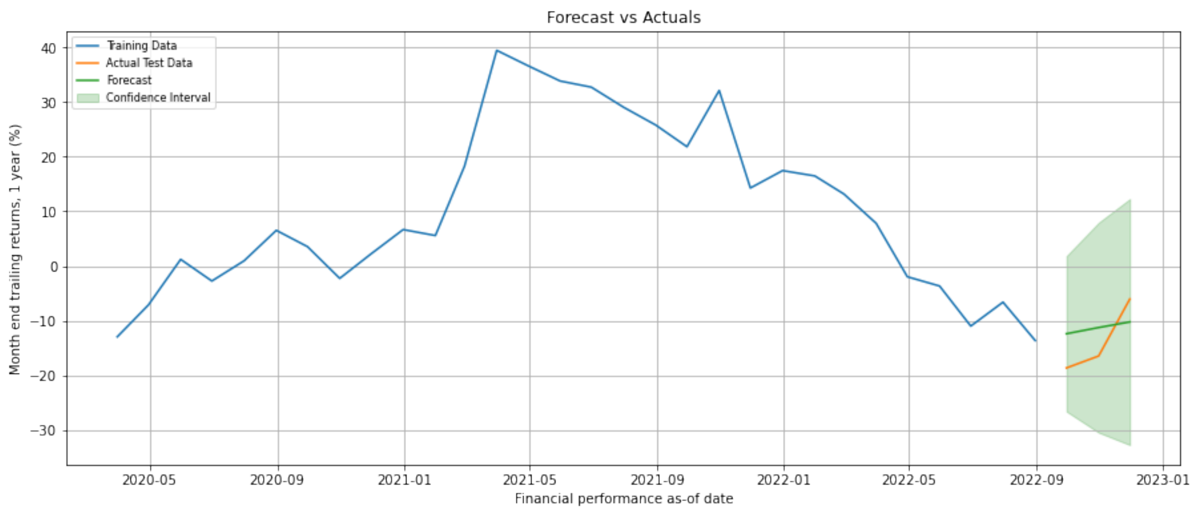


*Figure 4 Forecast vs. Actual values of month-end trailing one year return for BlackRock Global Dividend Inv A on a 3-month forecast horizon, using ARIMA model (unit: %), shaded region represents the 95% confidence interval.*

# Conclusion and Discussion

The stakeholders, investing individuals or financial institutions, will be interested in knowing whether the funds they are interested in are forecasted to gain or lose money. Our findings showed that using the ARIMA model in most cases can predict the month-end trailing one-year return for a 3-month horizon better than a regression model that includes many features. The result indicates that the time dependencies are crucial to the datasets we worked on. The stakeholders can use the function we built around the ARIMA model to forecast the future returns for any of the thousands of fund share classes in the datasets and use it as a reference.

Yet, our work is merely the tip of the iceberg. The Random Forest models we employed offer a wealth of untapped potential, from incorporating more relevant features to identifying more effective lagging windows and aggregation functions. Furthermore, exploring the impact of different regression models on the results could yield intriguing insights. Based on the ARIMA or improved/alternative regression models, we can also predict future returns for all the fund share classes in the dataset and rank them by their forecasted gains. The possibilities for further research are both vast and promising.

Our work also has limitations. The dataset's return variables are lastly updated by the end of 2022, making the target variable only available for less than three years during 2020-2022. Moreover, many ESG funds are new and evolving. Therefore, it is hard to justify long-term return patterns with such a short period. These data limitations may lead to robustness issues. In future studies on ESG funds, we need to acquire longer and more recent data to make a more reliable forecast for the stakeholders.