BANCA D'ITALIA

EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Mind the gap! Machine learning, ESG metrics and sustainable investment

by Ariel Lanza, Enrico Bernardini and Ivan Faiella

June 2020

Number 561

BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Mind the gap! Machine learning, ESG metrics and sustainable investment

by Ariel Lanza, Enrico Bernardini and Ivan Faiella

Number 561 – June 2020

The series Occasional Papers *presents studies and documents on issues pertaining to the institutional tasks of the Bank of Italy and the Eurosystem. The* Occasional Papers *appear alongside the* Working Papers *series which are specifically aimed at providing original contributions to economic research.*

The Occasional Papers *include studies conducted within the Bank of Italy, sometimes in cooperation with the Eurosystem or other institutions. The views expressed in the studies are those of the authors and do not involve the responsibility of the institutions to which they belong.*

The series is available online at www.bancaditalia.it .

ISSN 1972-6627 (print)
ISSN 1972-6643 (online)

Printed by the Printing and Publishing Division of the Bank of Italy

MIND THE GAP! MACHINE LEARNING, ESG METRICS AND SUSTAINABLE INVESTMENT

by Ariel Lanza,[1] Enrico Bernardini[2] and Ivan Faiella[3]

Abstract

This work proposes a novel approach for overcoming the current inconsistencies in ESG scores by using Machine Learning (ML) techniques to identify those indicators that better contribute to the construction of efficient portfolios. ML can achieve this result without needing a model-based methodology, typical of the modern portfolio theory approaches. The ESG indicators identified by our approach show a discriminatory power that also holds after accounting for the contribution of the style factors identified by the Fama-French five-factor model and the macroeconomic factors of the BIRR model. The novelty of the paper is threefold: a) the large array of ESG metrics analysed, b) the model-free methodology ensured by ML and c) the disentangling of the contribution of ESG-specific metrics to the portfolio performance from both the traditional style and macroeconomic factors. According to our results, more information content may be extracted from the available raw ESG data for portfolio construction purposes and half of the ESG indicators identified using our approach are environmental. Among the environmental indicators, some refer to companies' exposure and ability to manage climate change risk, namely the transition risk.

JEL Classification: C63, G11, Q56.
Keywords: portfolio construction, factor models, sustainable investment, ESG, machine learning.
DOI: 10.32057/0.QEF.2020.561

Contents

[1] Kellogg School of Management, Northwestern University (PhD student).
[2] Bank of Italy, Directorate General for Markets and Payment Systems.
[3] Bank of Italy, Directorate General for Economics, Statistics and Research.

1 Introduction *

There is a growing attention on how finance can contribute to sustainability objectives embedded in the 2030 agenda, and in particular to answer the urgency of climate adaptation and mitigation. The widening integration of sustainability criteria in the financial investing is increasingly prompted by regulators, business practices and investors.

According to Global Sustainable Investment Alliance the global assets managed with sustainability criteria have rapidly increased to 31 trillion USD at the beginning of 2018, spreading from traditional instruments to new specific assets such as green bonds. This trend is also supported by the search of stable risk-return profiles: there is an extensive literature showing that sustainable investment leads to risk-adjusted market returns that are often higher than those achieved using traditional financial models. Good ESG practices seem to provide firms with a competitive advantage stemming from different sources: they can contribute to innovation or to reduce operational, legal and reputational risks, leading to a more efficient resource allocation (as resources can be shifted from risk management to productive activities), thus favoring a better corporate financial and market performance, and lowering the cost of capital.

Market participants use EGS scores to account for these factors: this information is provided by private firms using methods that are not always consistent. In particular, the representation of the different ESG dimensions have different level of complexities (the "E" component being usually less heterogeneous and controversial because quantitative data and conceptual models are more easily available). In fact, there are neither broadly accepted rules for ESG data disclosure by individual firms, nor auditing standards to verify the reported data. ESG-score providers rely heavily on voluntary disclosures by firms and on subjective methodologies to select, assess and weight individual ESG indicators.

As a result, ESG scores of individual firms show a large heterogeneity across agencies compared, for example, with the credit ratings. There is also evidence of significant biases in ESG scores, which tend to overestimate companies that are larger and belong to specific industrial sectors and geographic regions.

Our research want to address the following question: to what extent are equity portfolio returns sensitive to ESG information? And how is ESG information absorbed in stock prices? In order to do so, we propose to overcome the current inconsistencies in the ESG scores by using Machine Learning (ML) techniques to better spot the most material E, S or G metrics for sustainable investing. In particular, ML techniques are used to identify those E, S, G indicators that better contribute to the construction

* We would like to thank Stefano Battiston and the other participants to the CEMLA and Banco de México "Conference on Climate Change and its Impact in the Financial System" (Mexico City, 5-6 December 2019). The opinions expressed are personal and do not necessarily reflect those of the Bank of Italy.

of efficient portfolios. ML can achieve this result without the need of a model-based methodology, which is typical of the modern portfolio theory approaches. Our strategy applies ML techniques using around 220 ESG metrics from two of the largest data providers, Refintiv-Asset 4 and MSCI ESG Research and sheds light on what are the prominent ESG indicators in determining risk and return differentials, along a time span from 2007 to 2019 for around 250 euro area listed companies. The novelty of the paper is threefold: a) the large array of ESG metrics being analyzed, b) the model-free methodology ensured by ML c) the disentangling of the contribution of ESG specific metrics to the portfolio performance from the traditional style and macroeconomic factors.

Our approach shows that the information content extracted by ML application from ESG indicators contributes considerably (in terms of size and statistical significance) to portfolio performance differentials (risk and return), even after taking into account the contribution of both standard Fama-French style factors and alternatively considering macroeconomic factors

The paper is organized as follows. In Chapter 2 we review the literature on equity returns, describing factor models used afterwards as a robustness check, we introduce ESG investing and the evidence of ESG contribution to risk and return of corporate financial performance, we discuss ESG limits and potential and we present some ML application to investment purposes. Chapter 3 describes our data set for financial data (index constituents and return time series) and ESG indicators, with a specific focus on the treatments adopted for completing the data set. In Chapter 4 our original setting of ML technique is presented together with the objectives of the portfolio construction framework. Chapter 5 reports the results and the analysis of main findings and presents a robustness exercise that checks if our results hold even after accounting for the two aforementioned factor models, using in-sample and out-of-sample tests. Finally, Chapter 6 concludes and discusses possible follow-ups.

2 Review of the literature

This work lies at the intersection of three different topics: modern portfolio theory and portfolio construction, ESG integration and applications of ML in portfolio allocation. We can find a vast literature about how factors, both fundamental and macroeconomic, affect stock returns and the relevant tests (one of the most relevant for our work is by Fama and MacBeth[23]); there is also a growing literature about the relation between ESG indicators and the risk-return features of securities (for example Khan et al. [39], MSCI [41] and Melas et al.[43]); finally there are papers about the effectiveness of ML techniques in predicting returns. However, combining

ESG integration with techniques of ML is a field which is still almost unexplored. There is some evidence that such a link might exist, for example in an article of the CFA Institute Research Foundation [26] the author touches both topics, however the focus is on the effectiveness of ML in retrieving ESG information rather than in analyzing stock returns or portfolio performance through ESG data. The originality of this research lies in the study of this kind of application.

2.1 Risk factors for equity returns

This work relies, as a background, on several factor models for equity returns based either on macroeconomic or financial variables. The first model is based on macroeconomic variables and was originally proposed by Burmeister, Ibbotson, Roll and Ross (Birr model) [10] in a formulation for the Euro area market proposed by Carboni [11]. This paragraph provides a review of factor models proposed within the portfolio theory, the specifications for the actual implementation in portfolio construction and their economic rationale. These class of models will be helpful in disentangling the contribution of ESG variables from the explanatory variables used by the factor models described below.

The Capital Asset Pricing Model (CAPM) has its foundations in the modern portfolio theory developed by Markowitz [42]. In his paper the basic setting is laid down: the investors are risk-averse and build their portfolios in one single period by minimizing the variance while selecting the preferred expected return or by maximizing the expected return after having selected the preferred level of variance, the concept of efficient frontier is introduced, this idea quantifies the intuitive result that, when portfolios are properly managed, more return is associated with more risk. The model is then extended by Sharpe [58] (who will received the Nobel Prize in 1990) and by Lintner [40]. They add the concepts of *complete agreement* between investors (on the real distribution of returns) and of *borrowing and lending at a risk-free rate*. Hence, they introduce the concept of a market portfolio, as the portfolio in which stocks are weighted proportionally to their capitalization. Some work relaxing the assumption on the existence of a risk-free rate for lending and borrowing is done later by Black [9], because of this work, the CAPM is sometimes referred to as the Sharpe, Lintner or the Black model.

Empirical tests, however, showed that returns are higher than those predicted by the CAPM for stocks issued by firms with the following financial characteristics:

- small price-earnings ratio [5],
- small capitalization [4],

- high debt-equity ratio [8],
- high book-to-market equity ratio [53].

Other critics to the CAPM regard the assumption of normality of returns and the use of historical variance as the only measure of risk (a good review of the limits of the Capital Asset Pricing Model can be found in Fama and French [22]).
To address this criticism to CAPM, the Arbitrage Pricing Theory (APT) was developed by Ross [54]. According to APT, in place of a single source of risk (identified as the exposure to the market in the CAPM), financial risk can be partially explained by different factors $f_i(t)$ and each stock is exposed to them through different factor loadings $\beta_{i,j}$-s. Specifically, the difference between the return on asset r_i (including dividends) and its expectation $\mathbb{E}[r_i(t)]$ at the beginning of the period is given by

$$r_i(t) - \mathbb{E}[r_i(t)] = \beta_{i,1} f_1(t) + \ldots + \beta_{i,k} f_k(t) + \varepsilon_i(t) \tag{1}$$

where $\varepsilon_i(t)$ is the value of the idiosyncratic shock at the end of the period. It is assumed that

$$\mathbb{E}[f_j(t)] = \mathbb{E}[\varepsilon_i(t)] = \mathrm{Cov}[\varepsilon_i(t), f_j(t)] = 0 \qquad \forall i,\ \forall j = 1, \ldots, k \tag{2}$$

and that the factors and the idiosyncratic shocks are uncorrelated across time, i.e.,

$$\mathrm{Cov}[\varepsilon_i(t), \varepsilon_i(t')] = \mathrm{Cov}[f_j(t), f_j(t')] = 0 \qquad \forall i,\ \forall j = 1, \ldots, k,\ \forall t, t' \text{ with } t \neq t'. \tag{3}$$

It can be shown that under the hypothesis of absence of arbitrage it is possible to find $k+1$ numbers $P_0, \ldots, P_k$ such that for every stock

$$\mathbb{E}[r_i] = P_0 + \beta_{i,1} P_1 + \ldots + \beta_{i,k} P_k. \tag{4}$$

Those P_j-s are interpreted as the *price of risk* or the *risk premium for the j-th risk factor*. In the special case in which a portfolio is not exposed to any risk factor (i.e., $\forall j\ \beta_{i,j} = 0$) we must conclude that the portfolio must be risk free. Since the only element left in this situation in the RHS is P_0, we must have that P_0 is the risk free rate TB. We can use for TB the 30-day Treasury bill rate $TB(t)$, known to investors at the beginning of the month.

From equations (1) and (4) we obtain

$$r_i(t) - TB(t) = \beta_{i,1}[P_1 + f_1(t)] + \ldots + \beta_{i,k}[P_k + f_k(t)] + \varepsilon_i(t). \tag{5}$$

Further refinements on how to choose the best equity risk factors led to the macroeconomic model proposed by Burmeister, Roll and Ross [10] (aka Birr model) and the five-style factor model proposed by Fama-French [20] (aka FF model).
The **Birr model** considers changes in fundamental economic variables such as investor confidence, interest rates, inflation, real business activity and a market index such in the CAPM. Burmeister et al. (2003) suggested to adopt the following risk factors:

Risk Factor	**Unanticipated change in**	**Measurement**
Confidence $f_1(t)$	Investors' willingness to undertake risky investments	Rate of return of relatively risky corporate bonds minus government bonds (twenty-year maturities).
Time Horizon $f_2(t)$	Investors' desired time to payouts	Twenty-year government bond minus 30-day treasury bill
Inflation $f_3(t)$	Short-run and long-run inflation rates	Actual inflation for the month minus predicted
Business Cycle $f_4(t)$	Level of real business activity	Change rate between expected value of a business activity index at the beginning and at the end of the month
Market Timing $f_5(t)$	Part of total return market portfolio which is not explained by the other risks and the intercept	Change rate between value of regressed index at the beginning and at the end of the month

In the work of Carboni [11], the variable **Confidence Risk** was firstly measured as the variation of Merrill Lynch - Total return index, EMU Corporates, BBB Rated, 7-10 Years (Bloomberg ticker ER44) and the variations of JPM Morgan Global Bond Index EMU 10 year + (Bloomberg ticker JNEU10), however his work concludes that a better indicator of Confidence Risk is given by the Option Adjusted Spread (OAS) built on the zero curve relative to government bonds of the euro area. The **Time Horizon Risk** is measured as the variations of JPM Morgan Global Bond Index EMU 10 year + (Bloomberg ticker JNEU10) diminished by the one-month LIBOR returns. The **Inflation Risk** is measured as the variation of the inflation index of the Eurozone computed by Eurostat (Bloomberg ticker CPEXEMU) diminished by its moving average of the previous 12 months, i.e.,

$$IR_t = \ln\left(\frac{CP_t}{CP_{t-1}}\right) - \sum_{i=1}^{12} \frac{IR_{t-i}}{12} \tag{6}$$

The **Business Cycle Risk** is found as the variation of the indicator that anticipates the economic activity seasonally adjusted and calculated by the OECD (Bloomberg ticker O11EA013)

$$BCR_t = \ln\left(\frac{I_t}{I_{t-1}}\right). \tag{7}$$

The **Fama-French five-factor model** is inspired by their earlier work on the three-factor model [21] and the basic idea is that firm's profitability and cash flows could have material effect on stock returns. That was based on the previous Gordon's dividend discount model [25] and reframed the relationship between expected return and internal rate of return derived from future dividend flows. Their approach combines previous findings that stocks with high profitability outperform [47], stocks being repurchased tend to do well [46], growing firms outperform firms with poor growth [45], and firms with high accruals are more likely to suffer subsequent earnings disappointments and their stocks tend to underperform peers with low accruals [52]. Furthermore, it was observed that small companies are considered more risky (the size effect) than the big ones, as they are less liquid, and companies with a high book-to-market price ratio (the value effect) generally outperform companies with a low book-to-market price ratio.

The Fama-French five-factor model used in the present analysis considers the equation for the excess return to the risk free rate as represented below (for simplicity the time reference is omitted):

$$\begin{aligned} R_a - R_{rf} = \alpha + \beta_{mkt}(R_{mkt} - Rrf) + \beta_{smb}SMB + \\ \beta_{hml}HML + \beta_{rmw}RMW + \beta_{cma}CMA + \varepsilon, \end{aligned} \tag{8}$$

in which R_a = asset return, R_{rf}= risk-free return , α = excess return over the benchmark, β_{mkt} = market factor loading (exposure to market risk, different from CAPM beta), Rmkt = market return, β_{smb} = size factor loading (the level of exposure to size risk), SMB = small-minus-big (the size premium), β_{hml} = value factor loading (the level of exposure to value risk), HML = high-minus-low (the value premium), β_{rmw} = profitability factor loading , RMW = robust-minus-weak (the profitability premium), β_{cma} = investment factor loading, CMA = conservative-minus-aggressive (the conservative investment premium) [55].

2.2 Approaches to sustainable investment

The increase in the investors' interest in Socially Responsible Investing (SRI) is a rather recent phenomenon which has grown swiftly. According to the Global Sustainable Investment Alliance [32], at the beginning of 2018, 31 trillion dollars (accounting for 26% of the world professionally managed investments) were managed in a sustainable way, half of which in Europe (approximately 14 trillions).

The rationale of the positive impact of Environmental, Social and Governance (ESG) issues on stock return, is that a "sustainable" company will face less risk related to environmental issues, regulation or lawsuits and could benefit more of the opportunities stemming from good ESG practices; the academic literature identifies those companies that adopt sustainable production methods are generally on the frontier of productive efficiency and they benefit from competitive advantage (e.g. from process/product innovation and customer satisfaction),and, by virtue of less exposure to operational, reputational and legal risks, they manage to achieve a lower cost of capital; they get higher valuation assigned by the investors which translates into superior market performance [15]. Data on sustainability has been vastly studied in the academic literature from many points of view, not only in relation to risk and return. For example, Cheng, Ioannou and Serafeim in [14] showed that firms that scored well in Corporate Social Responsibility (CSR) parameters had better access to finance and at a lower cost. From the point of view of risk management, an interesting result comes from Godfrey, Merrill and Hansen [31], where it is shown that there is an insurance-like property of CSR activity in case of negative events such as legal/regulatory actions.

Integrating sustainability issues in portfolio management is a delicate matter even from a theoretical point of view. As pointed out by Hoepner [36], at first researchers considered sustainability as a purely "ethical" choice, leaving aside any link with the traditional risk-return framework. According to this view, the responsible investment is limited to *screening* out some securities from the portfolio and at best could lead to a portfolio as efficient as the unscreened one, since adding constraints to a portfolio optimization problem could never lead to a diversification profile which is better than the unconstrained case. In fact, in accordance with the Capital Asset Pricing Model [19], with the reduction of the investable universe the diversification benefits are partly lost. Despite this was considered for many years the "inescapable conclusion", more recently Arnott [3] has shown that a series of equally weighted random portfolios of sample stocks taken from a benchmark outperform the same cap-weighted benchmark over a 40-year period. This leads to consider that the reduced universe portfolios have to use weighting schemes carefully adapted by using risk- and return-based factors. For the practical implementation, there is a tipping point in the threshold of (e.g. sustainability) filter beyond which the constraint is too strong and can reduce significantly the investment universe and hence negatively impact on diversification and performance. As for the case of sustainable investing, there are two further important considerations to make:

- as Hoepner [36] points out, the risk reduction due to diversification can be decomposed into 3 elements: the number of securities, their correlation and their specific risk. Showing that good ESG score is associated with a lower specific risk and this component might offset the negative effect of screening on the other first two elements, it is possible to avoid the "inescapable conclusion" showing that sustainability should be considered in a risk-return framework. Some empirical results have been provided by Verheiden et al. [59].

- As pointed out by Shoenmaker and Shramade [56], a substantial limitation of traditional analysis on the risk-return framework is that it involves mainly time-series analysis, which is *backward looking*. Evaluations on sustainability have the advantage of being *forward looking* since it is focused on the long term. This criticism is compatible with the hypotheses of adaptive markets, incomplete information and not-completely rational behaviors.

More recently other approaches to sustainable investing emerged, for example *impact investing*, in which the investor not only seeks a financial objective by optimizing the risk and/or the return of an investment, but also its social or environmental impact. Again, this choice should not be considered superficially, as Shoenmaker and Shramade [56] point out, there is a growing literature on the fact that corporations should have a broader objective than maximizing profits. In one of their articles Zingales and Hart argue that it is often too narrow to identify shareholder welfare as market value and that "money-making and ethical activities are often inseparable" therefore "companies should maximize shareholder welfare not market value" [34] . An enlightening example is about a company selling high capacity magazines (of the sort used in mass killings) having shareholders who are concerned about mass killings. For them it would be more efficient to ban the sales of ammunition rather than reinvesting the profits made by the company in gun control. This principle justifies the increasing popularity of impact funds, where investors can pursue financial returns while addressing social and environmental challenges.

An alternative is *ESG Integration*, the one investigated in this study, that consists in making investment decisions that include considerations of material ESG factors within the traditional financial modeling framework: ESG indicators are thus treated as other financial indicators in order to explain risk and returns.

Although the positive effect of ESG on returns is not-unanimous, a research conducted by Khan, Serafeim and Yoon [39], showed that firms with good ratings on material sustainability issues tend to outperform firms with poor ratings. [1] A similar study has been conducted on the European market by Giudici and Bonaventura [30] and shows that firms with better practices in all the three ESG pillars exhibit higher returns; furthermore, strategies that combine ESG tilt with fundamental indicators (like price-earnings ratio) seem to be more efficient. [2]

Although a complete review of this vast literature is beyond the scope of this work, we recall two meta-analyses produced in 2015 by Clark, Feiner and Viehs [15] (reviewing 200 studies) and by Friede, Busch and Bassen [27] (reviewing more than 2.000 studies); the latter concludes that in the majority of the studies (88%) the relationship between ESG practices and financial corporate performance is positive or non negative and the relationship is stable over time.

2.3 ESG: the silver bullet for sustainable investment or just lies?

While the first researches on corporate social responsibility (CSR) date back to the seventies, see for example Bowman and Haire [33], the acronym ESG was only introduced in 2005, and is only in recent times that ESG reporting became enough frequent and granular to allow for significant statistical analysis firm-wise. The ESG approach has the desirable property of providing the investor with a score or rating that factor in a huge amount of information on how a single firm is performing in terms of different dimensions of sustainability. Integrating ESG factors into equity investments is becoming the most common responsible investment practice and there is a general agreement on the multiple benefits of this strategy that encourages a growing number of investors to practice ESG integration. But how reliable is the information content of EGS scores? A very provocative article in 2018 by Allen [2] was expressing serious doubts on how much the investors are aware on the information they are conveying, creating a false sense of confidence regarding the stories that these numbers are actually telling. Recently also the IMF expressed concern

[1] Unfortunately applying those results to our work is not straightforward for two reasons, the first is that this study was conducted on data from Sustainalytics, but its reporting methodology changed recently, hence we have a limited timeseries to use with the new methodology and the coverage for European equities is rather limited. The second reason is that materiality was assessed through SASB tables, which has been originally designed for the US firms and it might be arguable to squarely apply them to European firms.

[2] The ESG scores are computed by means of ESG data points taken from Thompson Eikon ESG which have been weighted by the SASB matrix for 30 classes of variables.

regarding the quality and the consistency of the information they provide and called for a standardization of terminology and definitions [29]. In fact, the lack of generally agreed methodologies and formats in compiling ESG data and of auditing standards to verify what is reported by the firm is one of most pressing concerns on the quality of the ESG informational content. On top of that, ESG-score providers rely heavily on voluntary disclosures by firms that they have to complement with their own (different) estimates. Furthermore the providers apply subjective methodologies to select, assess and weight individual ESG indicators, which add to the arbitrary nature of ESG scores.
As a result, ESG scores of individual firms show a large heterogeneity across rating agencies compared, for example, with credit ratings [13], which end up with a low correlation among ESG rating (within intervals of 0.4-07),[6] and Table 1, in respect of correlations among credit ratings above 0.9.

Table 1: ESG score providers' cross-correlations

	Sustainalytics	MSCI	RobecoSAM	Bloomberg ESG
Sustainalytics	1	0.53	0.76	0.66
MSCI		1	0.48	0.47
RobecoSAM			1	0.68
Bloomberg ESG				1

Source: State Street Global Advisors (2019).

There is also evidence of possible biases in ESG scores, which tend to give prominence to companies that have larger size and belong to specific industrial sectors and geographic regions [17]. According to a recent decomposition analysis most of the disagreement is due to different measurement techniques; a different focus and weight of the individual E, S and G components play also a part together with "a priori" bias of the rating companies [7]

But although still lacking a standardized, detailed firm-level information they are central in order to design a portfolio that factor in the sustainable practices of the firms. With all their issues, ESG scores are a wealth of data that can complement (or even complete) the informative set for the investors and can play a role in shaping a more thorough asset pricing on the markets.

Burmeister et al. (2003) [10], when laid the foundations of their model, explicitly

Table 2: ESG score providers' cross-correlations in our dataset

Euroarea-ex Italy	RobecoSAM	Sustainalytics	Refinitiv-A4
MSCI	0,42	0,46	0,32
RobecoSAM		0,58	0,56
Sustainalytics			0,41

Italy	RobecoSAM	Sustainalytics	Refinitiv-A4
MSCI	0,54	0,54	0,60
RobecoSAM		0,67	0,53
Sustainalytics			0,56

Source: Our computations on providers' data.

discouraged the use of accounting data for reasons that partially could apply also to ESG data. One issue regards the reporting methodology. Many parameters are measured only on a yearly frequency and it is clear that a standard for ESG reporting is not going to be adopted soon, however, some data providers have been reporting ESG scores and raw data for several years. In addition to that, we have samples that are big enough for regressing each sector differently, choosing material indicators for each sector according to his business peculiarities. Thanks to this continuous evolution and improvement of data feeds, we could overcome the biggest differences among reports of different companies.

Starting from this consideration, and after checking that we have a similar low-correlation issue in our data (see Table 2), we devise a strategy that use ML techniques applied to the raw ESG data points in order to setup a heuristic selection process that will produce sample portfolios built up on the basis of their financial and sustainability performance.

2.4 Machine Learning in finance

Even if the use of ML in relation to ESG criteria is almost unexplored, sometimes is used for text mining like by Feiner [26], as briefly discussed previously. ML is a subject that undoubtedly experienced a surge in popularity in the last few years, and the term is often used more for advertising reasons that for practical reasons. It is not unusual to find some references in which the terms "Machine Learning" and

"ESG" are juxtaposed without a thorough explanation of the choices made and the specification of the chosen methods [16].

The applications of ML to portfolio choices is a wider field with a broader disclosure, see for example Chan et al. [12], however there are some issues specific to sustainable investment that we had to overcome and for which it was more convenient to develop our own model. The first issue is that we need a model that can be easily understood, i.e., if our model is a black-box it is difficult to check that is going in the proper "ESG" direction. If we have solid beliefs that sustainable investing will lead to better result on the long term we can not rely on a model which might produce as a result to invest in the "unsustainable" firms. Secondly it is important to focus on the long term rather then on the short term while many applications of ML are just focused on high frequency data.

3 Data description and treatment

The data for the analysis is made of descriptive and financial information of the companies and mainly timeseries related to stock returns and ESG data indicators as described below. For the two type of data (returns and ESG) one important step is the treatment of missing data. In this respect, we explain techniques that we have applied to overcome the issue with two objectives: safeguard the broader scope of securities for the analysis and ensure a proper significance of missing values.

3.1 Stock data and indices

The universe for our research is composed by the stocks that compose the EURO STOXX index, the index which tracks the top-300 stocks in the euro area with respect to capitalization, which is weighted by their capitalization and the free-float share. From the stocks in the index we excluded the companies of the financial sector. We used the monthly total return of each stock starting from 31/12/2000 up to 30/04/2019. For this work we consider as our universe the stocks which composed the index at 31/12/2010.This choice is based on the following considerations and however requires some attention. Let's assume that we decide to start the analysis as at 31/12/2000, using the universe of stock names available at the last date, 30/04/2019. What does it happen if we compare the index weighted with respect to capitalization with the equal weighted index? The result, shown in Figure 1 (left) might appear surprising: the equal weighted index outperforms the cap weighted index, recording an overall performance which is up to 30% bigger than the performance of the cap-weighted index. However, this is not what we would experience in a real setting,

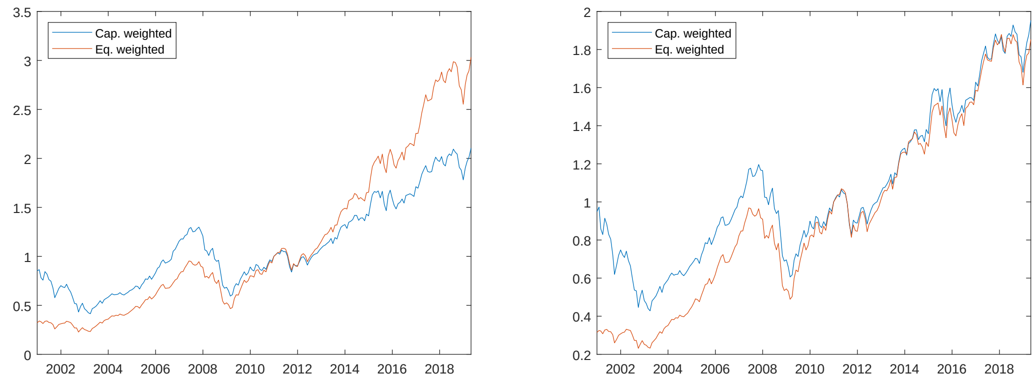


Figure 1: Comparing the equal weighted index with the index weighted by capitalization when the universe is chosen at the final date (on the left) and when the universe is chosen at 31/12/2010 (on the right). The price at 31/12/2010 is set to 1. Data from EURO STOXX 300.

this behavior is due to a specification problem called *survivorship bias*, i.e. we are picking stocks using information that is only available ex-post. Knowing that a stock is going to enter the index of the top 300 capitalized in a few years means that its price will grow more then the price of the stocks which are currently in the index, hence the equal weight portfolio made of those stocks at the end of the period is investing relatively much more in those stocks that are going to experience a significant growth. In practice, any investor that at 31/12/2000 had known which stock has been going to be among the top capitalized in 19 years would have had a huge advantage on other investors and might have experienced some peculiar dynamics as the one just shown. Furthermore, we did not need to take the universe at 31/12/2000 since the reporting of ESG data is absent at that date, hence we used the universe at 31/12/2010. In Figure 1 (right) we can clearly see that the price (that we set equal to 1 at the date of the choice of our universe) only shows this bias before 31/12/2010, while after the equally weighted and cap weighted portfolios do not show a significant difference. To summarize, we decided to use the 252 stocks comprising those existing at the beginning and at the end of the period, with time series available from 31/12/2006 to 30/4/2019 i. e. 125 dates, in order to use as much as possible of the available ESG data described hereafter.

3.2 ESG data retrieval and cleaning

3.2.1 Refinitiv-Asset 4

Refinitiv has expanded its offer of financial data with ESG ratings since 2009 with the acquisition of the Swiss provider Asset4 focused on environmental, social and governance data. After the acquisition, Asset4's ESG rating methodology was revised and improved. Reuters ESG team of 165 analysts covers about 1,700 companies in Europe, and its ESG scores time series begins from 2002. For each company, two numerical scores are drawn up, the "ESG score" and the "ESG Combined score", for the both a literal rating is also provided. The ESG score measures the performance, commitment and effectiveness demonstrated by companies with reference to the environmental, social and governance dimensions. The ESG Combined score complements the ESG score with the assessment of companies' controversies on ESG issues. The framework adopted divides the three pillars E-S-G into 10 categories, each of which is evaluated through a variable number of indicators based on the industry to which they belong, selected from a set of 178 indicators. To this end, the 54 "industry groups" of the Thomson Reuters Business Classification (TRBC) are taken as reference. In our study, after a first selection of 100 distinct reported ESG variables (such as the E, S and G scores, the level of carbon emissions, the number of accidents occurred to employees, etc.) available for the our investment universe made of 252. We added some economic variables (such as revenues, EBITDA, employees, etc.). We observed that some fields were missing (NaN) for some dates, whereas, after data editing and cleaning, we are left with 105 variables to explore. In dealing with missing values we should be careful in trying to understand what a missing value might mean. Usually the absence is related to the fact that the reported variable does not apply to the sector that we are considering or that the firm is not reporting the relevant information. In the case of variables that do not apply to firms, we often observe that many firms in the same sector have similar missing variables, hence we can proceed by removing whole columns; while in the case of the a firm not reporting the value, the reason might lie in the fact that the firm does not have the necessary resources to report a value even if the value is a good sign of sustainability, or that the firm prefers to provide no news rather than a bad news. Due to this potentially extreme alternatives, we have chosen to delete missing information rather than filling NaNs with any value (0, an average by sector or an average overall). An interesting analysis aimed at finding the ESG "exposure" for a company in order to extrapolate not reported ESG information is done in [35], however the results could hardly apply at granular level.

We stack the observations of each firm in a sector obtaining a tall matrix of regressors that could be used for linear regression. When we found that at some time a variable which was not reported for a firm but it was reported some time before, we suspected that it was a problem of how raw data was handled, however, in order no to make too strong assumptions, we left the NaN value instead of extending the last observation.

In order to have a matrix with only reported values, we needed to delete rows (observations) and columns (ESG data points) until the submatrix that is left for regression did not contain any NaN value. The problem of excluding as few properly reported values as possible, however, it is not a trivial one. It can be shown that it could be reduced to the maximum edge biclique problem[3], which is NP-complete [49]. Hence a good heuristic is needed. What we propose to do is to clean only the rows and columns with only NaN values at start, and then to carry on deleting the row or the column with the highest ratio of number on NaNs over size until no NaN is left. Finally we add all the rows that can be added. We decided to add the rows because we can obtain a better statistical evidence if our regression matrix has many more rows than columns, otherwise there is a higher risk of overfitting. We found that this approach produces a fairly good result, as sketched in Figure 2. It could be interesting, however, to further penalize the elimination of a row.

We also found important to modify some variables in order to obtain a better economic explanatory power. Variables such as CO2 equivalent emissions, waste, hazardous waste, environmental expenditures, energy use, coal energy purchased, coal energy produced, natural gas energy purchased, natural gas energy produced, oil energy purchased, oil energy produced and water used total were divided by the revenue. The injury rate, employee accidents, employees leaving and the training costs were divided by the number of employees. Contractor accidents were divided by the number of internal employee accidents.

Considering the *Utility* sector, which is composed by 22 firms, we are interested in data reported monthly from 31 December 2000, hence our dates are 221, giving rise to potentially $221 \cdot 22 = 4862$ observations. It is now critical how we proceed on variable selection with the application of our heuristics, we could obtain drastically different results depending on the order of the application of the selection process. We are aware of this difference by applying or not the substitutions of NaNs with the previous reported value.

[3]This is a problem in graph theory that consists in finding the clique with the maximum number of edges in a bipartite graph. Rewriting the problem in terms of adjacency matrix (or, more properly, biadjacency matrix) we obtain the reductions needed to show the equivalence with our problem.

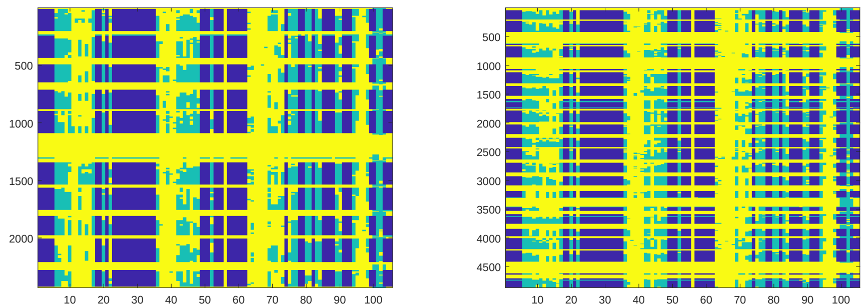


Figure 2: A graphical representation of the data picked respectively in the Oil&Gas sector (left side) and in the Utilities sector (right side). Each colored point represents a value in the regression matrix. The rows are the observations and the columns are the 100 chosen variables. The NaNs are painted in yellow, the reported numbers that are picked are painted in blue while the reported numbers that are not picked are represented in green.

Extending over NaNs For our 96 variables we obtain a total of not NaNs which is 264984 over the potential $221 \cdot 22 \cdot 96 = 466752$. Applying our heuristic to the *Utility* sector we are left with 56 variables from the 96 we had at the beginning and 2684 observations from the 4862 we were starting from. This gives a total of 150304 data points used over the 264984 which are not NaNs. Considering only the last date of the year over our starting value of 221 possible dates we are left with 19 dates, hence giving rise to potentially $19 \cdot 22 = 418$ observations. Only 217 dates are in our screened list of 2684 observations. However we still have to delete those rows for which we do not have an annual return (i.e., the rows relative to 31 Dec 2018), those are 20, we are therefore left with 197 total observations; while over the 56 variables, only 49 are non constant in the remaining set, hence we are left with $197 \cdot 49 = 9653$ data points. Diving this set in a training and a test set, where the test is given by the last two years of observations, we are left with a training matrix of dimension 157×49 and a test matrix of dimension 40×49. However, we notice that doing the row and column selection prior to the extension of the last reported value over the NaNs we obtain the data size reported below

	Utility	**Algorithm**	**Date**	**Returns**
Rows	4862	2684	217	197
Columns	96	56	56	49

Not extending over NaNs However, if we do not assign the previous valid value to NaNs, the total number of reported values is 225252 instead of 264984. After the algorithm selection we are left with 44 variables and 3012 observations, after the application of the date selection we are left with 251 dates, only for 249 we can find a return and only 38 dates vary over time. We are left with 9462 data points. Even if some data points are lost, the rows are more than 6 times the columns, while in the other case the rows were approximately 4 times the columns.

	Utility	**Algorithm**	**Date**	**Returns**
Rows	4862	3012	251	249
Columns	96	44	44	38

Selecting the rows as a first step Probably, the most robust approach would be to first select only the dates that we are going to use in the regression, eliminate constant columns, use our heuristics and finally eliminate (again) constant columns. In this case the final result is the same as in the preceding case:

	Utility	**Date**	**Algorithm**	**Column**
Rows	4862	321	249	249
Columns	96	96	44	38

Reporting dates As it can be inferred from Figure 2, less variables are reported as we go back in time. The first date with a reported variable for Refinitiv was 31 October 2001.

3.2.2 MSCI

The other data provider considered is MSCI ESG Research, which provided 172 ESG variables. MSCI ESG Research is a subsidiary of MSCI Inc., born in 2010 following the acquisition of RiskMetrics Group and the reorganization of the companies Innovest and KLD, both focused on ESG research. MSCI ESG Research is organized with a team of around 185 dedicated analysts covering approximately 1,500 companies in Europe. The ESG rating time series covers a time span of 20 years. MSCI ESG Research is currently the largest ESG rating provider, its analysis are used for the construction of around 600 equity and bond indices.MSCI provides a literal ESG rating scale from AAA to CCC grade that summarizes the exposure of companies to the risks and opportunities arising from specific key issues on the environmental, social and governance profiles and the ability to manage these issues. The rating is expressive of companies' ESG profile in comparative terms, as it results from the comparison of the scores of companies operating in the same industry. The MSCI framework divides the three E-S-G pillars into 10 themes, in turn divided into 37 key issues of risks and opportunities For our study, the data is available from January 2007 to June 2018. The reporting dates for ESG scores are not necessarily periodic and are not the same for every stock. As in the case of Refinitiv, a score for the E, S and G components was also provided. The other variables are defined as "key issues" (for example, raw material sourcing, product carbon footprint, etc.). Those key issues have an overall score which is obtained aggregating a risk-exposure score with a risk-management score, among the variables we also count the weight that is given to the key issue in the evaluation of a company. We have decided to exclude the weight of the key issues in our evaluation and we only conserved the three scores and the key issues for a total number of 112 data points.

Also for this provider less variables are reported as we go back in time, however older data is scarce since the first reported date was January 2007 which we have chosen as starting date for our analysis.

3.3 First trials

A first plain-vanilla approach has not resulted to be very promising. We used the MATLAB built-in regression learner in order to try several different regressions. Our dataset was the result of the heuristic selection applied to the full 56134×96 original regression matrix where in order to select less rows it was imposed that a row was eliminated only if its NaN ratio was greater than the NaN ratio of each column at the power of 0.1. The selection left 41 variables and 2841 observations. After that selection a constant column was added, as long as a dummy with a different value for each firm, a dummy with a different value for each sector and a variable with the return of the sector. Therefore 45 variables in total. In order to estimate the goodness of the fit we considered the RMSE on an 8-fold validation, where a RMSE of 0.35054 was obtained using only the constant value. The best RMSE, namely 0.2817, was reached in the regression done with bagged trees with the single variable sector return, that was by far the most explanatory variable. The same method with all the variables gave a slightly worse RMSE, namely 0.29615. However, removing only the sector return, the RMSE rose to 0.35291, even worse than the only constant value.

However, the fact that those first results were not promising does not imply that the data does not have explanatory power, i.e. "absence of evidence is not evidence of absence". We suspected that several aspects might have impacted negatively on this preliminary results. First of all, some data was lost in order to work with rectangular matrices, in addition to this, any regression-type of analysis affects only indirectly the choice on portfolio and therefore might not appreciate some properties that emerge only when stocks are grouped in a portfolio. In addition to this, we wanted to have the possibility to study different portfolio metrics, like Sharpe ratio, variance, or mean return. This lead us to develop a specific ML method.

4 Our Machine Learning approach

This section describes the approach that we have used to select the relevant ESG factors, the reasons that led us to the specific development and the practical choices we have made.

4.1 A promising approach

A useful experiment consisted in creating portfolios which are reallocated annually in which stocks are equally weighted. This allowed us to make a first comparison of the best ESG performers versus the worst ESG performers, factor by factor. We decided to create portfolios by dividing the stock in "best" and "worst" performer where "best" and "worst" are referred respectively to the top and the bottom quartile. We found that the aggregate ESG scores computed by the data providers on a subjective basis systematically led to lower returns for the most ESG-compliant companies. This happened also when we considered the "Environmental", "Social" and "Government" scores provided by each provided one by one instead of considering the aggregate ESG. This was quite surprising since for many factors, as the "CO2 emissions" divided by the revenues, the portfolio of the less polluting companies performed better than the portfolio of the worst companies.

In order to keep the model simple and informative, we decided to stick to the equally weighted portfolios. We noticed that a more flexible choice of the thresholds (rather than the arbitrary quartile choice) could lead to slightly different but sometimes significant results. For example, a particular choice of thresholds could lead to a group of highest scoring companies on the Refinitiv Environmental score to perform better than a group of worst scoring, despite the choice of the quartile showed the opposite situation. We decided as an additional contribution of this research to use our machine learning to find automatically those thresholds in order to obtain the highest possible performance for the ESG-compliant companies. It is worth noting that despite this choice could significantly increase the risk of false positives, it could be the only systematic way to appreciate the value of "weaker" factors. In a certain way we have decided to do the exact opposite to a subjective choice. For this reason a "human" way to check the resulting tresholds could be very valuable.

4.2 Tree-based approach: general idea

Our ML approach for portfolio construction is made of two steps: first we use ML to select the 10 most meaningful ESG indicators in three type of trials (for different financial objectives), second we combine those indicators together to construct portfolios, which are tested afterwards. In order to systematically find the most significant ESG indicators that could provide a portfolio extra performance, we decided to check for those indicators that can help for stock selection and portfolio construction aimed at maximizing the best-minus-worst (BmW) portfolio differential in terms of three financial metrics on a 12-month horizon, namely:

- mean absolute return;
- variance;
- Sharpe ratio;

Our first experiments suggested that a tree-like structure could have been the best way to automate our research. An early work on the use of trees for Corporate Governance factor selection can be that of Misangyi and Acharya [44]. However, decision tress were used for regression or classification. Our idea consists in building trees with the objective of optimize a variable which is not the RMSE, but a portfolio financial metric which implies maximizing the metric in the case of the mean absolute return or the Sharpe ratio, while minimizing the variance.

In order to go in the "ESG direction" (i.e. increasing with respect to the ESG scores, negative with respect, for example, to the carbon emissions), we impose to the tree to divide the stocks in a *best* portfolio and a *worst* portfolio where the stocks in the *best* portfolio are more "sustainable" than the stocks in the *worst* portfolio. The choice of the variable of selection and of the thresholds for the split is the one that gives the best optimization result for the chosen portfolio metrics, after having tried all the possible variables with all the possible thresholds in the set that are $\{20\%, 25\%, 30\%, 35\%, 40\%, 45\%, 50\%\}$ for the lower bound and, as a complement, $\{80\%, 75\%, 70\%, 65\%, 60\%, 55\%, 50\%\}$ for the upper bound. Formally every possible couple of bounds should be tested but a simple optimization argument allows the algorithm to be linear instead of quadratic in the number of different thresholds to try. When we deal with trees, we place a root (usually at the top) and start by

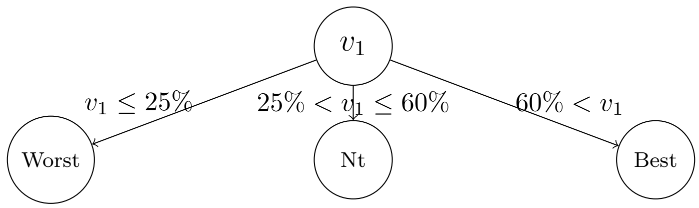


Figure 3: The first split of our tree is sketched. The lower threshold is 25%, meaning that all the stocks that have a score (given by the variable v_1) that falls in the lower quartile are assigned to the "worst" porfolio. While the stocks with score in the top 40% are assigned to the "best" portfolio.

creating splits that generates new branches. We will explain hereafter what our trees do by starting from the meaning of the first split.

The first split is illustrated in figure 3, is equivalent to dividing the stocks in the best percentiles and comparing them to the ones in the worst percentile. We write on each branch the values of the thresholds. We notice that unlike the most used decision or regression trees our splits are not necessarily binary (i.e., with only two branches per split) but allow for a "neutral" node in which we put all the stocks which are not in the best nor in the worst portfolio.

The power of the tree methods comes from the interaction between the variables, that can be perceived by adding new splits. However, adding too many splits could complicate the understanding of the model. We have decided to limit our structure to a 2-level tree in order to maintain a good level of interpretability. We added a second split identical to the first one, but that sorts our stocks with respect to a second variable starting from the neutral node, this split can promote stocks that were put in the neutral portfolio after the first split if the score relative to the second variable is high in the ESG direction, it can leave them in the neutral zone or put them in the worst portfolio if the score is low. A third split is added using the same second variable in order to have the possibility to downgrade to neutral (but not to worst) stocks that were put in the best portfolio at the first step (see Figure 4). The idea behind these choices was to leave space for the second variable to "correct" the sorting of the first, paying attention not to introduce enormous change (such as going directly from best to worst) or not to recover stocks with a particularly poor score.

The strength of this approach is that all the available data at each time is used and that it looks straight at portfolio performance rather than at those indicators that could suggest a good portfolio performance. The model also allows for a simple

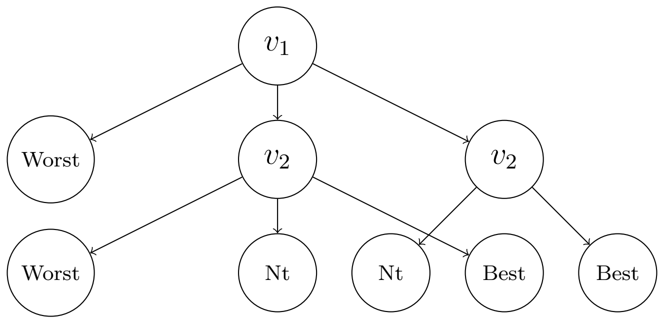


Figure 4: The second split for our trees.

interpretation of the results. The drawback is typical of the ML techniques whereas with this approach we are only concerned in learning from empirical evidence, explanations and corrections are left to humans. However, unlike some recent uses of ML in finance, our approach has the advantage of being tailored to look for long term performance rather than for the study of high frequency data, given that our model has been set up for one-year performance metrics as an objective. Overall although we tried to keep our exercise as much comprehensive and parsimonious as possible, the burden of numerical calculation is quite remarkable as it needs to consider 252 stocks, 125 dates, 217 ESG indicators with 7x2 (best and worst) thresholds; every combination is repeated three times, according to the three financial objectives.

4.3 Training the trees

We have chosen to consider the time period 2007-2016 as the training period, while the test period was 2017-2019.

After finding the best first split for each ESG variable, for each of those first split the tree best second splits were selected and the best thresholds for the third splits were computed. We have given a score in order to weight each ESG factor (weighting score). In order to include the impact of a variable also in interaction with other variables, we found the base score as the difference between the best and the worst portfolio for the chosen financial metric at the first split. We added to this base score one third of the increase in score given at every positive contribution at a second or a third split, excluding those contribution that left in the last five years less than 5 stocks in any portfolio (best or worst).

Then the ESG variables were sorted by their overall score and a worst and a best portfolios were made using the top 10 variables, selecting the stocks classified as best first split for each and weighted with respect to the score of the variable in such a way that, starting from equal weight, no difference in score could provide a tilt greater that one fourth of the weight to each portfolio. The same analysis was afterwards repeated using only environmental variables in order to focus on the matters that look incleasingly considered by investors as a source of material risk. Finally the portfolios were tested in-sample and out-of-sample for each of the portfolio financial metric and the relevant returns were regressed with respect to the Fama-French 5 style factors and with the macroeconomic variables in the Birr model. We can anticipate that, as expected, we found a strong correlation with the market portfolio (it should not be surprising since we are working inside the universe of the benchmark). The alpha (intercept) in each regression was always higher for the best portfolio, with the higher levels of statistical significance for the mean absolute return optimizations.

5 Analysis of results

We present results of our analysis separately for the cases where we considered the 3 risk/return metrics used as an objective of the portfolio construction, namely:

- mean absolute return;
- variance;
- Sharpe ratio.

Using equation 8 we test if portfolios built upon the ML-selected ESG indicators show a return or risk differential between the Best minus Worst (BmW) portfolios not fully explained by the Fama-French risk factors (or style factors), such as market, size, value (B/M), operating profitability, the conservativeness and then the residual extra-return can be attributed to the alpha generated by the ESG key indicator.[4] A similar factor analisys is performed to disentangle the contribution of macroeconomic variables of the Birr model from the BmW portfolios' risk and return metrics using the equation 5.
For each of the cases we provide information about the ESG indicators (first exercise, commented in par.5.1) and the environmental indicators only (second exercise in par.5.2), that we found the most material and we show the following information:

[4]The F-F five factors for the regressions of our portfolios are taken from the Kenneth French data library for Europe available on his website and converted in EUR terms with the correspondent USD/EUR rates (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

- the tables with the 10 ESG indicators, showing the score (weight) of each indicator in combination with an other indicator or alone, whether the indicator is a bivariate variable or not, the type (environmental, social or governance), the threshold we found significant for discriminating best over worst portfolios at first and second split, the minimum size (number of securities) of the best and worst portfolios;
- the graphs of the price return and the number of stocks for the best and worst portfolios, which show the overall simulation and in-sample and out-of-sample exercises;
- the metrics of the monthly return, variance, Sharpe ratio and maximum draw-down for the best and worst portfolios, both for in-sample and out-of-sample exercises;
- the tables with statistics for the regressions of the (best/worst) portfolios returns with the factor models (Fama-French 5 style factors and Birr) in order to assess the additional contribution of the ESG indicators (as the intercept of the regression can be considered as the alpha of the ESG component) and their significance (pValue and other statistics).

We found that the best portfolios in-sample were the best also out-of-sample, each one with respect to each portfolio metric. Only the return out-of-sample of the best portfolio obtained optimizing the difference BmW in variance was less than the return out-of-sample of the worst portfolio. Good results were obtained also with respect to the draw-down, which was always smaller for the best portfolios than for the worst ones, both in-sample and out-of-sample.

5.1 Results for ESG indicators

The analysis of the portfolio construction with 10 ESG indicators highlights that the indicators selected for maximizing the difference BmW of **absolute return** provide positive outcome, that holds true in-sample and out-of-sample with a yearly return difference around 4.5% and 1.2% respectively (which are computed as 38 and 10 bps on a monthly basis)(Table 3) . Given a pretty small increase in the variance, the Sharpe ratio difference BmW improves by 0.039 (see A.1).

In disentangling the factor contribution with the Fama-French (FF) model, the alpha generated by the ESG indicators provides a return difference BmW of 3.7% annualised (31 bps per month) and a similar magnitude with the Birr model (3.3%), which are both statistically significant. The graph on the right shows that number

Table 3: ESG indicators

	Absolute return		**Variance**		**Sharpe ratio**	
	in-sample	*out-of-sample*	*in-sample*	*out-of-sample*	in-sample	*out-of-sample*
Return BmW (annualised)	4.5%	1.2%	1.2%	-0.6%	2.4%	0.5%
Variance BmW (ann.)	0.01%	-0.02%	-0.12%	-0.09%	-0.18%	-0.09%
Sharpe ratio BmW	0.073927	0.03856	0.026612	0.02058	0.048533	0.046937
alpha FF BmW	3.66%		0.81%		1.70%	
alpha Birr BmW	3.28%		0.23%		1.07%	

of stocks of the best and worst portfolios increases along the time, as more data are available for the selected ESG indicators. This pattern is similar in all the exercises we have done and it underlines how helpful would be for the investors to broaden the universe of disclosing companies.

In the optimization of the difference BmW for the **variance**, the results show that the 10 ESG indicators contribute to construction of best portfolios which lower slightly the variance both in-sample and out-of-sample (-12 bps and -9 bps on yearly basis respectively), but also provide with a better Sharpe ratio (by 0.020 out-of-sample), as return is substantially similar. In disentangling the factor contribution with the FF factor model and Birr model, the alpha generated by the ESG construction provides a difference BmW of 0.8% annualised (7 bps per month) and 0.2% (2 bps monthly) respectively, which are both statistically significant for the best portfolios.

As for maximizing the difference BmW of **Sharpe ratio** the results in-sample and out-of sample are similar, with a difference in Sharpe ratio of 0.049 and 0.047 respectively; this case provides also positive results in the return difference BmW (+2.4% yearly in-sample and +0.5% out-of-sample) and annualised variance (-18 bps and -9 bps). The disentangling of the factor contribution with the FF factor model and Birr model shows that the alpha generated by the ESG indicators provides a difference BmW of 1.7% annualised (14 bps per month) and 1.1% (9 bps monthly) respectively, which are both statistically significant for the best portfolios.

Among the most material ESG indicators which are used in our portfolio construction, 9 out of 17 are related to environmental issues. That underlines the relevance of environmental issues for the equity portfolio performance.The environmental indicators relates not only to the carbon emissions (carbon intensity) but also to the waste management, recycling and eco-innovation. Interestingly the environmental score of one of the providers is identified as material but is not on the first ones. On the others indicators, 5 are related to social profiles (mainly about employees' safety) and 3 to governance factors, with a material role of diversity. Only 4 ESG variables

Table 4: The most material ESG indicators

	Return	Variance	Sharpe	Total	Biv	Type
CO2 Emissions/Revenue	0.037744	-0.0081091	1.0305	3	0	ENV
Waste/Revenue	0.033685	-0.013162	0.85332	3	0	ENV
Hazardous Waste/Revenue	0.016597	-0.012686	0.49405	3	0	ENV
Employee Accidents	0.011874	-0.0023094	0.20353	2	0	SOC
Specific Board Skills	0.011174	-0.00026921	0.29618	2	0	GOV
Controversial Sourcing Exposure	0.0099531	-0.00038647	0.29592	2	0	SOC
Total Injury Rate	0.0095307	-0.0025926	0.17775	2	0	SOC
Bribery, Corruption, Fraud Controversies	0.0082332	-0.0074764	0.33058	2	1	SOC
Nuclear	0.0054778	-0.0085057	0.20614	2	1	ENV
Energy Use/Revenue	0.0049003	-0.0035541	0.22408	2	0	ENV
Eco-Design Products	0.014168	0.0014373	0.12906	1	1	ENV
Long-term Compensation Incentives	0.0086402	-0.00011075	0.10977	1	0	GOV
Environmental Score	0.0083413	0.00071022	0.073592	1	0	ENV
Waste Recycling Ratio	0.0072689	0.00011708	0.2286	1	0	ENV
Board Diversity	0.0063854	-0.00026604	0.24095	1	0	GOV
Women Employees	0.0053944	-0.0022783	0.18377	1	0	SOC
Animal Testing	0.0029715	-0.0053909	0.10969	1	1	ENV

are bivariate (Table 4).

The exercises with the 17 indicators show that the Best portfolio over-performed the Worst portfolio both in-sample and out-of-sample for the three financial objectives, with lower performance for the objective of variance optimization (out-of-sample), while positive results are provided with the objective of Sharpe ratio difference maximization, and remarkable good results for the objective of absolute return where also the variance (out-of-sample) and alphas are in favour of BmW.

5.2 Results for Environmental indicators

The analysis of the portfolio construction with 10 environmental indicators finds, beside those indicators identified before, others complementary indicators. In the case of maximizing the difference BmW of **absolute return** the environmental indicators have proved to bring higher differential return out-of-sample compared with the ESG indicators, with a annualised return difference of 1.8% (compared to 1.2% for ESG indicators), lower variance and therefore a higher Sharpe ratio to 0.07 (Table 5). Also the in-sample results show a positive BmW difference for the return (+2.8% on annual basis) and Sharpe ratio (0.04). The analysis of the factors contribution shows that the alpha generation by portfolio construction with environmental indicators is significant both with the Fama-French model (2.8% annual and 24 bps monthly) and with the Birr model (2.0% annual and 17 bps monthly)(see Appendix).

In the optimization of BmW difference in **variance**, the results show that the

Table 5: Environmental indicators

	Absolute return		**Variance**		**Sharpe ratio**	
	in-sample	*out-of-sample*	*in-sample*	*out-of-sample*	in-sample	*out-of-sample*
Return BmW (annualised)	2.8%	1.8%	0.2%	0.8%	3.2%	1.8%
Variance BmW (ann.)	0.03%	-0.05%	-0.06%	-0.07%	-0.26%	-0.10%
Sharpe ratio BmW	0.04461	0.069083	0.0080234	0.047756	0.070629	0.08947
alpha FF BmW	2.84%		0.63%		2.91%	
alpha Birr BmW	2.01%		-0.19%		1.37%	

10 environmental indicators contribute not only to reduce the variance but also to a positive annualised return difference (0.2% in-sample and 0.8% out-of-sample) and to Sharpe ratio increase (+0.08 and +0.05 respectively). The alpha generated seems not relevant, as it is positive with the FF factor decomposition (+0.63% annualised) and slightly negative with the Birr model (-0.19%), which is statistically more significant.

The maximization of the difference BmW for the **Sharpe ratio** shows very positive results in-sample and out-of sample for all the financial measures: the annualised return increase is 3.2% and 1.8% respectively, the variance reduction is 26bps and 10 bps, the Sharpe ratio increase is about 0.07 and 0.09. Also the disentangling of the factor contribution shows that the alpha generated by the environmental indicators is remarkably of 2.9% annualised with the FF factor model and 1.4% with Birr model, which the best portfolios statistically significant in both cases.

Table 6: The most material Environmental indicators

	Return	**Variance**	**Sharpe**	**Total**	**Biv**
Waste/Revenue	0.013808	-0.0072546	0.35684	3	0
CO2 Emissions/Revenue	0.013171	-0.0057402	0.35125	3	0
Hazardous Waste/Revenue	0.0051957	-0.0079682	0.16283	3	0
Climate Change Theme Score	0.00338	-0.0016526	0.11476	3	0
Waste Recycling Ratio	0.0080097	8.1759e-05	0.2737	2	0
Prod. Carbon Footprint Score	0.0041826	0.0002437	0.14035	2	0
Prod. Carbon Footprint Mgmt	0.0038396	0.00025081	0.14645	2	0
Emission Reduction Objectives	0.0038287	-0.00071986	0.071566	2	1
Water Use/Revenue	0.0018263	-0.00064089	0.1227	2	0
Eco-Design Products	0.0075791	0.0014373	0.10354	1	1
Environmental Score	0.0068444	0.00079796	0.083197	1	0
Energy Use/Revenue	0.0030538	-0.0021816	0.095028	1	0
Opportunities in Renewable Energy Score	0.0029098	7.8753e-05	0.11179	1	0
Nuclear	0.0025489	-0.0031798	0.068439	1	1
Opportunities in Clean Tech Score	0.0024122	0.00036662	0.11353	1	0
Opportunities in Renewable Energy Exp	-0.00052838	-0.00047105	-0.011381	1	0
Animal Testing	-0.0026548	-0.0018501	-0.077037	1	1

Among the most material environmental indicators, besides the indicators already analysed for the ESG case study, some refer to the assessment of providers, highlighting how it is significant the forward-looking evaluation for the environmental issues and climate-change risk, namely the transition risk, which implies the assessment both of the exposure and the ability of the corporates to manage such risks and move forward adaptation techniques, like renewables and clean technologies (Table 6). As we are not able to measure to what extent the evaluation by providers integrates climate-related scenario analysis, it may be worth checking for additional indicators that represent such scenario and possibly stress test considerations.

6 Conclusions and future developments

The ESG investing is experiencing a remarkable growth in terms of assets and number of investors, but the transparency and consistency in the ESG assessment is still developing. Once it reaches a higher degree of harmonization and consistency, included a convergence with the Taxonomy on Sustainable Activities that is being finalized at the European level, it can trigger the appetite of investors for integrating ESG factors in their strategies and thus fostering the growth of sustainable finance. Our research proposes a model-free approach that overcomes some of the limits of ESG scores and that sheds light on the positive financial results from a proper integration of ESG and more specifically environmental risks into investors' portfolio choices.

According to our results, more information content may be extracted from the available ESG indicators: of those selected by our ML technique half are environmental and some refers to companies' exposure and ability to manage climate change risk, namely the transition risk. Among the environmental factors selected, only one corresponds to the environmental score of a provider: this means that the ESG scores do not exhaust the information available in the raw data disclosed by the firms. As we were not able to measure to what extent the evaluation by providers integrates climate-related scenario analysis, it may be worth checking for additional indicators that represent such scenario and possibly stress test considerations.

The portfolio construction based on ESG and environmental indicators identified with our ML application contributes considerably (in terms of size and statistical significance) to portfolio performance differentials (risk and return), even after taking into account the contribution of both a standard Fama-French style factor model and an alternative model that controls for macroeconomic factors.

On the basis of the risk/return metrics we have selected - Return, Sharpe ratio and Variance - our strategy provides the best results over the first two, while the contribution to the latter is mixed.

Since the proposed method is fairly new, more can be done in order to test its robustness. The validation was only done by comparing the result of training at the first period with the results on out-of-sample dates (last period). We could think of a form of cross validation, however we would need to overcome the general problem that the best reporting is at the later period. The disentangling methodology to detect the contribution of ESG and environmental indicators has been carried out by using the Fama-French and Birr models, while also a test for a "naive" portfolio can be added in the follow-up research. Moreover, an analysis of the relevance of the ESG variables (including eventually an imputation mechanism for missing information) by sector could be carried out.

Finally a deeper understanding of our model would be warranted by experimenting different methodologies in splitting and variable choice. For instance one can develop a bootstrap technique that suits the portfolio construction (bagging) and experiment how to restrict the number of variables at each split (random forest).

References

[1] State Street Global Advisors. *The ESG Data Challenge*. `https://www.ssga.com/investment-topics/environmental-social-governance/2019/03/esg-data-challenge.pdf`. "[Online; accessed 24-September-2019]". 2019.

[2] Kate Allen. *Lies, damned lies and ESG rating methodologies*. `https://ftalphaville.ft.com/2018/12/06/1544076001000/Lies--damned-lies-and-ESG-rating-methodologies/`. "[Online; accessed 19-August-2019]". 2018.

[3] Robert Arnott et al. "The surprising Alpha from Malkiel's monkey and upside-down strategies". In: *The Journal of Portfolio Management* 39.4 (2013).

[4] Rolf W Banz. "The relationship between return and market value of common stocks". In: *Journal of financial economics* 9.1 (1981), pp. 3–18.

[5] Sanjoy Basu. "Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis". In: *The journal of Finance* 32.3 (1977), pp. 663–682.

[6] Jennifer Bender et al. "A Blueprint for Integrating ESG into Equity Portfolios". In: *Journal of Investment Management* 16.1 (2018), pp. 44–58.

[7] Florian Berg, Julian Kölbel, and Roberto Rigobon. "Aggregate Confusion: The Divergence of ESG Ratings". In: *MIT Sloan Research Paper No. 5822-19* (Aug. 2019). DOI: `10.2139/ssrn.3438533`.

[8] Laxmi Chand Bhandari. "Debt/equity ratio and expected common stock returns: Empirical evidence". In: *The journal of finance* 43.2 (1988), pp. 507–528.

[9] Fischer Black. "Capital market equilibrium with restricted borrowing". In: *The journal of business* 45.3 (1972), pp. 444–455.

[10] Edwin Burmeister, Richard Roll, and Stephen A Ross. "Using macroeconomic factors to control portfolio risk". In: *A Practitioner's Guide to Factor Models* 9 (2003), pp. 1–27.

[11] Paolo Carboni. "Modelli a fattori macroeconomici per i rendimenti azionari". Final paper of the internship within the Financial Risk Management Directorate - Bank of Italy - supervisor Enrico Bernardini. 2017.

[12] Thim Chan et al. "Optimizing portfolio construction using artificial intelligence". In: 3 (Apr. 2011), pp. 168–175. DOI: `10.4156/ijact.vol3.issue3.16`.

[13] A. K. Chatterji et al. “Do ratings of firms converge? Implications for managers, investors and strategy researchers”. English. In: *Strategic Management Journal* 37.8 (2016), 1597–1614.

[14] Beiting Cheng, Ioannis Ioannou, and George Serafeim. “Corporate social responsibility and access to finance”. In: *Strategic management journal* 35.1 (2014), pp. 1–23.

[15] Gordon L Clark, Andreas Feiner, and Michael Viehs. *From the stockholder to the stakeholder: How sustainability can drive financial outperformance*. 2015.

[16] Carmine De Franco. “Performance of ESG and Machine Learning investment approaches”. In: *Ossiam* (2019).

[17] Timothy M. Doyle. *Ratings that don't Rate. The Subjective World of ESG Ratings Agencies*. English. Journal. 2018.

[18] Jeff Dunn, Shaun Fitzgibbons, and Lukasz Pomorski. “Assessing risk through environmental, social and governance exposures”. In: *AQR Capital Management* (2017).

[19] Eugene F Fama. “Efficient capital markets: a review of theory and empirical work”. In: *The journal of Finance* 25.2 (1970), pp. 383–417.

[20] Eugene F Fama and Kenneth French. “A five-factor asset pricing model”. In: *Journal of Financial Economics* 116.1 (2015), pp. 1–22.

[21] Eugene F Fama and Kenneth French. “Common risk factors in the returns on stocks and bonds”. In: *Journal of Financial Economics* 33 (1993), pp. 3–56.

[22] Eugene F Fama and Kenneth R French. “The capital asset pricing model: Theory and evidence”. In: *Journal of economic perspectives* 18.3 (2004), pp. 25–46.

[23] Eugene F Fama and James D MacBeth. “Risk, return, and equilibrium: Empirical tests”. In: *Journal of political economy* 81.3 (1973), pp. 607–636.

[24] Mingyu Fang, Ken Seng Tan, and Tony S Wirjanto. “Sustainable portfolio management under climate change”. In: *Journal of Sustainable Finance & Investment* 9.1 (2019), pp. 45–67.

[25] James Farrell. “The Dividend Discount Model: A Primer”. In: *Financial Analysts Journal* 41 (Nov. 1985), pp. 16–25.

[26] Andreas Feiner. “Machine Learning and Big Data enable a quantitative approach to ESG investing”. English. In: *Mainstreaming Sustainable Investing* 4.7 (2018).

[27] Gunnar Friede, Timo Busch, and Alexander Bassen. “ESG and financial performance: aggregated evidence from more than 2000 empirical studies”. In: *Journal of Sustainable Finance & Investment* 5.4 (2015), pp. 210–233.

[28] Mark Fulton, Bruce Kahn, and Camilla Sharples. “Sustainable Investing: Establishing Long-Term Value and Performance”. In: *SSRN Electronic Journal* (June 2012). DOI: 10.2139/ssrn.2222740.

[29] International Monetary Fund. *The Link Between Sustainable Finance and Financial Stability.* Tech. rep. 2019.

[30] Giancarlo Giudici and Matteo Bonventura. “La relazione fra rating ESG e performance di mercato: uno studio sui titoli dell’indice Stoxx Europe 600”. In: *Quaderno di ricerca Banor SIM* 1.1 (2018), pp. 1–36.

[31] Paul C Godfrey, Craig B Merrill, and Jared M Hansen. “The relationship between corporate social responsibility and shareholder value: An empirical test of the risk management hypothesis”. In: *Strategic management journal* 30.4 (2009), pp. 425–445.

[32] Global Sustainable Investment Alliance GSIA. “Global sustainable investment review 2018”. In: *http://www.gsi-alliance.org/wp-content/uploads/2019/03/GSIR Review2018.3.28.pdf* (2019), p. 29.

[33] Edward H. Bowman and Mason Haire. “A Strategic Posture Toward Corporate Social Responsibility”. In: *California Management Review* 18 (Dec. 1975), pp. 49–58. DOI: 10.2307/41164638.

[34] Oliver Hart and Luigi Zingales. “Companies Should Maximize Shareholder Welfare Not Market Value”. In: *Journal of Law, Finance, and Accounting* 2.2 (2017), pp. 247–275. URL: https://EconPapers.repec.org/RePEc:now:jnllfa:108.00000022.

[35] Roy Henriksson et al. “Integrating ESG in Portfolio Construction”. In: *The Journal of Portfolio Management* 45.4 (2019), pp. 67–81. ISSN: 0095-4918. eprint: https://jpm.pm-research.com/content/45/4/67.full.pdf.

[36] Andreas GF Hoepner. “Portfolio diversification and environmental, social or governance criteria: Must responsible investments really be poorly diversified”. In: *Available at SSRN 1599334* (2010).

[37] Ravi Jagannathan, Ashwin Ravikumar, and Marco Sammon. *Environmental, Social, and Governance Criteria: Why Investors are Paying Attention.* Tech. rep. National Bureau of Economic Research, 2017.

[38] Lars Kaiser. "ESG Integration: Value, Growth and Momentum". In: *SSRN Electronic Journal* (Jan. 2018). DOI: 10.2139/ssrn.2993843.

[39] Mozaffar Khan, George Serafeim, and Aaron Yoon. "Corporate sustainability: First evidence on materiality". In: *The accounting review* 91.6 (2016), pp. 1697–1724.

[40] John Lintner. "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets". In: *Stochastic optimization models in finance*. Elsevier, 1975, pp. 131–155.

[41] MSCI ESG RESEARCH LLC. *Foundations of ESG investing / Part 1: How ESG Affects Equity Valuation, Risk and Performance*. English. Journal. 2017.

[42] Harry Markovitz. "Portfolio selection: Efficient diversification of investments". In: *NY: John Wiley* (1959).

[43] Dimitris Melas, Zoltàn Nagy, and Padmakar Kulkarni. *Factor investing and ESG integration*. English. Journal. 2016.

[44] Vilmos Misangyi and Abhijith Acharya. "Substitutes or Complements? A Configurational Examination of Corporate Governance Mechanisms". In: *The Academy of Management Journal* 57 (Dec. 2014), pp. 1681–1705. DOI: 10.5465/amj.2012.0728.

[45] P. S. Mohanram. "Separating Winners from Losers among Low Book-to-Market Stocks using Financial Statement Analysis". In: *Review of Accounting Studies* 10 (2-3 2008), pp. 171–184.

[46] S. Mohanty, J. Pontiff, and J. Woodgate. "Share Issuance and Cross sectional Returns". In: *Journal of Finance* 68 (2 2008), 921—945.

[47] R. Novy-Marx. "The other side of value: The gross profitability premium". In: *Journal of Financial Economics* 108 (1 2013), pp. 1–28.

[48] Fabio Panetta. "The Stability of the Relation Between the Stock Market and Macroeconomic Forces". In: *Economic Notes* 31.3 (2002), pp. 417–450. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0300.00093.

[49] René Peeters. "The maximum edge biclique problem is NP-complete". In: *Discrete Applied Mathematics* 131.3 (2003), pp. 651 –654. ISSN: 0166-218X. DOI: https://doi.org/10.1016/S0166-218X(03)00333-0. URL: http://www.sciencedirect.com/science/article/pii/S0166218X03003330.

[50] Richard Priestley. "The arbitrage pricing theory, macroeconomic and financial factors, and expectations generating processes". In: *Journal of Banking & Finance* 20.5 (1996), pp. 869 –890. ISSN: 0378-4266. URL: http://www.sciencedirect.com/science/article/pii/0378426695000356.

[51] Deutsche Bank Research. *Konzept / Big data shakes up ESG investing*. English. Journal. 2018.

[52] S. R. Richardson, R. G. Sloan, and M. T. Soliman. "Separating Winners from Losers among Low Book-to-Market Stocks using Financial Statement Analysis". In: *Journal of Accounting and Economics* 39 (3 2005), pp. 437–485.

[53] Barr Rosenberg, Kenneth Reid, and Ronald Lanstein. "Persuasive evidence of market inefficiency". In: *The Journal of Portfolio Management* 11.3 (1985), pp. 9–16.

[54] Stephen Ross. "The arbitrage theory of capital asset pricing". In: *Journal of Economic Theory* 13.3 (1976), pp. 341–360. URL: https://EconPapers.repec.org/RePEc:eee:jetheo:v:13:y:1976:i:3:p:341-360.

[55] Mohanty S. "Does one model fit all in global equity markets? Some insight into market factor based strategies in enhancing alpha". In: *International Journal of Financial Economics* 24 (1 2019), 1170—1192.

[56] Dirk Schoenmaker and Willem Schramade. *Investing for Long-Term Value Creation*. CEPR Discussion Papers 13175. C.E.P.R. Discussion Papers, Sept. 2018. URL: https://ideas.repec.org/p/cpr/ceprdp/13175.html.

[57] Dirk Schoenmaker and Willem Schramade. *Investing for Long-Term Value Creation*. Tech. rep. 2018.

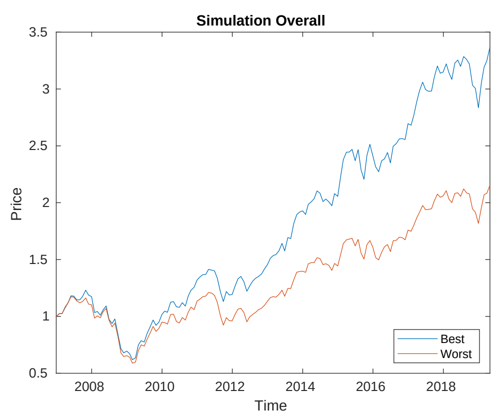
[58] William F Sharpe. "Capital asset prices: A theory of market equilibrium under conditions of risk". In: *The journal of finance* 19.3 (1964), pp. 425–442.

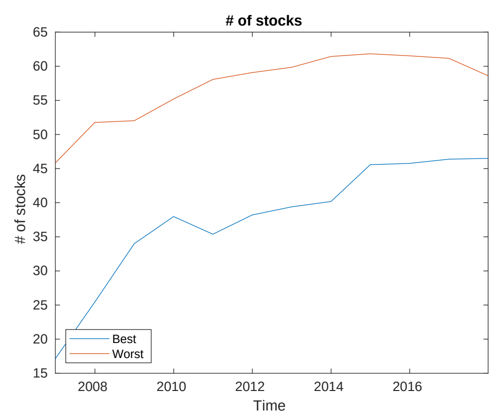
[59] Tim Verheyden, Robert G. Eccles, and Andreas Feiner. "ESG for All? The Impact of ESG Screening on Return, Risk, and Diversification". In: *Journal of Applied Corporate Finance* 28.2 (2016), pp. 47–55. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jacf.12174.

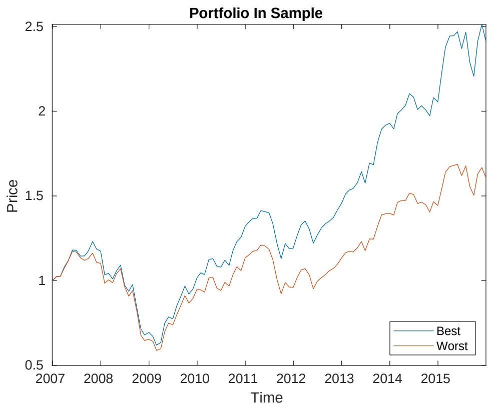
A Portfolios Obtained with ESG indicators

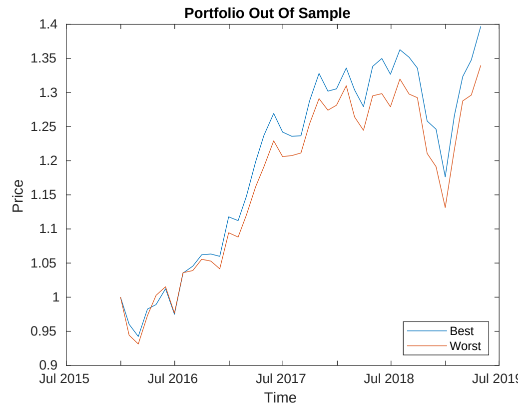
A.1 Optimizing Return, with 10 ESG variables

	VarNum	Score	LoneScore	Biv	Mater	Type
Carbon Intensity	6	0.037744	0.0048375	0	3	ENV
Waste Intensity	7	0.033685	0.0035924	0	3	ENV
Hazardous Waste Intensity	8	0.016597	-0.00024426	0	3	ENV
Eco-Design Products	35	0.014168	0.0049799	1	2	ENV
Employee Accidents	71	0.011874	0.0061546	0	3	SOC
Value - Board Structure/Specific Skills	89	0.011174	0.0036441	0	3	GOV
CONTROV_SRC_EXP_SCORE	208	0.0099531	0.0099531	0	3	SOC
Total Injury Rate	44	0.0095307	0.0060275	0	3	SOC
Senior Exec. Long-term Compensation incentives	84	0.0086402	0.0039069	0	3	GOV
Environm. Perform. Score	3	0.0083413	0.0017736	0	3	ENV
	Tr1	**Tr2**	**minB**	**minW**		
Carbon Intensity	0.3	0.8	9	6		
Waste Intensity	0.35	0.8	9	5		
Hazardous Waste Intensity	0.5	0.75	10	5		
Eco-Design Products	0.5	0.5	14	22		
Employee Accidents	0.25	0.5	5	9		
Value - Board Structure/Specific Skills	0.25	0.8	8	10		
CONTROV_SRC_EXP_SCORE	0.45	0.5	5	5		
Total Injury Rate	0.35	0.75	6	5		
Senior Exec. Long-term Compensation incentives	0.2	0.65	8	28		
Environm. Perform. Score	0.5	0.8	7	18		









In Sample

	Best	Worst	Diff
Mean	0.0094857	0.0056998	0.0037859
Var	0.0026063	0.0025737	3.2607e-05
Sharpe	0.16575	0.091828	0.073927
Draw	0.49681	0.49774	NaN

Out of Sample

	Best	Worst	Diff
Mean	0.0088811	0.0078496	0.0010316
Var	0.0010018	0.0010459	-4.4016e-05
Sharpe	0.29414	0.25558	0.03856
Draw	0.13677	0.14268	NaN

Best FF

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.0060001	0.0014944	0.010506	0.0022793	2.6325	0.0094146
Mkt-RF	0.6834	0.58366	0.78313	0.050453	13.5454	2.3484e-27
SMB	-0.077394	-0.31458	0.15979	0.11998	-0.64505	0.51994
HML	0.14594	-0.19548	0.48737	0.17272	0.845	0.39953
RMW	0.24882	-0.20718	0.70482	0.23067	1.0786	0.28258
CMA	-0.3935	-0.77894	-0.0080523	0.19498	-2.0181	0.045463

# obs.	148	Degrees of freedom	142
F-statistic	75.020088	p-value	0.000000
R^2	0.725392	R^2-Adj	0.715723
RMSE	0.024971		

Worst FF

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.0029504	-0.0009466	0.0068473	0.0019713	1.4966	0.13671
Mkt-RF	0.70281	0.61655	0.78907	0.043636	16.1061	7.5182e-34
SMB	0.034553	-0.17059	0.23969	0.10377	0.33297	0.73965
HML	0.22014	-0.075158	0.51544	0.14938	1.4737	0.14278
RMW	0.23395	-0.16044	0.62834	0.19951	1.1726	0.24291
CMA	-0.35818	-0.69155	-0.024809	0.16864	-2.1239	0.035409

# obs.	148	Degrees of freedom	142
F-statistic	109.152800	p-value	0.000000
R^2	0.793534	R^2-Adj	0.786264
RMSE	0.021598		

Best Birr

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.011103	0.00875	0.013456	0.0011902	9.3283	1.9599e-16
CF	-0.25811	-0.28134	-0.23488	0.011749	-21.9684	1.6013e-47
TH	-0.023619	-0.11864	0.0714	0.048067	-0.49137	0.62392
I	0.29727	-0.028125	0.62266	0.16461	1.8059	0.073044
BC	3.2409	2.221	4.2608	0.51592	6.2818	3.8452e-09
MT	0.8885	0.82377	0.95324	0.032747	27.1326	4.7967e-58

# obs.	148	Degrees of freedom	142
F-statistic	293.601158	p-value	0.000000
R^2	0.911802	R^2-Adj	0.908696
RMSE	0.014152		

Worst Birr

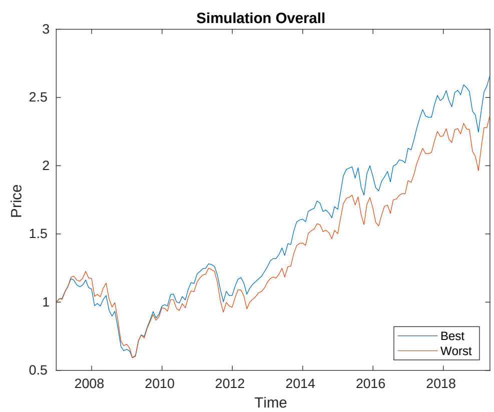
	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.00837	0.0062107	0.010529	0.0010923	7.6625	2.5962e-12
CF	-0.27824	-0.29956	-0.25693	0.010783	-25.8047	1.8078e-55
TH	-0.079704	-0.16691	0.0074985	0.044113	-1.8068	0.072907
I	0.42293	0.12431	0.72156	0.15106	2.7997	0.0058274
BC	3.0154	2.0794	3.9514	0.47348	6.3686	2.4803e-09
MT	0.84449	0.78508	0.9039	0.030053	28.1002	7.1928e-60

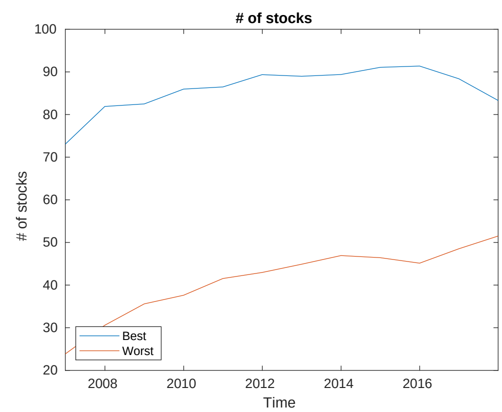
# obs.	148	Degrees of freedom	142
F-statistic	351.976287	p-value	0.000000
R^2	0.925337	R^2-Adj	0.922708
RMSE	0.012988		

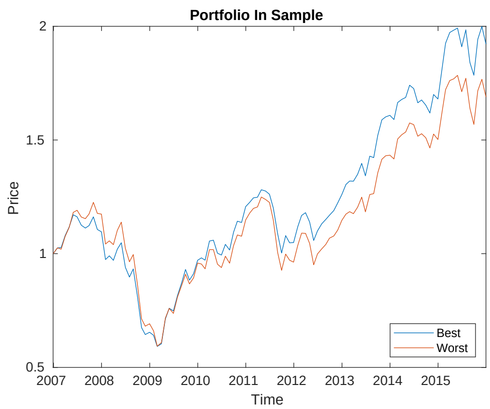
A.2 Optimizing Variance, with ESG 10 ESG variables

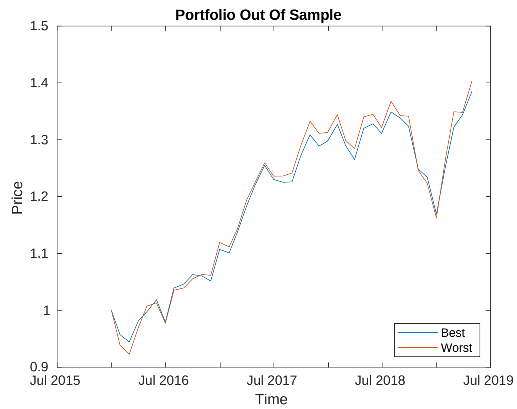
	VarNum	Score	LoneScore	Biv	Mater	Type
Waste Intensity	7	-0.013162	-0.00074149	0	3	ENV
Hazardous Waste Intensity	8	-0.012686	-0.00048242	0	3	ENV
Nuclear	28	-0.0085057	-0.00068621	1	1	ENV
Carbon Intensity	6	-0.0081091	-0.00025696	0	3	ENV
Community/Bribery, Corruption/Fraud Controv.-val	61	-0.0074764	-0.00037344	1	2	SOC
Animal Testing	32	-0.0053909	0.00062545	1	1	ENV
Energy Use/Revenues	10	-0.0035541	-0.00042974	0	3	ENV
Total Injury Rate	44	-0.0025926	0.00027167	0	3	SOC
Employee Accidents	71	-0.0023094	0.0015916	0	3	SOC
Women Employees	47	-0.0022783	-0.0010164	0	3	SOC

	Tr1	Tr2	minB	minW
Waste Intensity	0.5	0.75	12	6
Hazardous Waste Intensity	0.5	0.7	10	6
Nuclear	0.5	0.5	30	6
Carbon Intensity	0.2	0.75	6	7
Community/Bribery, Corruption/Fraud Controv.-val	0.5	0.5	27	9
Animal Testing	0.5	0.5	28	8
Energy Use/Revenues	0.2	0.5	6	13
Total Injury Rate	0.35	0.65	6	6
Employee Accidents	0.25	0.5	5	9
Women Employees	0.5	0.75	9	17









In Sample

	Best	Worst	Diff
Mean	0.0073566	0.006337	0.0010197
Var	0.0025715	0.0029187	-0.00034718
Sharpe	0.12469	0.098078	0.026612
Draw	0.49438	0.51595	NaN

Out of Sample

	Best	Worst	Diff
Mean	0.0086475	0.0091074	-0.00045993
Var	0.00096221	0.0012306	-0.0002684
Sharpe	0.29251	0.27193	0.02058
Draw	0.13333	0.15001	NaN

Best FF

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.0044687	0.00033853	0.0085989	0.0020893	2.1388	0.034159
Mkt-RF	0.68984	0.59842	0.78126	0.046248	14.9162	7.2928e-31
SMB	0.014723	-0.20269	0.23214	0.10998	0.13387	0.8937
HML	0.17107	-0.1419	0.48404	0.15832	1.0805	0.28174
RMW	0.21716	-0.20084	0.63515	0.21145	1.027	0.30617
CMA	-0.40126	-0.75458	-0.047946	0.17873	-2.2451	0.026309

# obs.	148	Degrees of freedom	142
F-statistic	92.752365	p-value	0.000000
R^2	0.765584	R^2-Adj	0.757330
RMSE	0.022890		

Worst FF

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.0037923	-0.00041137	0.007996	0.0021265	1.7834	0.076664
Mkt-RF	0.73678	0.64373	0.82983	0.047071	15.6525	1.0176e-32
SMB	-0.086901	-0.30819	0.13438	0.11194	-0.77632	0.43885
HML	0.27984	-0.038704	0.59838	0.16114	1.7366	0.084622
RMW	0.25227	-0.17316	0.67771	0.21521	1.1722	0.24308
CMA	-0.38706	-0.74666	-0.027448	0.18191	-2.1277	0.03509

# obs.	148	Degrees of freedom	142
F-statistic	106.080138	p-value	0.000000
R^2	0.788816	R^2-Adj	0.781380
RMSE	0.023297		

Best Birr

	Estimate	**CIleft**	**CIright**	**SE**	**tStat**	**pValue**
(Intercept)	0.0095652	0.007328	0.011802	0.0011317	8.4521	3.0983e-14
CF	-0.27012	-0.2922	-0.24803	0.011171	-24.1795	3.3884e-52
TH	-0.036425	-0.12677	0.053921	0.045702	-0.79699	0.42678
I	0.33382	0.024436	0.64321	0.15651	2.1329	0.034649
BC	3.0527	2.083	4.0225	0.49054	6.2232	5.1613e-09
MT	0.85279	0.79124	0.91433	0.031136	27.3891	1.5599e-58

# obs.	148	Degrees of freedom	142
F-statistic	322.195588	p-value	0.000000
R^2	0.918995	R^2-Adj	0.916143
RMSE	0.013456		

Worst Birr

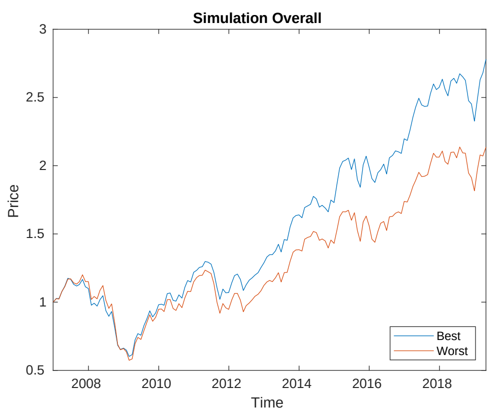
	Estimate	**CIleft**	**CIright**	**SE**	**tStat**	**pValue**
(Intercept)	0.0093749	0.0072069	0.011543	0.0010967	8.5482	1.7902e-14
CF	-0.29036	-0.31176	-0.26896	0.010826	-26.8211	1.8947e-57
TH	-0.10881	-0.19636	-0.02126	0.04429	-2.4568	0.015222
I	0.44834	0.14851	0.74816	0.15167	2.956	0.0036506
BC	3.016	2.0763	3.9557	0.47538	6.3444	2.8036e-09
MT	0.93795	0.8783	0.9976	0.030173	31.0853	3.135e-65

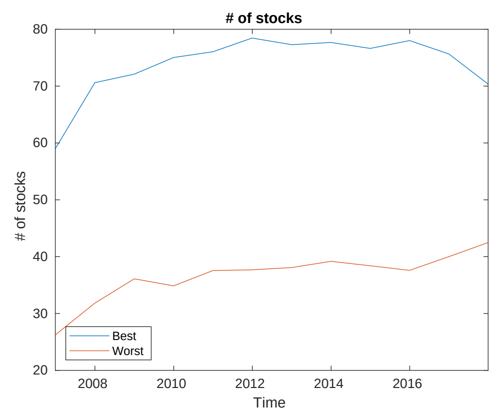
# obs.	148	Degrees of freedom	142
F-statistic	400.872565	p-value	0.000000
R^2	0.933842	R^2-Adj	0.931512
RMSE	0.013040		

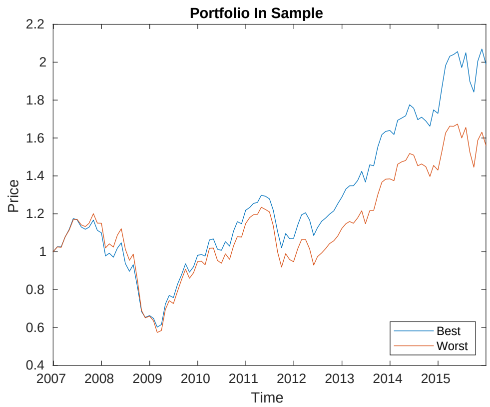
A.3 Optimizing Sharpe, with 10 ESG variables

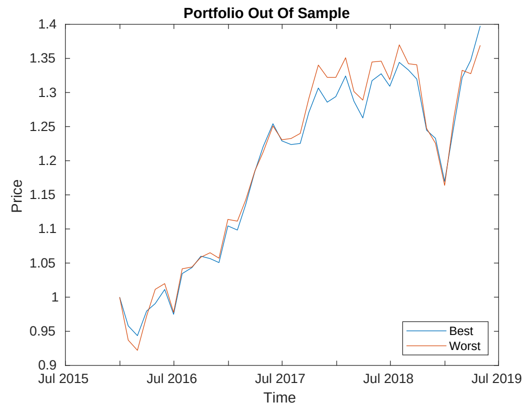
	VarNum	Score	LoneScore	Biv	Mater	Type
Carbon Intensity	6	1.0305	0.11519	0	3	ENV
Waste Intensity	7	0.85332	0.10673	0	3	ENV
Hazardous Waste Intensity	8	0.49405	0.016071	0	3	ENV
Community/Bribery, Corruption/Fraud Controv.-val	61	0.33058	0.022376	1	2	SOC
Value - Board Structure/Specific Skills	89	0.29618	0.091705	0	3	GOV
CONTROV_SRC_EXP_SCORE	208	0.29592	0.29064	0	3	SOC
Board Structure/Board Diversity	79	0.24095	0.031644	0	3	GOV
Waste Recycling Ratio	22	0.2286	0.033572	0	3	ENV
Energy Use/Revenues	10	0.22408	0.079164	0	3	ENV
Nuclear	28	0.20614	0.024375	1	1	ENV

	Tr1	Tr2	minB	minW
Carbon Intensity	0.3	0.8	9	6
Waste Intensity	0.35	0.8	9	5
Hazardous Waste Intensity	0.35	0.5	7	10
Community/Bribery, Corruption/Fraud Controv.-val	0.5	0.5	27	9
Value - Board Structure/Specific Skills	0.25	0.8	8	10
CONTROV_SRC_EXP_SCORE	0.45	0.5	5	5
Board Structure/Board Diversity	0.3	0.5	13	12
Waste Recycling Ratio	0.2	0.6	10	6
Energy Use/Revenues	0.2	0.8	6	5
Nuclear	0.5	0.5	30	6









In Sample

	Best	Worst	Diff
Mean	0.0076312	0.0056434	0.0019878
Var	0.0024835	0.0030097	-0.00052617
Sharpe	0.13241	0.083882	0.048533
Draw	0.4876	0.52209	NaN

Out of Sample

	Best	Worst	Diff
Mean	0.0088626	0.0084772	0.00038545
Var	0.00094644	0.0012177	-0.00027121
Sharpe	0.30202	0.25508	0.046937
Draw	0.12981	0.15031	NaN

Best FF

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.0046817	0.00058585	0.0087775	0.0020719	2.2596	0.025371
Mkt-RF	0.68508	0.59442	0.77575	0.045863	14.9377	6.4339e-31
SMB	-0.02061	-0.23622	0.195	0.10907	-0.18896	0.85039
HML	0.13168	-0.17868	0.44205	0.157	0.83872	0.40304
RMW	0.22057	-0.19395	0.63509	0.20969	1.0519	0.29464
CMA	-0.38755	-0.73793	-0.037169	0.17724	-2.1865	0.030416

# obs.	148	Degrees of freedom	142
F-statistic	90.880546	p-value	0.000000
R^2	0.761906	R^2-Adj	0.753522
RMSE	0.022700		

Worst FF

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.0032657	-0.00099098	0.0075224	0.0021533	1.5166	0.13159
Mkt-RF	0.7268	0.63258	0.82103	0.047664	15.2484	1.0541e-31
SMB	-0.013494	-0.23757	0.21058	0.11335	-0.11905	0.90541
HML	0.35336	0.030808	0.67592	0.16317	2.1656	0.03201
RMW	0.25342	-0.17737	0.68422	0.21793	1.1629	0.24682
CMA	-0.45751	-0.82165	-0.093365	0.18421	-2.4837	0.014167

# obs.	148	Degrees of freedom	142
F-statistic	106.115224	p-value	0.000000
R^2	0.788871	R^2-Adj	0.781437
RMSE	0.023591		

Best Birr

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.0097227	0.0075711	0.011874	0.0010884	8.9328	1.9574e-15
CF	-0.26457	-0.28581	-0.24333	0.010744	-24.6244	4.1752e-53
TH	-0.027045	-0.11394	0.059846	0.043955	-0.61529	0.53935
I	0.37127	0.073711	0.66883	0.15052	2.4665	0.014834
BC	2.9469	2.0143	3.8796	0.47179	6.2463	4.5976e-09
MT	0.84834	0.78914	0.90754	0.029946	28.3294	2.699e-60

# obs.	148	Degrees of freedom	142
F-statistic	338.584195	p-value	0.000000
R^2	0.922612	R^2-Adj	0.919888
RMSE	0.012941		

Worst Birr

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.0088298	0.0065218	0.011138	0.0011676	7.5626	4.4951e-12
CF	-0.29704	-0.31982	-0.27426	0.011525	-25.773	2.088e-55
TH	-0.11226	-0.20547	-0.019055	0.047151	-2.3809	0.018595
I	0.40859	0.089399	0.72778	0.16147	2.5305	0.012481
BC	3.3149	2.3145	4.3154	0.50609	6.5501	9.8199e-10
MT	0.92299	0.85949	0.98649	0.032123	28.7334	4.8613e-61

# obs.	148	Degrees of freedom	142
F-statistic	360.067187	p-value	0.000000
R^2	0.926892	R^2-Adj	0.924318
RMSE	0.013882		

The most material ESG indicators

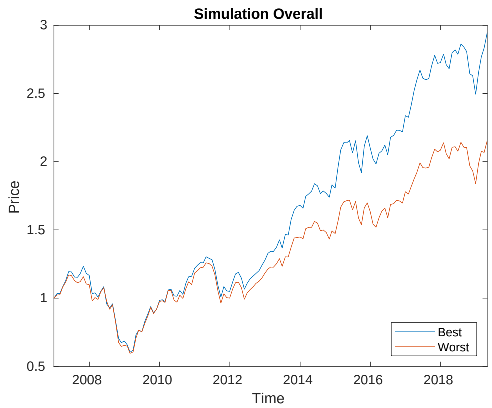
	VarNum	Return	Variance	Sharpe	Total	Biv	Type
CO2 Emissions/Revenue	6	0.037744	-0.0081091	1.0305	3	0	ENV
Waste/Revenue	7	0.033685	-0.013162	0.85332	3	0	ENV
Hazardous Waste/Revenue	8	0.016597	-0.012686	0.49405	3	0	ENV
Employee Accidents	71	0.011874	-0.0023094	0.20353	2	0	SOC
Specific Board Skills	89	0.011174	-0.00026921	0.29618	2	0	GOV
Controversial Sourcing Exposure	208	0.0099531	-0.00038647	0.29592	2	0	SOC
Total Injury Rate	44	0.0095307	-0.0025926	0.17775	2	0	SOC
Bribery, Corruption, Fraud Controversies	61	0.0082332	-0.0074764	0.33058	2	1	SOC
Nuclear	28	0.0054778	-0.0085057	0.20614	2	1	ENV
Energy Use/Revenue	10	0.0049003	-0.0035541	0.22408	2	0	ENV
Eco-Design Products	35	0.014168	0.0014373	0.12906	1	1	ENV
Long-term Compensation Incentives	84	0.0086402	-0.00011075	0.10977	1	0	GOV
Environmental Score	3	0.0083413	0.00071022	0.073592	1	0	ENV
Waste Recycling Ratio	22	0.0072689	0.00011708	0.2286	1	0	ENV
Board Diversity	79	0.0063854	-0.00026604	0.24095	1	0	GOV
Women Employees	47	0.0053944	-0.0022783	0.18377	1	0	SOC
Animal Testing	32	0.0029715	-0.0053909	0.10969	1	1	ENV

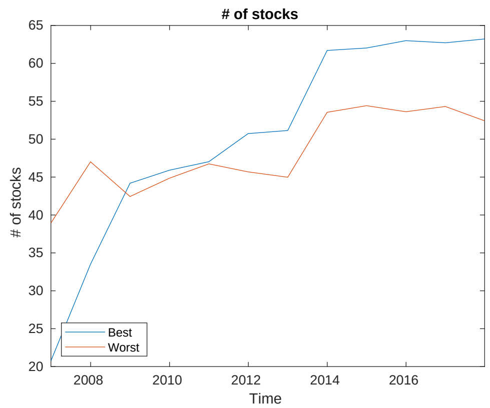
B Portfolios Obtained with Environmental indicators

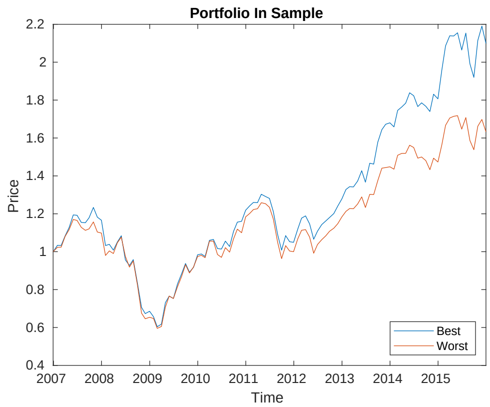
B.1 Optimizing Only Environmental indicators - Return, with 10 variables

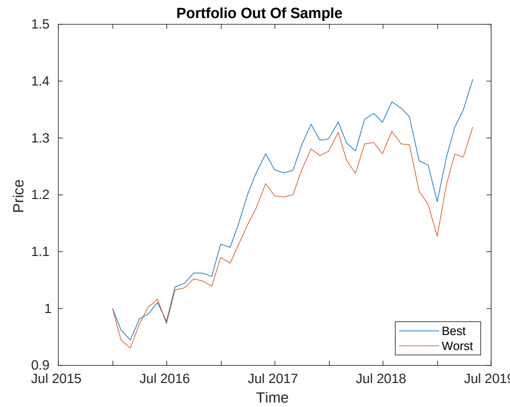
	VarNum	Score	LoneScore	Biv	Mater	Type
Waste Intensity	7	0.013808	0.0035924	0	3	ENV
Carbon Intensity	6	0.013171	0.0048375	0	3	ENV
Waste Recycling Ratio	22	0.0080097	0.0010198	0	3	ENV
Eco-Design Products	35	0.0075791	0.0049799	1	2	ENV
Environm. Perform. Score	3	0.0068444	0.0017736	0	3	ENV
Hazardous Waste Intensity	8	0.0051957	-0.00024426	0	3	ENV
Product Carbon Footprint score	144	0.0041826	0.0039622	0	3	ENV
Product Carbon Footprint Mgmt score	147	0.0038396	0.0038396	0	3	ENV
Emission Reduction Obj	20	0.0038287	0.0007293	1	2	ENV
Climate Change Theme Score	117	0.00338	0.0011116	0	3	ENV

	Tr1	Tr2	minB	minW
Waste Intensity	0.35	0.8	9	5
Carbon Intensity	0.3	0.8	9	6
Waste Recycling Ratio	0.2	0.6	10	6
Eco-Design Products	0.5	0.5	14	22
Environm. Perform. Score	0.5	0.8	7	18
Hazardous Waste Intensity	0.5	0.75	10	5
Product Carbon Footprint score	0.4	0.8	7	16
Product Carbon Footprint Mgmt score	0.2	0.8	7	8
Emission Reduction Obj	0.5	0.5	30	6
Climate Change Theme Score	0.25	0.5	102	53









In Sample

	Best	Worst	Diff
Mean	0.0082289	0.0058633	0.0023656
Var	0.0026796	0.0026064	7.3286e-05
Sharpe	0.13908	0.094467	0.04461
Draw	0.51029	0.49116	NaN

Out of Sample

	Best	Worst	Diff
Mean	0.0089502	0.0074672	0.001483
Var	0.00091414	0.0010672	-0.00015305
Sharpe	0.31025	0.24116	0.069083
Draw	0.12863	0.14075	NaN

Best FF

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.0052095	0.00084593	0.009573	0.0022074	2.36	0.019633
Mkt-RF	0.69388	0.59729	0.79047	0.048861	14.2012	4.839e-29
SMB	-0.11961	-0.34931	0.11009	0.1162	-1.0294	0.30505
HML	0.16477	-0.16588	0.49542	0.16727	0.98507	0.32626
RMW	0.22812	-0.21349	0.66973	0.2234	1.0212	0.30892
CMA	-0.37858	-0.75186	-0.0052992	0.18883	-2.0049	0.046877

# obs.	148	Degrees of freedom	142
F-statistic	83.428058	p-value	0.000000
R^2	0.746039	R^2-Adj	0.737096
RMSE	0.024183		

Worst FF

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.0028397	-0.001066	0.0067453	0.0019757	1.4373	0.15284
Mkt-RF	0.70827	0.62181	0.79472	0.043734	16.1949	4.526e-34
SMB	0.091803	-0.11379	0.2974	0.104	0.88269	0.3789
HML	0.21956	-0.076396	0.51552	0.14971	1.4665	0.14471
RMW	0.2529	-0.14238	0.64817	0.19996	1.2648	0.20803
CMA	-0.38253	-0.71664	-0.048412	0.16902	-2.2632	0.025137

# obs.	148	Degrees of freedom	142
F-statistic	110.322396	p-value	0.000000
R^2	0.795275	R^2-Adj	0.788066
RMSE	0.021646		

Best Birr

	Estimate	**CIleft**	**CIright**	**SE**	**tStat**	**pValue**
(Intercept)	0.010143	0.0080364	0.012249	0.0010656	9.5183	6.4374e-17
CF	-0.2597	-0.2805	-0.23891	0.010519	-24.6887	3.0905e-53
TH	0.0048586	-0.080213	0.08993	0.043035	0.1129	0.91027
I	0.35098	0.059652	0.64231	0.14737	2.3816	0.018564
BC	3.4485	2.5353	4.3616	0.46191	7.4657	7.6376e-12
MT	0.90471	0.84675	0.96267	0.029318	30.8582	7.7735e-65

# obs.	148	Degrees of freedom	142
F-statistic	378.988721	p-value	0.000000
R^2	0.930288	R^2-Adj	0.927833
RMSE	0.012670		

Worst Birr

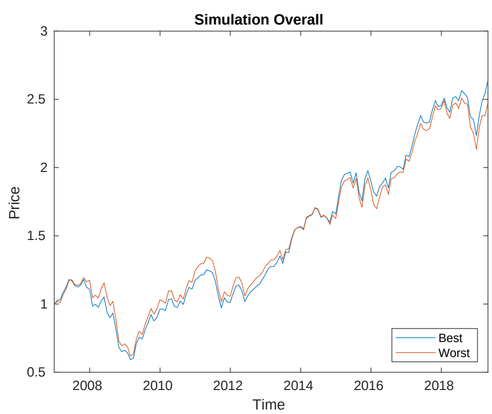
	Estimate	**CIleft**	**CIright**	**SE**	**tStat**	**pValue**
(Intercept)	0.008467	0.006114	0.01082	0.0011903	7.1135	5.1146e-11
CF	-0.28389	-0.30712	-0.26067	0.01175	-24.162	3.6803e-52
TH	-0.095696	-0.19072	-0.00067433	0.048068	-1.9908	0.048418
I	0.37125	0.045846	0.69665	0.16461	2.2553	0.025642
BC	2.9299	1.91	3.9498	0.51594	5.6788	7.3855e-08
MT	0.82879	0.76406	0.89353	0.032748	25.3085	1.7465e-54

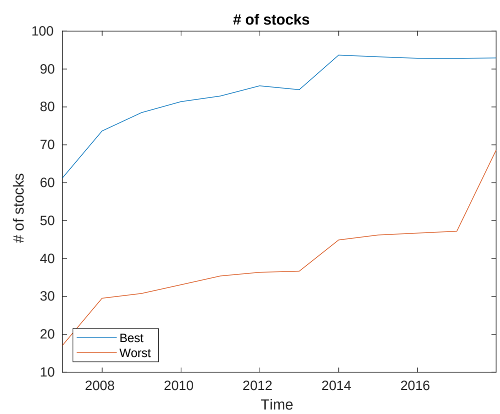
# obs.	148	Degrees of freedom	142
F-statistic	296.119807	p-value	0.000000
R^2	0.912486	R^2-Adj	0.909405
RMSE	0.014152		

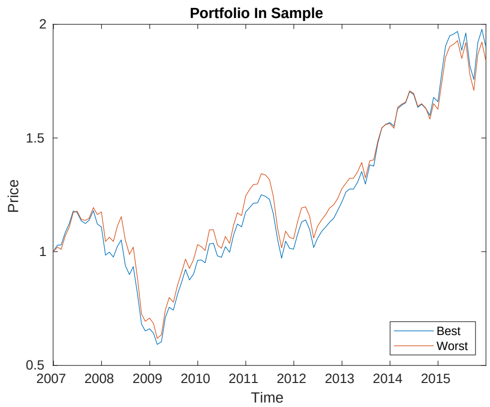
B.2 Optimizing Only Environmental indicators - Variance, with 10 variables

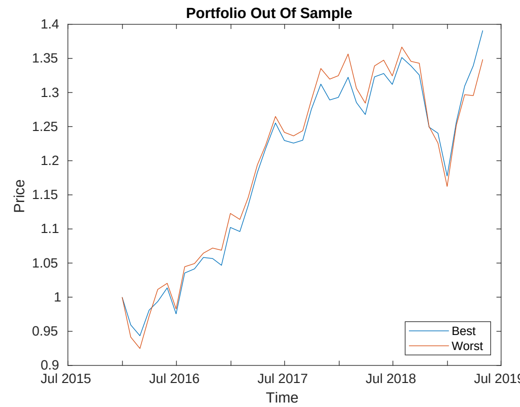
	VarNum	Score	LoneScore	Biv	Mater	Type
Hazardous Waste Intensity	8	-0.0079682	-0.00048242	0	3	ENV
Waste Intensity	7	-0.0072546	-0.00074149	0	3	ENV
Carbon Intensity	6	-0.0057402	-0.00025696	0	3	ENV
Nuclear	28	-0.0031798	-0.00068621	1	1	ENV
Energy Use/Revenues	10	-0.0021816	-0.00042974	0	3	ENV
Animal Testing	32	-0.0018501	0.00062545	1	1	ENV
Climate Change Theme Score	117	-0.0016526	-0.00015463	0	3	ENV
Emission Reduction Obj	20	-0.00071986	-0.00056938	1	2	ENV
Water Use Total/Revenues	17	-0.00064089	1.5117e-05	0	3	ENV
Opportun. Renew. Energy Exposure Score	188	-0.00047105	0.00023768	0	3	ENV

	Tr1	Tr2	minB	minW
Hazardous Waste Intensity	0.5	0.7	10	6
Waste Intensity	0.5	0.75	12	6
Carbon Intensity	0.2	0.75	6	7
Nuclear	0.5	0.5	30	6
Energy Use/Revenues	0.2	0.5	6	13
Animal Testing	0.5	0.5	28	8
Climate Change Theme Score	0.3	0.6	83	66
Emission Reduction Obj	0.5	0.5	30	6
Water Use Total/Revenues	0.5	0.5	13	12
Opportun. Renew. Energy Exposure Score	0.35	0.75	5	6









In Sample

	Best	Worst	Diff
Mean	0.0072493	0.007043	0.00020626
Var	0.0025948	0.002778	-0.0001832
Sharpe	0.12201	0.11399	0.0080234
Draw	0.49801	0.48155	NaN

Out of Sample

	Best	Worst	Diff
Mean	0.0087329	0.0080474	0.00068549
Var	0.00092484	0.0011157	-0.00019087
Sharpe	0.30121	0.25345	0.047756
Draw	0.12856	0.14945	NaN

Best FF

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.0044081	0.00031308	0.0085032	0.0020715	2.1279	0.035069
Mkt-RF	0.69262	0.60197	0.78326	0.045854	15.1047	2.4311e-31
SMB	-0.034661	-0.25023	0.1809	0.10905	-0.31785	0.75106
HML	0.18302	-0.12729	0.49333	0.15697	1.1659	0.24561
RMW	0.22848	-0.18596	0.64293	0.20965	1.0898	0.27763
CMA	-0.39218	-0.74249	-0.041866	0.17721	-2.2131	0.02849
	# obs.	148	Degrees of freedom	142		
	F-statistic	95.248432	p-value	0.000000		
	R^2	0.770317	R^2-Adj	0.762229		
	RMSE	0.022695				

Worst FF

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.0038835	-0.00016713	0.0079342	0.0020491	1.8953	0.060092
Mkt-RF	0.73001	0.64035	0.81967	0.045357	16.0946	8.0324e-34
SMB	-0.050854	-0.26408	0.16238	0.10787	-0.47146	0.63804
HML	0.22797	-0.078979	0.53491	0.15527	1.4682	0.14427
RMW	0.25713	-0.15281	0.66708	0.20738	1.2399	0.21705
CMA	-0.36147	-0.70799	-0.014952	0.17529	-2.0621	0.04102
	# obs.	148	Degrees of freedom	142		
	F-statistic	108.486873	p-value	0.000000		
	R^2	0.792529	R^2-Adj	0.785224		
	RMSE	0.022449				

Best Birr

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.0094216	0.0073657	0.011477	0.00104	9.0593	9.4031e-16
CF	-0.26873	-0.28902	-0.24843	0.010266	-26.1763	3.3682e-56
TH	-0.012827	-0.095852	0.070198	0.041999	-0.30541	0.7605
I	0.37379	0.089475	0.65811	0.14383	2.5989	0.010339
BC	3.1422	2.251	4.0333	0.4508	6.9703	1.0956e-10
MT	0.8682	0.81164	0.92476	0.028613	30.3427	6.2178e-64

# obs.	148	Degrees of freedom	142
F-statistic	388.126541	p-value	0.000000
R^2	0.931817	R^2-Adj	0.929416
RMSE	0.012365		

Worst Birr

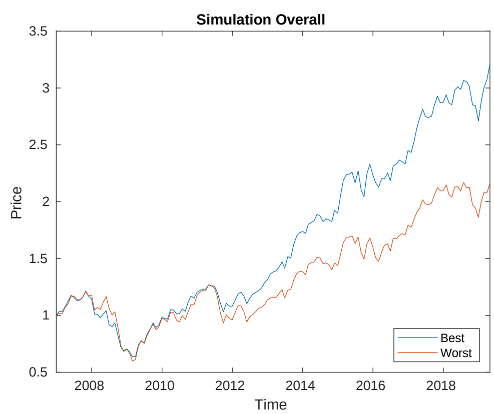
	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.0095772	0.00738	0.011774	0.0011115	8.6167	1.2091e-14
CF	-0.28397	-0.30566	-0.26228	0.010972	-25.8822	1.2715e-55
TH	-0.10972	-0.19845	-0.020993	0.044886	-2.4445	0.01573
I	0.48565	0.18179	0.78951	0.15371	3.1595	0.0019315
BC	3.0432	2.0908	3.9956	0.48178	6.3166	3.2266e-09
MT	0.8958	0.83535	0.95625	0.030579	29.2944	4.6272e-62

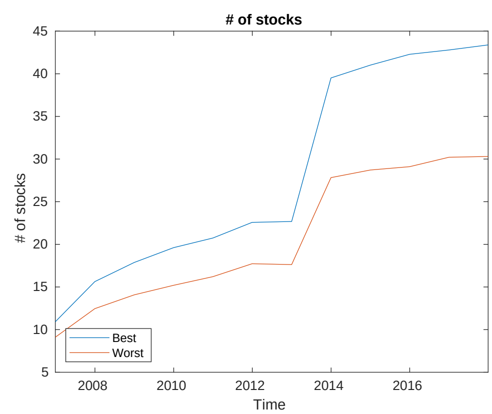
# obs.	148	Degrees of freedom	142
F-statistic	366.621588	p-value	0.000000
R^2	0.928105	R^2-Adj	0.925574
RMSE	0.013215		

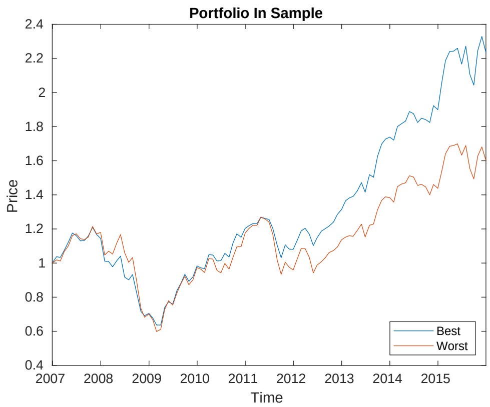
B.3 Optimizing Only Environmental indicators - Sharpe, with 10 variables

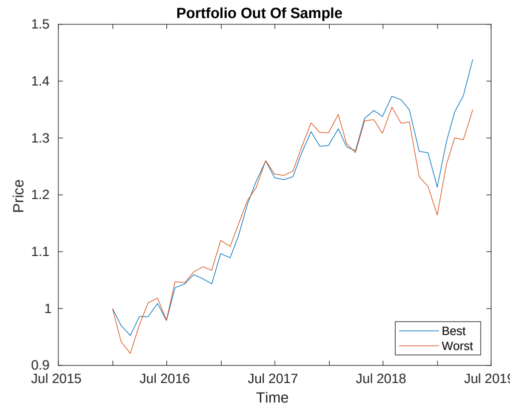
	VarNum	Score	LoneScore	Biv	Mater	Type
Waste Intensity	7	0.35684	0.10673	0	3	ENV
Carbon Intensity	6	0.35125	0.11519	0	3	ENV
Waste Recycling Ratio	22	0.2737	0.033572	0	3	ENV
Hazardous Waste Intensity	8	0.16283	0.016071	0	3	ENV
Product Carbon Footprint Mgmt score	147	0.14645	0.14645	0	3	ENV
Product Carbon Footprint score	144	0.14035	0.14035	0	3	ENV
Water Use Total/Revenues	17	0.1227	0.040102	0	3	ENV
Climate Change Theme Score	117	0.11476	0.072899	0	3	ENV
Opportunities in Clean Tech Score	178	0.11353	0.074936	0	3	ENV
Opportunities in Renewable Energy Score	186	0.11179	0.11179	0	3	ENV

	Tr1	Tr2	minB	minW
Waste Intensity	0.35	0.8	9	5
Carbon Intensity	0.3	0.8	9	6
Waste Recycling Ratio	0.2	0.6	10	6
Hazardous Waste Intensity	0.35	0.5	7	10
Product Carbon Footprint Mgmt score	0.2	0.8	7	8
Product Carbon Footprint score	0.4	0.65	14	16
Water Use Total/Revenues	0.3	0.7	8	8
Climate Change Theme Score	0.25	0.5	102	53
Opportunities in Clean Tech Score	0.2	0.5	39	17
Opportunities in Renewable Energy Score	0.25	0.65	6	5









In Sample

	Best	Worst	Diff
Mean	0.0086035	0.0058966	0.0027069
Var	0.0022809	0.0030456	-0.00076469
Sharpe	0.15863	0.087996	0.070629
Draw	0.47368	0.50718	NaN

Out of Sample

	Best	Worst	Diff
Mean	0.0095463	0.0080847	0.0014616
Var	0.00085815	0.0011446	-0.00028649
Sharpe	0.34081	0.25134	0.08947
Draw	0.11637	0.14019	NaN

Best FF

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.0055969	0.0013981	0.0097957	0.002124	2.6351	0.0093467
Mkt-RF	0.662	0.56905	0.75494	0.047016	14.0802	9.8794e-29
SMB	-0.1864	-0.40742	0.034629	0.11181	-1.6671	0.097699
HML	0.032983	-0.28519	0.35115	0.16095	0.20492	0.83792
RMW	0.20217	-0.22277	0.62711	0.21496	0.9405	0.34856
CMA	-0.26419	-0.62338	0.094994	0.1817	-1.454	0.14815

# obs.	148	Degrees of freedom	142
F-statistic	75.776911	p-value	0.000000
R^2	0.727387	R^2-Adj	0.717788
RMSE	0.023270		

Worst FF

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.0031748	-0.0010224	0.0073719	0.0021232	1.4953	0.13706
Mkt-RF	0.73853	0.64562	0.83143	0.046998	15.7141	7.1321e-33
SMB	-0.088932	-0.30987	0.13201	0.11177	-0.7957	0.42754
HML	0.33757	0.019527	0.65562	0.16089	2.0982	0.03766
RMW	0.2879	-0.13687	0.71268	0.21488	1.3398	0.18244
CMA	-0.43529	-0.79434	-0.076245	0.18163	-2.3966	0.01785

# obs.	148	Degrees of freedom	142
F-statistic	110.167796	p-value	0.000000
R^2	0.795046	R^2-Adj	0.787829
RMSE	0.023261		

Best Birr

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.010045	0.0079973	0.012092	0.0010357	9.6982	2.2331e-17
CF	-0.2354	-0.25561	-0.21518	0.010224	-23.0238	8.724e-50
TH	0.069345	-0.01334	0.15203	0.041827	1.6579	0.099547
I	0.36694	0.083786	0.65009	0.14324	2.5618	0.011457
BC	3.0388	2.1514	3.9263	0.44895	6.7688	3.1612e-10
MT	0.8532	0.79687	0.90953	0.028496	29.9413	3.1968e-63

# obs.	148	Degrees of freedom	142
F-statistic	343.578635	p-value	0.000000
R^2	0.923652	R^2-Adj	0.920963
RMSE	0.012315		

Worst Birr

	Estimate	CIleft	CIright	SE	tStat	pValue
(Intercept)	0.0089056	0.0065458	0.011266	0.0011938	7.46	7.8765e-12
CF	-0.29222	-0.31551	-0.26892	0.011784	-24.7975	1.86e-53
TH	-0.12904	-0.22434	-0.03374	0.04821	-2.6767	0.0083112
I	0.54701	0.22065	0.87337	0.16509	3.3133	0.0011699
BC	3.1233	2.1003	4.1462	0.51746	6.0358	1.3095e-08
MT	0.94212	0.8772	1.0071	0.032844	28.6847	5.9714e-61

# obs.	148	Degrees of freedom	142
F-statistic	343.750925	p-value	0.000000
R^2	0.923687	R^2-Adj	0.921000
RMSE	0.014194		

The most material Environmental indicators

	VarNum	Return	Variance	Sharpe	Total	Biv
Waste/Revenue	7	0.013808	-0.0072546	0.35684	3	0
CO2 Emissions/Revenue	6	0.013171	-0.0057402	0.35125	3	0
Hazardous Waste/Revenue	8	0.0051957	-0.0079682	0.16283	3	0
Climate Change Theme Score	117	0.00338	-0.0016526	0.11476	3	0
Waste Recycling Ratio	22	0.0080097	8.1759e-05	0.2737	2	0
Prod. Carbon Footprint Score	144	0.0041826	0.0002437	0.14035	2	0
Prod. Carbon Footprint Mgmt	147	0.0038396	0.00025081	0.14645	2	0
Emission Reduction Objectives	20	0.0038287	-0.00071986	0.071566	2	1
Water Use/Revenue	17	0.0018263	-0.00064089	0.1227	2	0
Eco-Design Products	35	0.0075791	0.0014373	0.10354	1	1
Environmental Score	3	0.0068444	0.00079796	0.083197	1	0
Energy Use/Revenue	10	0.0030538	-0.0021816	0.095028	1	0
Opportunities in Renewable Energy Score	186	0.0029098	7.8753e-05	0.11179	1	0
Nuclear	28	0.0025489	-0.0031798	0.068439	1	1
Opportunities in Clean Tech Score	178	0.0024122	0.00036662	0.11353	1	0
Opportunities in Renewable Energy Exp	188	-0.00052838	-0.00047105	-0.011381	1	0
Animal Testing	32	-0.0026548	-0.0018501	-0.077037	1	1