

Machine Learning, ESG Indicators, and Sustainable Investment



Ariel A. G. Lanza, Enrico Bernardini, and Ivan Faiella

1 Introduction

Finance can make a key contribution to the sustainability objectives embedded in the United Nations 2030 agenda, in particular by channelling resources into adaptation and mitigation measures. The integration of sustainability criteria in investment decision-making is fostered by regulators, corporate practices, and investors. This trend has accelerated during the outbreak of the Covid-19 pandemic, with inflows to sustainable investment outpacing those of the standard financial instruments (Ferriani and Natoli 2021). The COP26 held in Glasgow in 2021 recorded a widespread commitment of the private financial sector, representing globally more than USD 130 trillion, to support energy transition and the fight against climate change. The decrease in global carbon emissions due to the Covid outbreak and the shift in renewable energy development (Adebayo et al. 2022) was short-lived. More efforts and capital are needed to mitigate environmental degradation and accelerate the energy transition (Fareed et al. 2022). The Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) has highlighted the need for urgent action to tackle the already apparent consequences of climate-related acute and chronic events, by fostering investments in mitigation and adaptation measures (IPCC 2022).

According to Global Sustainable Investment Alliance, the global assets managed with sustainability criteria have increased to USD 35 trillion at the beginning of

A. A. G. Lanza (✉)

Kellogg School of Management, Northwestern University, Evanston, IL, USA

e-mail: ariel.lanza@kellogg.northwestern.edu

E. Bernardini · I. Faiella

Bank of Italy, Climate Change and Sustainability Hub, Rome, Italy

e-mail: enrico.bernardini@bancaditalia.it; ivan.faiella@bancaditalia.it

2021, almost double than in 2016, ranging from traditional instruments to new assets such as green bonds. This market trend is also driven by the search for long-term investments with less volatile risk-return profiles. An extensive literature shows that sustainable investment leads in most of the cases to risk-adjusted market returns that are often higher than those achieved using traditional financial models (Atz et al. 2022; Friede et al. 2015).

The importance of the environmental, social, and governance (ESG) profiles has been underlined since the 2004 UN Global Compact report ‘Who Cares Wins’ (Global Compact 2004). The integration of ESG principles into corporate management can innovate business practices and provide firms with a competitive edge. It contributes to reducing operating, legal and reputational risks; it leads to a more efficient allocation of resources, which can be shifted from risk management to productive activities, and a more motivated workforce. This favours in turn a better operational and market performance, thus lowering the cost of capital.

ESG scores have become popular among investors as a tool for setting sustainable investment strategies and selecting instruments and market indices in the equity and bond space. For this reason, scores are very important in driving the choices of market participants. However, the assessment of ESG practices embedded in these scores raises some concerns. ESG scores are computed using the information provided by private firms using heterogeneous methods. In particular, the representation of each ESG pillar has different levels of complexity, with the E component being usually less heterogeneous and controversial owing to the greater availability of quantitative data and conceptual models. Furthermore, there are neither broadly accepted rules for ESG data disclosure by individual firms nor auditing standards for the verification of the reported data. ESG score providers rely heavily on voluntary disclosure by firms and on proprietary methodologies to select, assess, and weigh individual ESG indicators. As a result, ESG scores of individual firms show a large heterogeneity across agencies compared, for example, with credit ratings. There is also evidence of significant biases in ESG scores, which tend to overestimate the score of companies that are larger and belong to specific industrial sectors and geographic regions.

This chapter investigates the sensitivity of stock returns to ESG information. We propose to (partially) overcome the current inconsistencies and fill the gaps in the ESG scores by using Machine Learning (ML) techniques to spot the most significant E, S, and G indicators that better contribute to the construction of efficient portfolios. ML does not need a model-based methodology, unlike portfolio theory. Our strategy applies ML techniques using over 220 ESG indicators from two of the largest data providers, Refinitiv-Asset 4 and MSCI ESG Research, for around 250 listed companies in the euro area in the period from 2007 to 2019, and sheds light on the main ESG indicators associated with risk and return differentials. The novelty of this study is threefold: (a) we analyze a very large array of ESG indicators; (b) we employ a model-free ML methodology; and (c) we disentangle the additional contribution of ESG indicators to portfolio performance, beyond the traditional style, and macroeconomic factors.

The study shows that a European equity market investor who had developed the proposed ML technique in 2016 and applied it using the ESG indicators in the period from January 2017 to April 2019 would have achieved an average annualized extra

return between 0.5 and 1.2 percentage points (depending on the different risk/return objectives), compared with the Eurostoxx index. Applying ML techniques to the environmental indicators only, the extra return would have been between 0.8 and 1.8 percentage points.

Even taking into account the contribution of standard Fama-French (FF) (2015) style factors and, alternatively, of macroeconomic factors, the information content extracted from ESG indicators with ML significantly contributes, economically and statistically, to portfolio performance.

The rest of the chapter is organized as follows. In Sect. 2, we review the literature on equity returns, introduce the notion of ESG investing and some key evidence, discuss the current ESG data gaps and present some ML applications for investment purposes. Section 3 describes our data set (index constituents and return time series) and ESG indicators, with a focus on the treatment of missing data. In Sect. 4, we present the setting of the ML technique together with the framework for portfolio construction. Section 5 shows the results and presents a set of robustness checks. Section 6 concludes and discusses possible avenues for future research.

2 Literature Review

This section deals with the juncture 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. Two of the most important studies for our work are those by Fama and MacBeth (1973) and Burmeister et al. (2003). ESG data have become prominent in sustainable investment decision-making, although there is no uniform definition of sustainability. According to Meuer et al. (2019), there are over 33 definitions of corporate sustainability. ESG data can be generally defined as every information and indicator of environmental, social, and governance profiles related to corporate operations. ESG scores have become popular sustainability indicators among financial professionals. Based on information obtained from publicly available documents, questionnaires, data or news archives, and other sources, some private-sector data providers have developed ESG scores of firms relating to areas not strictly connected to their core business. By aggregating these elements, weighted according to different criteria to obtain a single final score, the providers sell valuations in two areas: (1) the firm's ability to deal with risks stemming from these three dimensions, e.g. market risks arising from climate regulation, risk of litigation with consumers or of penalties for illegal conduct, reputational risks, etc.; (2) the firm's capacity to seize new opportunities, in terms of innovation and efficiency in its processes and of competitiveness of its products, through sound practices, like internalizing negative environmental externalities with low levels of waste or having a high share of women in managerial positions.

Some studies show the effectiveness of ML techniques in filling the sustainable data gap, such as Nguyen et al. (2021). Other studies perform textual analysis of the ESG investing literature as Kumar et al. (2022). To the best of our knowledge, the possibility of combining ESG data with ML techniques for portfolio construction seems unexplored. A study by Feiner (2018) considers that such a link might exist and focuses on the effectiveness of ML in retrieving ESG information. In applying ML techniques, we look inside the ESG scores and try to enhance the understanding of the materiality of the individual ESG raw indicators for investment purposes. We employ decision trees, which are simply framed and easy to interpret in economic terms.

2.1 Risk Factors for Equity Returns

The first factor model relies on macroeconomic variables and was originally proposed by Burmeister et al. (2003) (hereafter BIRR) for the US equity market. We apply the model to the euro area market as proposed by Carboni (2017). The second-factor model is based on financial variables and is inspired by Fama-French (1993). The two models are derived from the general Asset Pricing Theory model by Ross (1976), according to the following equation:

$$r_i(t) - R_{rf}(t) = \beta_{i,1}[P_1 + f_1(t)] + \dots + \beta_{i,k}[P_k + f_k(t)] + \varepsilon_i(t) \quad (1)$$

where the return of security i in excess of the risk-free rate R_{rf} in period t is explained by several factors $f_k(t)$ to which the security is exposed through the factor coefficients, β_i , with ε_i as an idiosyncratic error term.

The models are described below. They help disentangle the contribution of the ESG variables, and check whether their role is not already captured by macro or financial factors identified by literature.

The BIRR model considers changes in fundamental economic variables such as investor confidence, interest rates, inflation, real business activity, and a market index as in the CAPM. Burmeister et al. (2003) suggest the adoption of the risk factors shown in Table 1.

In the FF five-factor model, the firm's profitability and cash flows may have a material effect on stock returns, as in Gordon's model (Farrell 1985). Other factors that may generate outperformance are profitability (as in Novy-Marx 2013), share buy-backs (Mohanty et al. 2008), and growth (Mohanram 2008). Furthermore, small companies are generally less liquid and riskier than big ones (size effect), and companies with a high book-to-market price ratio generally outperform companies with a low ratio (value effect).

The FF five-factor model for the present analysis employs the following equation for the excess return (the time reference is omitted for simplicity):

Table 1 Risk factors in the Birr model

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 (20-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 the 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 of the market portfolio which is not explained by the other risks and the intercept	Change rate between the value of regressed index at the beginning and at the end of the month

Source: Burmeister et al. (2003)

$$R_i - R_{rf} = a_i + b_i(R_{mkt} - R_{rf}) + s_i\text{SMB} + h_i\text{HML} + r_i\text{RMW} + c_i\text{CMA} + \varepsilon_i \quad (2)$$

in which R_i is the asset return, R_{rf} is the risk-free rate, a_i is the excess return over the benchmark, b_i is the market factor loading (exposure to market risk, different from the CAPM beta), R_{mkt} is the market return, s_i is the size factor loading (the level of exposure to size risk, SMB), h_i is the value factor loading (the level of exposure to value risk, HML), r_i is the profitability (RMW) factor loading, and c_i is the investment (CMA) factor loading (Mohanty 2019).

2.2 Sustainable Investment: Foundations and Issues

The investors' interest in Socially Responsible Investing (SRI) is a recent phenomenon and is growing fast. According to the Global Sustainable Investment Alliance (GSIA 2020), since 2016 sustainable investment has almost doubled and it has reached USD 35 trillion at the beginning of 2021 (around 36 per cent of professionally managed funds), one-third of which is located in Europe.

The rationale for the positive impact of ESG profiles on stock return is that a sustainable company will face less risk related to environmental issues, regulation, or lawsuits and can benefit more from the opportunities stemming from good ESG practices. Some studies find that the companies that adopt sustainable production methods are generally on the frontier of productive efficiency and benefit from a competitive advantage, e.g. from process/product innovation and customer satisfaction, with a lower exposure to operational, reputational and legal risks. These companies achieve a lower cost of capital; they get higher valuation assigned by the investors which translates into superior market performance (Clark et al. 2015).

ESG scores are widely used in sustainable finance for selecting financial instruments, building investment portfolios, creating market indices, and reporting (Bernardini et al. 2021a, b). The growing use of ESG scores goes together with a high heterogeneity among the scores computed by different providers for the same company. This phenomenon depends primarily on the different viewpoints of the providers as concerns the risk exposure to and risk management of the sustainability factors. Besides, the divergence stems from different procedures for data collection and selection of ESG indicators, as well as different assessment methodologies. Overall, this leads to some confusion (Berg et al. 2022).

Sustainability data have been studied in the literature from many angles, including, but not limited to, risk and return. Cheng et al. (2014) show that firms that score well in Corporate Social Responsibility (CSR) parameters have better access to finance at a lower cost. As concerns risk management, Godfrey et al. (2009) show that there is an insurance-like property of CSR activity in case of negative events such as legal/regulatory actions.

Integrating sustainability issues into portfolio management is a complex matter even from a theoretical point of view. As pointed out by Hoepner (2010), initially researchers viewed sustainability as a purely ethical choice, leaving aside any link with the traditional risk-return framework. According to this view, responsible investment is limited to *screening* the securities in the portfolio; at best this would lead to a portfolio as efficient as the unscreened one, since adding constraints to a portfolio optimization problem can never improve diversification and investment choices (Fama 1970). Although the previous general principle has been considered for many years as the ‘inescapable conclusion’, more recently Arnott (2013) has shown that a series of equally weighted random portfolios of sample stocks taken from a benchmark outperform the same cap-weighted benchmark over 40 years. This leads to the consideration that the reduced universe portfolios have to carefully adapt the weighting scheme for risk- and return-based factors. For practical purposes, there is a tipping point in the threshold of the sustainability filter beyond which the constraint is too strong and can significantly reduce the investment universe, with a negative impact on diversification and performance.

Two further considerations are in order. As argued by Hoepner (2010), the risk reduction due to diversification can be decomposed into three elements: the number of securities, their correlation, and their specific risk. If a good ESG score is associated with lower specific risk and this component offsets the negative effect of screening on the first two elements, it is possible to avoid the ‘inescapable conclusion’. Sustainability should then be considered in a risk-return framework. Some empirical results are provided by Verheyden et al. (2016).

As pointed out by Schoenmaker and Schramade (2018), a substantial limitation of traditional analysis with the risk-return framework is that it involves mainly time-series analysis, which is *backward-looking*. Sustainability assessment is inherently *forward-looking*, partly owing to its long-term perspective. This criticism is compatible with the hypotheses of adaptive markets, incomplete information, and not completely rational behaviour.

Other approaches to sustainable investing have been put forward recently. For example, under *impact investing* the investor not only seeks a financial objective, but he also aims at a social or environmental impact. This choice should not be considered superficially. A growing literature argues that corporations should have a broader objective than simple profit maximization. Hart and Zingales (2017) argue that it is often too narrow to identify shareholder welfare with market value and that ‘money-making and ethical activities are often inseparable’ therefore ‘companies should maximize shareholder welfare not market value’. An enlightening example is about the shareholders of a company selling high-capacity gun magazines. If the shareholders are concerned about mass killings, it would be more efficient for them to ban the sales of ammunition rather than reinvest the profits made by the company in gun control. This principle explains 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, which consists in making investment decisions that include ESG factors within the traditional financial modelling framework: ESG indicators are thus treated like other financial indicators to explain risk and return.

Although the literature on the effect of ESG factors on returns is not unanimous, research conducted by Khan et al. (2016) shows that firms with a fair rating on sustainability issues tend to outperform firms with poor ratings.¹ Giudici and Bonventura (2018) conduct a similar study for the European market and show that firms with better practices in all of the three ESG pillars exhibit higher returns; strategies that combine the ESG tilt with fundamental indicators, like the price-earning ratio, seem more efficient.²

A review of this vast literature is beyond the scope of this chapter. We just recall the two meta-analyses published by Friede et al. (2015), reviewing over 2000 studies and by Atz et al. (2022), reviewing over 1000 studies from 2015 and 2020. The latter finds a positive relationship for 58 per cent of the studies on the corporate performance (proxied by ROE, ROA, and stock return), and 59 per cent of the studies on the investment performance (measured by alpha and Sharpe ratio).

¹Unfortunately 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 time series 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 have been originally designed for the US firms and it might be arguable to squarely apply them to European firms.

²This 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.

2.3 ESG: The Silver Bullet for Sustainable Investment?

While initial research on corporate social responsibility dates back to the 1970s (e.g. Bowman and Haire 1975), the ESG acronym was introduced in 2005. Only recently has ESG reporting become regular and granular, such as to allow statistical analysis at firm level. The ESG approach has the desirable property of providing the investor with a score, or a rating, that factors in a large amount of information about how a firm performs along several sustainability dimensions. Integrating ESG factors into equity investments is becoming a common responsible investment practice and there is a general agreement on its benefits. But how reliable is the information content of ESG scores? In a provocative article, Allen (2018) expresses doubts on the investors' awareness of the information they are employing, creating a false sense of confidence on ESG figures. The IMF (2019) expresses concern regarding the quality and consistency of the information in ESG scores and calls for a standardization of terminology and definitions.

The lack of generally agreed methodologies in compiling ESG data and of auditing standards to verify what is reported by the firm is a pressing concern for the quality of ESG information. Besides, ESG score providers rely on voluntary disclosure by firms, which they complement with their own estimates. 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 ratings show a rather low correlation, between 0.4 and 0.7 (Chatterji et al. 2016; Table 2). This is in sharp contrast with the high correlation among credit ratings, which is above 0.9.

There is also evidence of possible biases in ESG scores, which tend to give prominence to companies that have a larger size and belong to specific industrial sectors and geographic regions (Doyle 2018). Most of the disagreement is due to different measurement techniques; a different weight of the individual E, S, and G components also plays a part, together with the a priori bias of the rating companies (Berg et al. 2022). There is clearly a gap between ESG indicators and other standard accounting variables that follow well-established principles (e.g. GAAP) and lead to lower variability between accounting data providers. With our innovative technique, we try to overcome these problems, thus providing a useful tool for decision-making.

With all the above caveats, ESG scores are key to designing a portfolio that factors in the sustainable practices of the firms. ESG scores contain a wealth of data

Table 2 ESG score providers' cross-correlations

	Sustainalytics	MSCI	Robeco SAM	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)

Table 3 ESG score cross-correlations

	Robeco SAM	Sustainalytics	Refinitiv-A4
Euroarea-exItaly			
MSCI	0.42	0.46	0.32
RobecoSAM		0.58	0.56
Sustainalytics			0.41
Italy			
MSCI	0.54	0.54	0.60
RobecoSAM		0.67	0.53
Sustainalytics			0.56

Source: Own calculations on ESG scores

that can complement the investors' information and play a role in shaping a thorough asset pricing on the markets.

Burmeister et al. (2003) warn against using accounting data for reasons that can also partially apply to ESG data. Our data samples are large enough for regressing each sector separately, choosing indicators for each sector according to its business peculiarities. Thanks to the continuous improvement of data feeds, we can overcome the largest differences among reports of different companies.

After checking that we have a similar low correlation issue in our data (Table 3), we devise a strategy that applies ML techniques to the raw ESG data to set up a heuristic selection process and create sample portfolios on the basis of their financial and sustainability performance.

2.4 Machine Learning in Finance

Even if the use of ML on ESG data for portfolio choice is little explored, it is sometimes used for text mining, e.g. by Feiner (2018) as previously recalled, and by Kumar et al. (2022). ML has become popular in recent years. One can find instances in which Machine Learning techniques are mentioned with regard to sustainable finance (Allen et al. 2017) or applied to ESG indicators for investment purposes (Erhardt 2020) or to ESG scoring (Sokolov et al. 2021), although there is not always a transparent specification of the methods (De Franco 2019).

The application of ML to portfolio choices is a wide field (see for example Chan et al. 2011). In the development of our model, we face some general issues. The first one is that we would like its results to be easily interpretable. If we have a strong a priori belief that sustainable investing will lead to better results in the long term, we cannot rely on a model which might suggest to invest in 'unsustainable' firms. Second, while many applications of ML employ high-frequency data and have a short-term use, we have a long-term orientation.

3 Data

The data for the analysis are time series at the company level on stock returns and ESG indicators. For both data types (returns and ESG data), the first step is the treatment of missing values. Below we explain the techniques to overcome this issue.

3.1 Returns and Indices

The sample is composed of the stocks in the EURO STOXX 300 index, which tracks the top 300 stocks in the euro area by capitalization. From the constituent stocks, we exclude the companies of the financial sector due to their business model, which differentiates them from non-financial firms. We first use the monthly total return of each stock starting from 31 December 2000 to 30 April 2019.

The sample includes the stocks in the index as of 31 December 2010. This choice requires some caution. Let us hypothesize for a moment to start the analysis on 31 December 2000, using the stocks in the index on the last date, 30 April 2019. A comparison of the cap-weighted index with the equal-weighted index reveals that the latter outperforms the cap-weighted index by 30 percent (Fig. 1).

This is the result of the well-known *survivorship bias*, because we are picking stocks based on information that is only available ex-post. Knowing that a stock is going to enter the index of the top 300 companies by capitalization in future years implies that its price will grow more than the price of the stocks which are currently in the index. Besides, we do not need to select the sample as of the end of 2000, since the reporting of ESG data was absent on that date. We use the sample as of the end of 2010. Figure 1 (right) shows that from 31 December 2010 onwards the equally weighted and cap-weighted portfolios do not show a significant return difference.

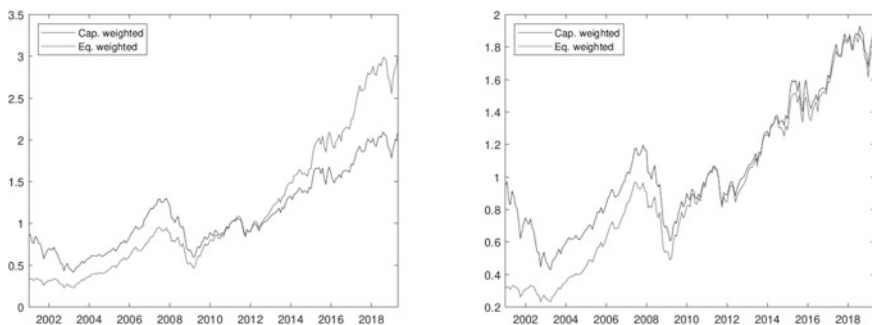


Fig. 1 We compare the return of the equal-weighted index with that of the index weighted by capitalization. On the left panel, the sample of stocks is chosen on the final date; on the right panel, the sample is chosen on 31 December 2010. The index value is normalized to 1 as of 31 December 2010. The data are those from EURO STOXX 300

We thus decided to use the 252 stocks that were in the index at the beginning and at the end of the period. We employ the time series from 31 December 2006 to 30 April 2019, i.e. 125 observations.

3.2 *ESG Data*

Refinitiv-Asset 4

Refinitiv has expanded its offer of financial data with ESG ratings since 2009 with the acquisition of the Swiss provider Asset4, devoted to environmental, social, and governance data. After the acquisition, Asset4's ESG rating methodology was revised and improved. The Refinitiv ESG team of 165 analysts covers about 1700 companies in Europe, and its ESG time series start from 2002. For each company, two numerical scores are drawn up, the 'ESG score' and the 'ESG combined score'; for both a literal rating is also provided. The ESG score measures the performance, commitment, and effectiveness demonstrated by companies regarding the environmental, social, and governance dimensions. The ESG combined score complements the ESG score with the assessment of companies' controversies on ESG issues. This framework divides the three pillars E-S-G into ten categories, each of which is evaluated through a variable number of indicators based on the industry to which they belong to, and selected from a set of 178 indicators. To this end, the 54 industry groups of the Thomson Reuters Business Classification (TRBC) are used as reference. In our study, after the initial selection of 100 distinct reported ESG variables (such as the E, S, and G scores, the level of carbon emissions, the number of accidents that occurred to employees, etc.) available for our investment sample of 252 companies, we added some economic variables (such as revenues, EBITDA, employees, etc.). We observe that some fields are missing (reported as 'Not a Number' or NaN) for some dates. After some data cleansing, we are left with 105 variables to explore.

We decided to modify some variables to compare different companies on a fair ground. 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 normalized using firm revenue. The injury rate, employee accidents, employees leaving, and training costs were normalized by the number of employees. Contractor accidents were normalized by the number of internal employee accidents.

MSCI

The other data provider is MSCI ESG Research, which produces 172 ESG variables. MSCI ESG Research is a subsidiary of MSCI Inc., created in 2010 after the

acquisition of RiskMetrics Group and the reorganization of the companies Innovest and KLD, both devoted to ESG research. MSCI ESG Research is organized with a team of around 185 analysts covering approximately 1500 companies in Europe. The ESG rating time series covers 20 years. MSCI ESG Research is currently the largest ESG rating provider; its analysis is 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 key issues on the environmental, social, and governance profiles and the ability to manage these issues. The rating is expressive of the company's ESG profile in comparative terms, as it results from the comparison of the scores of firms operating in the same industry. The MSCI framework divides the three E–S–G pillars into ten themes; in turn, these are 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 regular and are not the same for every stock. As in the case of Refinitiv, a score for the E, S, and G components is also provided. The other variables are defined as ‘key issues’ (for example, raw material sourcing, product carbon footprint, etc.). Key issues have an overall score which is obtained by 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 decided to exclude the weight of the key issues in our evaluation and we only employ the three scores and the key issues for a total number of 112 ESG indicators.

3.3 First Trials with Standard Approaches

The first plain-vanilla ML approach was not very promising because of missing data. Standard approaches work with full rectangular matrices of factors. Because of changes and improvements in methodologies and reporting, our matrices lack several fields. When dealing with missing values, we should be careful in trying to understand the reason for the absence. Usually, it is either because a reported variable does not apply to the sector under consideration, or because the firm has not disclosed relevant information. We often observe that many firms in the same sector have similar missing variables. In the case of a firm not reporting the relevant information, the reason might be that the firm does not have the necessary resources to disclose, even in the cases in which the information would be ‘good’. Another reason could be that the firm prefers to provide no news rather than bad news. Against these possible explanations, we have chosen to delete missing information rather than filling NaNs with some value as is often done in previous empirical studies (filling with zeros, extending the last available observation, and using the

sector average or the overall average).³ This choice implies that with standard approaches, to obtain a rectangular matrix without missing data, we will have to discard some pieces of information that are available to us.

To obtain a fully rectangular matrix, we start from the available data, and whenever we get a NaN, we either delete its row (time observations) or column (ESG indicator) until the submatrix that is left contains no missing value. The problem of excluding as few available data as possible is not trivial. As shown by Peeters (2003), it can be reduced to the maximum edge biclique problem, which is NP-complete.

We used the MATLAB built-in regression learner to try several alternative regressions. Our dataset is the result of the heuristic selection applied to the full $56,134 \times 96$ original regression matrix (given by the combination of securities, dates, and indicators). To select fewer rows, we eliminate a row if its NaN ratio was greater than the NaN ratio of each column at the power of 0.1. The selection left us with 41 variables and 2841 observations. After the selection, a constant column was added, as well 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, yielding 45 variables in total. To estimate the goodness of fit we considered the RMSE on an eight-fold validation, where an RMSE of 0.35054 is obtained using only the constant value. The best RMSE (0.2817) was reached in the regression with bagged trees with the single variable sector return, which was by far the best explanatory variable. The same method with all the variables gave a slightly worse RMSE (0.29615).

The fact that these initial results were not promising does not imply that the data has no explanatory power, that is ‘absence of evidence is not evidence of absence’. We suspected that several aspects might have negatively impacted these preliminary results. First of all, some data was lost in the construction of the rectangular matrices. In addition, any regression analysis affects only indirectly the portfolio choice and thus it might not capture 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 indicators, like the Sharpe ratio, variance, and mean return. This led us to develop a specific ML method.

4 A Tailored Machine Learning Approach

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

³Henriksson et al. (2019) carry out an interesting analysis aimed at finding the ESG exposure for a company that does not report ESG information; however, the results could hardly apply at granular level.

4.1 *The Proposed Approach*

A standard practice in the literature consists in creating portfolios where stocks are equally weighted and selected according to the ESG scores of the providers, and portfolios are rebalanced annually. This allows 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 stocks into ‘best’ and ‘worst’ performers where ‘best’ and ‘worst’ refer, respectively, to the top and the bottom quartile of the ESG score distribution. We found that the aggregate ESG scores computed by the data providers systematically led to lower returns for the most ESG-compliant companies. This happened also when we separately considered the ‘Environmental’, ‘Social’, and ‘Governance’ variables instead of considering the aggregate ESG variable. However, the same experiment done with single ESG variables (e.g. CO₂ emissions divided by revenue), yielded opposite results, i.e. the portfolio of the less polluting companies performed better than the portfolio of the most polluting ones.

To keep the model simple and informative, we stick to the equally weighted portfolios. We notice that a more flexible choice of the thresholds (rather than the standard quartile choice used in other studies) could lead to slightly different results. For example, a particular choice of thresholds could lead to a group of highest-scoring companies on the Refinitiv Environmental score performing better than a group of lowest-scoring companies, even though the choice of the quartile is showing the opposite situation. We set out to automatically find those thresholds to obtain the highest possible performance for the ESG-compliant companies. We note that, although this choice could increase the risk of false positives, it could be the only way to appreciate the information embedded in ‘weaker factors’ (according to the standard quartile method). This approach is fundamentally different from selecting the threshold subjectively. By automatically selecting the best ones, we put all our ESG variables on the same level playing field.

4.2 *Tree-Based Approach, the General Idea*

Our ML approach for portfolio construction has two steps: (1) we use an optimized algorithm to select the ten most meaningful ESG indicators in three types of trials, for different financial objectives; (2) we combine those indicators to select and weight stocks to construct portfolios, which are tested afterwards.

To systematically find the most significant ESG indicators that could provide portfolio extra performance, we check for the indicators that can help towards stock selection aimed at maximizing the best–minus–worst (BmW) differential in terms of three financial indicators on a 12-month horizon, namely:

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

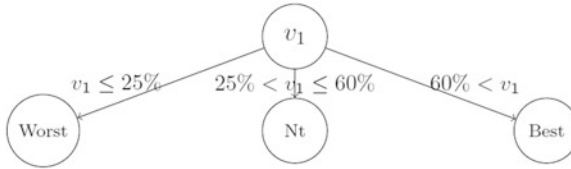


Fig. 2 The first split of decision tree. The lower threshold is 25 per cent, meaning that all the stocks that have a score (given by the variable v_1) that falls in the lower quartile are assigned to the ‘worst’ portfolio. While the stocks with a score in the top 40 per cent are assigned to the ‘best’ portfolio

From our initial trials, a tree-like structure arises naturally as one of the best ways to automate our research and keep the model as simple as possible, allowing the decision-maker to understand the economic meaning of the results. This addresses one of the greatest concerns about ML solutions, which is the lack of interpretability of the results.⁴ Our idea consists in building trees by setting thresholds that aim at the optimization of a variable that is not the RMSE, but a portfolio financial variable. Specifically, we maximize (minimize) the mean absolute return and the Sharpe ratio (the variance).

To go in the ‘ESG direction’, we impose the tree to allocate the stocks to the *best* and the *worst* portfolio (where the stocks in the best portfolio are more sustainable than the stocks in the worst). The choice of the ESG variable and the relevant thresholds for the split is made by our ML approach. This yields the best optimization result for the chosen portfolio metric, after having tried all the possible variables with all the possible thresholds in the set. These are 20, 25, 30, 35, 40, 45, and 50 per cent for the lower bound and, as a complement, 80, 75, 70, 65, 60, 55, and 50 per cent for the upper bound. A simple optimization argument allows the algorithm to be linear instead of quadratic in the number of different thresholds.

With decision trees, we start from a root (graphically it is often at the top) and we create splits that generate new branches. We explain hereafter what our trees do by starting from the meaning of the first split.

The first split consists in dividing the stocks in the best percentile and comparing them to the ones in the worst percentile (Fig. 2). We write on each branch the values of the thresholds. We highlight 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 neither in the best nor in the worst portfolio.

The power of the decision tree approach stems from the interaction between the variables, which can be grasped by adding more splits at each node. However, adding too many splits could complicate the understanding of the model. We thus decided to limit our structure to a 2-level tree for the benefit of interpretability. We added a second split identical to the first one, to sort our stocks with respect to a

⁴Early work on the use of decision trees for corporate governance factor selection can be found in Misangyi and Acharya (2014).

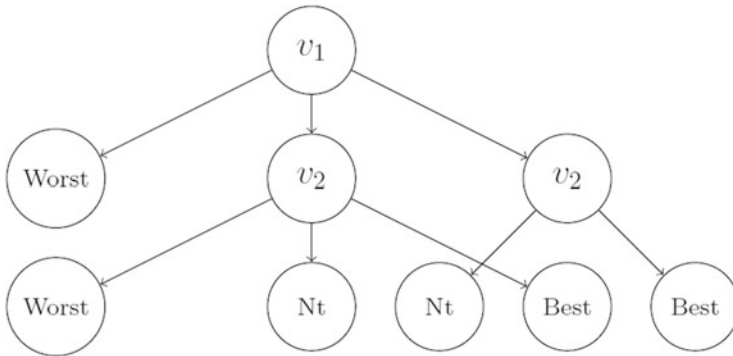


Fig. 3 The second split for decision trees

second ESG 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 relating to the second variable is high from the ESG viewpoint, the split can leave the stocks in the neutral zone or put them in the worst portfolio if the score is low. A third split (on the same level) is added by using the second variable, to introduce the possibility to downgrade to neutral (but not to worst) stocks that were put in the best portfolio at the first step (Fig. 3). The idea behind these choices is to leave space for the second variable to ‘correct’ the sorting of the first one, by leaving to the first variable the leading role in the decision.

The strength of this approach is twofold: (i) it looks straight at portfolio performance rather than at indirect indicators that could suggest a good portfolio performance; (ii) all the available data are used at each time. The model allows us to grasp a simple interpretation of the results. Despite the strong appeal of the empirical results, the explanations and possible correction mechanisms are left to the choice of the interpreter of results. Unlike some recent uses of ML in finance, our approach has the advantage of being tailored for long-term performance rather than the study of high-frequency data, since the objective has been set up as one-year performance.

Overall, although we tried to keep our exercise as parsimonious as possible, the burden of numerical calculation is quite significant as it involves 252 stocks, 125 dates, and 217 ESG indicators with 7×2 (best and worst) thresholds; in addition, every combination is repeated three times, according to the three financial objectives.

4.3 Training the Trees

We have chosen the period 2007–2016 as the training period, while the test period is 2016–2019.

Once the best first split for each ESG variable is found, the best ESG variables in the second split are selected, and only afterwards are the best thresholds for the third split computed. We have given a score to weight each ESG factor according to its importance in this process. To include the impact of a variable also in interaction with other variables, we compute the base score as the difference between the best and the worst portfolio for the chosen financial variable at the first split. We add to this base score one-third of the increase in score given at every positive contribution at the second or third split, excluding those contributions that leave in the last 5 years less than five stocks in any portfolio (best or worst).

Finally, the ESG variables are sorted by their overall score and the worst and the best portfolios are constructed using the top and bottom ten variables, selecting the stocks classified as best first split for each variable 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 than one-fourth of the weight in each portfolio.

The same analysis was repeated afterwards using only environmental variables to focus on the profiles that attract a growing consideration of the investors as an important source of climate-related risks.

Finally, the portfolios are tested in-sample and out-of-sample for each of the portfolio financial indicators, and the returns are regressed to the FF five factors and with the macroeconomic variables in the BIRR model. As expected, we find a strong correlation with the market portfolio. This is not surprising, since we are working inside the universe of the benchmark. The alpha intercept in each regression is always larger for the best portfolio, with the highest statistical significance for the mean absolute return optimizations.

5 Results

We present the results of our analysis separately for the three indicators of risk/return considered as the objective of portfolio construction, namely:

- mean absolute return
- variance
- Sharpe ratio.

By using Eq. (2), 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, and conservativeness; then we test whether the residual extra-return can be attributed to the alpha generated by the ESG key

indicator.⁵ A similar factor analysis is performed to disentangle the contribution of macroeconomic variables of the BIRR model from the BmW portfolios' risk and return indicators using Eq. (1).

For each case, we provide information about the ESG indicators (the first exercise, commented in Sect. 5.1) and the environmental indicators only (second exercise in Sect. 5.2) that we found as the most significant. For both exercises, we show the following information:

- the tables with the ten ESG indicators, showing the score (weight) of each indicator in combination with another indicator or alone, whether the indicator is a bivariate variable or not, the type (environmental, social, or governance), the threshold we found as significant for discriminating best over worst portfolios at the 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- and out-of-sample exercises;
- the value of the monthly return, variance, Sharpe ratio, and maximum drawdown for the best and worst portfolios, over a one-year horizon, for both in- and out-of-sample exercises; and
- the statistics for the regressions of the best/worst portfolio returns with the factor models (FF five-style factors and BIRR) to assess the additional contribution of the ESG indicators (where the intercept of the regression can be considered as the alpha of the ESG component) and their significance (p Value and other statistics).

We found that the best portfolios in-sample were the best also out-of-sample, with better results in each portfolio variable. Only the out-of-sample return of the best portfolio obtained by optimizing the difference BmW in variance was below the out-of-sample return of the worst portfolio. Good results were obtained also for the drawdown, 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 portfolio construction with ten ESG indicators shows that those selected for maximizing the difference BmW of absolute return provide a positive outcome; this holds true in-sample and out-of-sample, with a yearly return difference of around 4.5 per cent and 1.2 per cent, respectively (38 and 10 basis points, or bps,

⁵The 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).

Table 4 ESG indicators

	Absolute return		Variance		Sharpe ratio	
	In-sample	Out-of-sample	In-sample	Out-of-sample	In-sample	Out-of-sample
Return BmW (annualized)	4.5%	1.2%	1.2%	-0.6%	2.4%	0.5%
Variance BmW (annualized)	0.01%	-0.02%	-0.12%	-0.09%	0.18%	-0.09%
Sharpe ratio BmW	0.07393	0.03856	0.02661	0.02058	0.04853	0.046937
Alpha FF BmW	3.66%		0.81%		1.70%	
Alpha BIRR BmW	3.28%		0.23%		1.07%	

on a monthly basis; Table 4). Given a very small increase in the variance, the Sharpe ratio difference BmW improves by 0.039 (see Appendix 1).

Looking at the factor contribution with the FF model, we note that the alpha generated by the ESG indicators provides an annualized return difference BmW of 3.7 per cent (31 bps per month) and a similar magnitude with the BIRR model (3.3 per cent). Both are statistically significant. The graph on the right shows that the number of stocks of the best and worst portfolios increases over time, as more data at security level are available for the selected ESG indicators. This pattern is similar through all the exercises we have carried out and it underscores how helpful it 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 ten ESG indicators contribute to the construction of the best portfolios which slightly lower the variance both in-sample and out-of-sample (-12 bps and -9 bps on a yearly basis, respectively) and also display a better Sharpe ratio (by 0.02 out-of-sample), as the 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 an annualized difference BmW of 0.8 per cent (7 bps per month) and 0.2 per cent (2 bps per month), respectively, which are both statistically significant for the best portfolios.

For the maximization of the difference BmW of the Sharpe ratio the in-sample and out-of-sample results are similar, with a difference of 0.049 and 0.047, respectively; this case also yields positive results in the return difference BmW (+2.4 per cent yearly in-sample and +0.5 per cent out-of-sample) and in annualized variance (-18 bps and -9 bps). Disentangling the factor contribution with the FF factor model and BIRR model shows that the alpha generated by the ESG indicators provides an annualized difference BmW of 1.7 per cent annualized (14 bps per month) and 1.1 per cent (9 bps monthly), respectively, which are both statistically significant for the best portfolios.

Among the most material ESG indicators in our portfolio construction, 9 out of 17 are related to environmental issues. This finding highlights the relevance of the environmental issues for equity portfolio performance. The environmental indicators relate not only to carbon emissions (via the carbon intensity) but also to waste

Table 5 The most significant ESG indicators

	Return	Variance	Sharpe	Tot	Biv	Type
CO ₂ 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

Note: Tot: number of financial objectives for which the indicator is significant; Biv: flag for bivariate indicator; Type: indicator's group (E, S, or G)

management, recycling, and eco-innovation. Interestingly, the environmental score of one of the providers is identified as material but it is not on the first ones. Of the other indicators, five are related to social profiles (mainly about employee safety) and three to governance factors, with a prominent role for diversity. Only four ESG variables are bivariate (Table 5).

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 a lower over-performance for the objective of variance optimization (out-of-sample), while positive results are provided with the objective of Sharpe ratio difference maximization. Remarkably good results are obtained for the objective of absolute return, where also the variance (out-of-sample) and alphas are clearly in favour of BmW.

Our findings, obtained with a novel ML approach, are consistent with previous evidence from several studies which apply alternative models and techniques. In particular, these studies find extra performance for stocks with better indicators relating to environmental issues (carbon intensity, as in Bernardini et al. 2021a, b; Mats et al. 2016; In et al. 2019), social profiles (employee satisfaction, as in Edmans 2011), governance structure (Li and Li 2018), and gender diversity (Nguyen 2020). The empirical relevance of ESG factors in building efficient portfolios, as shown in our study, is in line with the findings of Kaiser (2020), Kumar et al. (2016), Giese

et al. (2019), and Maiti (2021). Other studies find mixed results (Billio et al. 2021) or show opposite results (Pedersen et al. 2021; De Spiegeleer et al. 2021).

5.2 Results for Environmental Indicators

The analysis of portfolio construction with ten environmental indicators, besides those identified in the previous section, finds some complementary indicators. The maximization of the difference BmW of absolute return shows that the environmental indicators bring larger differential return out-of-sample compared with the ESG indicators, with an annualized return difference of 1.8 per cent (compared with 1.2 per cent for ESG indicators), lower variance, and thus a higher Sharpe ratio (0.07, see Appendix 2). Besides, the in-sample results show a positive BmW difference for the return (+2.8 per cent on annual basis) and Sharpe ratio (0.04). The analysis of the factor contribution shows that the alpha generation by constructing portfolios with environmental indicators is significant both with the FF model (2.8 per cent annually and 24 bps monthly) and with the BIRR model (2.0 per cent annually and 17 bps monthly; Table 6).

The optimization of BmW difference in variance shows that the ten environmental indicators contribute not only to reducing the variance but also to a positive annualized return difference (0.2 per cent in-sample and 0.8 per cent out-of-sample) and a Sharpe ratio increase (+0.08 and +0.05, respectively). The alpha provides mixed results, as it is positive with the FF factor decomposition (+0.63 per cent annualized) and slightly negative with the BIRR model (−0.19 per cent), 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 annualized return increase is 3.2 per cent and 1.8 per cent, respectively; the variance reduction is 26 bps and 10 bps; the Sharpe ratio increase is 0.07 and 0.09. The factor contribution exercise shows that the alpha generated by the environmental indicators is remarkably large: it is 2.9 per cent on an annualized basis with the FF factor model

Table 6 Environmental indicators

	Absolute return		Variance		Sharpe ratio	
	In-sample	Out-of-sample	In-sample	Out-of-sample	In-sample	Out-of-sample
Return BmW (annualized)	2.8%	1.8%	0.2%	0.8%	3.2%	1.8%
Variance BmW (annualized)	0.03%	−0.05%	−0.06%	−0.07%	0.26%	−0.10%
Sharpe ratio BmW	0.04461	0.06908	0.00802	0.04786	0.07063	0.08947
Alpha FF BmW	2.84%		0.63%		2.91%	
Alpha BIRR BmW	2.01%		−0.19%		1.37%	

Table 7 The most material environmental indicators

	Return	Variance	Sharpe	Tot	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

and 1.4 per cent with the BIRR model, and the best portfolios are statistically significant.

Among the most significant environmental indicators, besides those already found in the ESG case study, some are based on the assessment of providers. This highlights the role of forward-looking evaluation of the environmental issues and climate-change risks. In turn, this strengthens the notion that corporates should manage such risks and move forward adaptation techniques, like renewables and clean technologies (Table 7).

6 Conclusions

ESG investing is enjoying a remarkable growth in terms of supply and demand. This creates a general interest in the transparency and consistency of the ESG assessment of firms. In the absence of standardized methodologies, the providers of ESG scores and ratings adopt a variety of proprietary techniques, which results in the low correlation of the ESG scores across different providers. Our research proposes a model-free approach that overcomes some of the limits of ESG scores. We identify a strategy that directly employs ESG indicators, and more specifically environmental factors, to build equity portfolios that generate efficient financial results, with

superior return and lower risk than those obtained with traditional factor models of the stock market.

The risk and return differentials are statistically and economically significant even after taking into account the contribution of the standard Fama-French model with style factors and of the BIRR model with macroeconomic factors. Among the risk/return indicators we have chosen—return, Sharpe ratio, and variance—our strategy provides the best results for the first two, while the contribution to variance is mixed. Our results are consistent with previous evidence, showing a positive performance differential for stocks with better indicators for the ESG profiles.

Our findings indicate that an investor in the European equity market who had developed the proposed ML technique in 2016 and applied it in the period from January 2017 to April 2019 would have achieved on average an extra annualized return between 0.5 and 1.2 percentage points over the Eurostoxx index, depending on the different risk/return objectives, and using the ESG indicators identified for portfolio construction; the extra return would have been between 0.8 and 1.8 percentage points using the environmental indicators only.

These findings prompt three remarks. First, the direct use of ESG indicators seems to have a significant payoff in terms of financial performance. Second, our findings support the notion that quantitative information on the company sustainability profiles is quite important and should be improved, by means of greater corporate disclosure, possibly via regulation aimed at wider consistency and comparability. Useful information may be extracted from the available ESG indicators other than the scores sold by professional providers. Among the ESG variables selected with our ML technique, half are environmental and some refer to the company exposure and ability to manage climate change risk. Among the selected environmental variables, only one corresponds to the environmental score of a provider. This means that the ESG scores do not exhaust the information available in the data disclosed by the firms.

As we were not able to measure the extent to which the evaluation by providers integrates climate-related scenarios, if at all, future research could investigate additional firm-level indicators based on climate scenarios and possibly perform a stress test analysis under different transition pathways.

Since the proposed ML methodology is fairly new, more can be done to test its robustness. Our validation was done by comparing the results of training in the first period with the out-of-sample results. Future research could try some form of cross-validation. As an alternative, one could try a shorter training period. The disentangling methodology to detect the specific contribution of ESG and environmental indicators was implemented by means of the Fama-French and BIRR models. A test for a naive portfolio could be carried out in future research. Furthermore, an analysis of the relevance of the ESG variables by sector could be carried out. Finally, a deeper understanding of our model would be warranted by experimenting with different methodologies in splitting and variable choice. For instance, one can develop a bootstrap technique that suits the portfolio construction (bagging) and experiment with restrictions on the number of variables at each split (random forest).

Appendices

Appendix 1: Portfolios Obtained with ESG Indicators

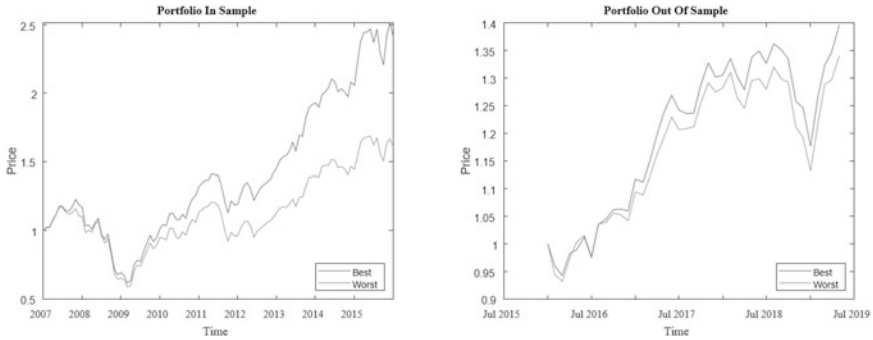


Fig. 4 Cumulative returns, in sample and out of sample, of best and worst portfolios built by optimizing return

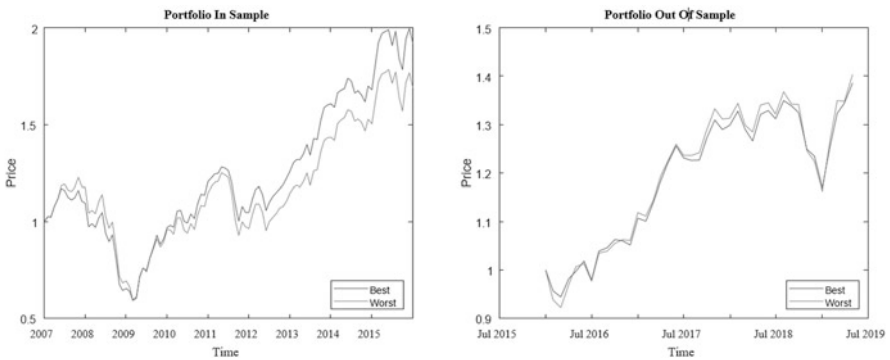


Fig. 5 Cumulative returns, in sample and out of sample, of best and worst portfolios built by optimizing variance

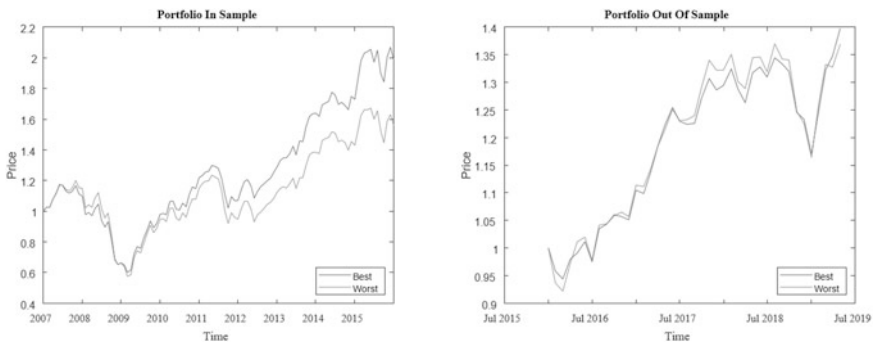


Fig. 6 Cumulative returns, in sample and out of sample, of best and worst portfolios built by optimizing Sharpe ratio

Appendix 2: Portfolios Obtained with Environmental Indicators

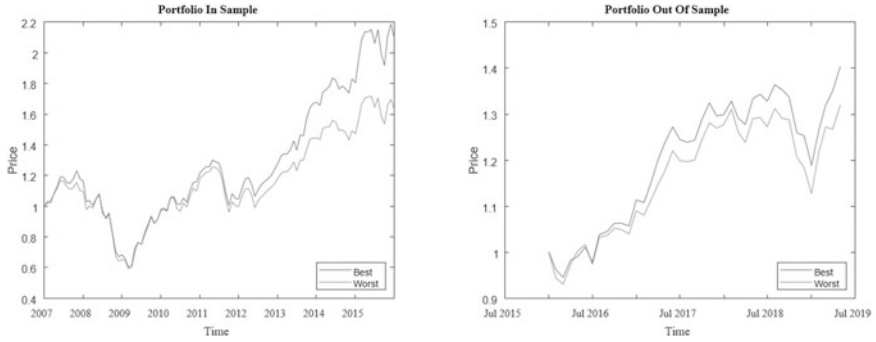


Fig. 7 Cumulative returns, in sample and out of sample, of best and worst portfolios built by optimizing return

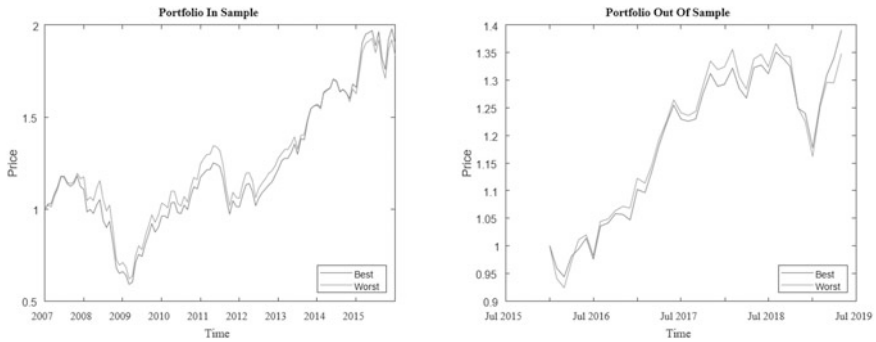


Fig. 8 Cumulative returns, in sample and out of sample, of best and worst portfolios built by optimizing variance

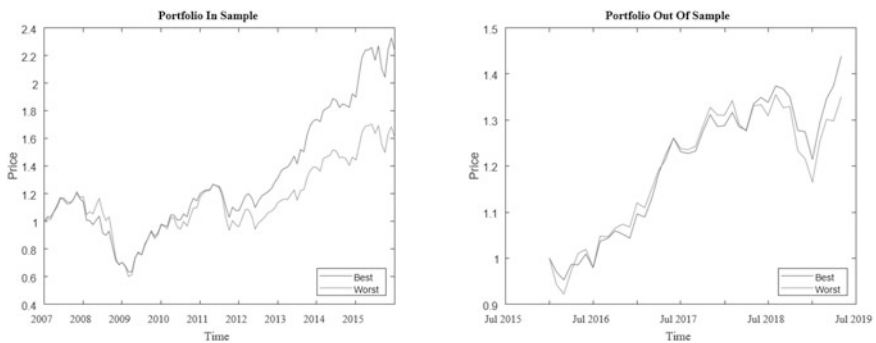


Fig. 9 Cumulative returns, in sample and out of sample, of best and worst portfolios built by optimizing Sharpe ratio

References

- Adebayo TS, Abdul Kareem HKK, Bilal, Kirikkaleli D, Shah MI, Abbas S (2022) CO2 behavior amidst the Covid-19 pandemic in the United Kingdom: The role of renewable and non-renewable energy development. *Renew Energy* 189:492–501
- Allen K (2018) Lies, damned lies and ESG rating methodologies. <https://www.ft.com/content/2e49171b-a018-3c3b-b66b-81fd7a170ab5>
- Allen E, Lyons K, Tavares R (2017) The application of machine learning to sustainable finance. *J Environ Invest* 8(1)
- Arnott R et al (2013) The surprising Alpha from Malkiel’s monkey and upsidedown strategies. *J Portf Manag* 39(4)
- Atz U, Van Holt T, Liu ZZ, Bruno C (2022) Does sustainability generate better financial performance? Review, meta-analysis, and propositions. *J Sustain Finance Investment*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3708495
- Berg F, Kolbel J, Rigobon R (2022) Aggregate confusion: the divergence of ESG ratings. *Rev Finance* 26(6):1315–1344. <https://doi.org/10.1093/rof/rfac033>
- Bernardini E, Di Giampaolo J, Faiella I, Poli R (2021a) The impact of carbon risk on stock returns: evidence from the European electric utilities. *J Sustain Finance Investment* 11(1):1–26.
- Bernardini E, Faiella I, Lavecchia L, Mistretta A, Natoli F (2021b) Central banks, climate risks and sustainable finance. Occasional Papers 608 Bank of Italy
- Billio M, Costola M, Hristova I, Latino C, Pelizzon L (2021) Inside the ESG ratings: (dis)agreement and performance. *Corp Soc Responsib Environ Manag* 28(5):1426–1445. <https://doi.org/10.1002/csr.2177>
- Bowman EH, Haire M (1975) A strategic posture toward corporate social responsibility. *Calif Manag Rev* 18:49–58. <https://doi.org/10.2307/41164638>
- Burmeister E, Roll R, Ross SA (2003) Using macroeconomic factors to control portfolio risk. In: The Institute of Chartered Financial Analysts (ed) *A practitioner’s guide to factor models*, vol 9. The Research Foundation of the Institute of Chartered Financial Analysts, Charlottesville, VA, pp 1–27
- Carboni P (2017) 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. Draft
- Chan T et al (2011) Optimizing portfolio construction using artificial intelligence. *J Inf Commun Technol* 3(16):168–175
- Chatterji AK et al (2016) Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strateg Manag J* 37(8):1597–1614
- Cheng B, Ioannou I, Serafeim G (2014) Corporate social responsibility and access to finance. *Strateg Manag J* 35(1):1–23
- Clark GL, Feiner A, Viehs M (2015) From the stockholder to the stakeholder: how sustainability can drive financial outperformance. <https://doi.org/10.2139/ssrn.2508281>
- De Franco C (2019) Performance of ESG and machine learning investment approaches. Ossiam, Paris, France
- De Spiegeleer J, Höcht S, Jakubowski D, Reyners S, Schoutens W (2021) ESG: a new dimension in portfolio allocation. *J Sustain Finance Invest*
- Doyle TM (2018) Ratings that don’t rate. the subjective world of ESG ratings agencies. American Council for Capital Formation, ACCF. https://accfcorgov.org/wp-content/uploads/2018/07/ACCF_RatingsESGReport.pdf
- Edmans A (2011) Does the stock market fully value intangibles? Employee satisfaction and equity prices. *J Financial Econ* 101(3):621–640
- Erhardt J (2020) The search for ESG alpha by means of machine learning - a methodological approach. <https://ssrn.com/abstract=3514573>
- Fama EF (1970) Efficient capital markets: a review of theory and empirical work. *J Finance* 25(2):383–417

- Fama EF, French K (1993) Common risk factors in the returns on stocks and bonds. *J Financial Econ* 33:3–56
- Fama EF, French K (2015) A five-factor asset pricing model. *J Financial Econ* 116(1):1–22
- Fama EF, MacBeth JD (1973) Risk, return, and equilibrium: empirical tests. *J Polit Econ* 81(3): 607–636
- Fareed Z, Rehman MA, Adebayo TS, Wang Y, Ahmad M, Shahzad F (2022) Financial inclusion and the environmental deterioration in Eurozone: the moderating role of innovation activity. *Technol Soc* 69
- Farrell J (1985) The dividend discount model: a primer. *Financial Anal J* 41:16–25
- Feiner A (2018) Machine learning and big data enable a quantitative approach to ESG investing. In: Greis MJ (ed) *Mainstreaming sustainable investing*, vol 7. CFA, Charlottesville, VA, pp 14–20. <https://www.cfainstitute.org/-/media/documents/article/rf-brief/mainstreaming-sustainable-investing.pdf>
- Ferriani F, Natoli F (2021) ESG Risk in Times of Covid-19. Covid-19 Notes, Bank of Italy (June 15 2020)
- Friede G, Busch T, Bassen A (2015) ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *J Sustain Finance Invest* 5(4):210–233
- Giese G, Lee L, Melas D, Nagy Z, Nishikawa L (2019) Foundations of ESG investing how ESG affects equity valuation, risk and performance. *J Portf Manag* 45(5)
- Giudici G, Bonventura M (2018) La relazione fra rating ESG e performance di mercato: uno studio sui titoli dell'indice Stoxx Europe 600. *Quaderno di ricerca Banor SIM* 1(1):1–36
- Global Compact (2004) The global compact leaders summit: final report. United Nations Headquarters, 24 June
- Global Sustainable Investment Alliance (2020) GSIA. Global sustainable investment review 2020. <https://www.gsi-alliance.org/trends-report-2020/>
- Godfrey PC, Merrill CB, Hansen JM (2009) The relationship between corporate social responsibility and shareholder value: an empirical test of the risk management hypothesis. *Strateg Manag J* 30(4):425–445
- Hart O, Zingales L (2017) Companies should maximize shareholder welfare not market value. *J Law Finance Account* 2(2):247–275. https://scholar.harvard.edu/files/hart/files/108.00000022-hart-vol2no2-jlfa-0022_002.pdf
- Henriksson R et al (2019) Integrating ESG in portfolio construction. *J Portf Manag* 45(4):67–81. <https://jpm.pm-research.com/content/45/4/67.full.pdf>
- Hoepner AGF (2010) Portfolio diversification and environmental, social or governance criteria: Must responsible investments really be poorly diversified?. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1599334.
- In SY, Park KY, Monk AH (2019) Is 'being green' rewarded in the market?: An empirical investigation of decarbonization and stock returns. Stanford Global Project Center Working Paper. <https://ssrn.com/abstract=3020304>
- Intergovernmental Panel on Climate Change (2022) Synthesis report of the sixth assessment report. IPCC. <https://www.ipcc.ch/ar6-syr/>.
- International Monetary Fund (2019) Sustainable finance: looking farther. The Global Financial Stability Report. <https://www.imf.org/-/media/Files/Publications/GFSR/2019/October/English/ch6.ashx>
- Kaiser L (2020) ESG integration: value, growth and momentum. *J Asset Manag* 21(4):32–51
- Khan M, Serafeim G, Yoon A (2016) Corporate sustainability: first evidence on materiality. *Account Rev* 91(6):1697–1724
- Kumar NCA, Smith C, Badis L, Wang N, Ambrosy P, Tavares R (2016) ESG factors and risk-adjusted performance: a new quantitative model. *J Sustain Finance Investment* 6(4):292–300
- Kumar S, Sharma D, Rao S et al. (2022). Past, present, and future of sustainable finance: insights from big data analytics through machine learning of scholarly research. *Ann Oper Res*
- Li D, Li EXN (2018) Corporate governance and costs of equity: theory and evidence. *Manag Sci* 64(1):83–101

- Maiti M (2021) Is ESG the succeeding risk factor? *J Sustain Finance Investment* 11(3):199–213
- Mats A, Bolton P, Samama F (2016) Hedging climate risk. *Financial Anal J* 72:13–32
- Meuer J, Koebel J, Hoffmann V H (2019) On the nature of corporate sustainability. *Organization and Environment*.
- Misangyi V, Acharya A (2014) Substitutes or complements? A configurational examination of corporate governance mechanisms. *Acad Manag J* 57:1681–1705
- Mohanram PS (2008) separating winners from losers among low book-to-market stocks using financial statement analysis. *Rev Account Stud* 10:171–184
- Mohanty S (2019) Does one model fit all in global equity markets? Some insight into market factor based strategies in enhancing alpha. *Int J Finance Econ* 24(1):1170–1192
- Mohanty S, Pontiff J, Woodgate J (2008) Share issuance and cross sectional returns. *J Finance* 63(2):921–945
- Nguyen P (2020) Board gender diversity and cost of equity. *Appl Econ Lett* 27(18):1522–1526
- Nguyen Q, Diaz-Rainey I, Kurupparachchi D (2021) Predicting corporate carbon footprints for climate finance risk analyses: a machine learning approach. *Energy Econ* 95
- Novy-Marx R (2013) The other side of value: the gross profitability premium. *J Financial Econ* 108(1):1–28
- Pedersen LH, Fitzgibbons S, Pomorski L (2021) Responsible investing: the ESG-efficient frontier. *J Financial Econ* 142(2):572–597
- Peeters R (2003) The maximum edge biclique problem is NP-complete. *Discret Appl Math* 131(3): 651–654
- Ross S (1976) The arbitrage theory of capital asset pricing. *J Econ Theory* 13(3):341–360.
- Schoenmaker D Schramade W (2018) Investing for long-term value creation. CEPR Discussion Papers 13175 <https://ideas.repec.org/p/cpr/ceprdp/13175.html>
- Sokolov A, Mostovoy J, Ding J, Seco L (2021) Building machine learning systems for automated ESG scoring. *J Impact ESG Invest.* <https://doi.org/10.3905/jesg.2021.1.010>
- State Street Global Advisors (2019) The ESG data challenge. <https://www.ssga.com/library-content/products/esg/esg-data-challenge.pdf>
- Verheyden T, Eccles RG, Feiner A (2016) ESG for all? The impact of ESG screening on return, risk, and diversification. *J Appl Corp Finance* 28(2):47–55