Computational Economics
https://doi.org/10.1007/s10614-024-10602-8

Can Machine Learning Explain Alpha Generated by ESG Factors?

Vittorio Carlei[1] · **Piera Cascioli**[1] · **Alessandro Ceccarelli**[1] · **Donatella Furia**[1]

Accepted: 7 April 2024

Abstract

This research explores the use of machine learning to predict alpha in constructing portfolios, leveraging a broad array of environmental, social, and governance (ESG) factors within the S&P 500 index. Existing literature bases analyses on synthetic indicators, this work proposes an analytical deep dive based on a dataset containing the sub-indicators that give rise to the aforementioned synthetic indices. Since such dimensionality of variables requires specific processing, we deemed it necessary to use a machine learning algorithm, allowing us to study, with strong specificity, two types of relationships: the interaction between individual ESG variables and their effect on corporate performance.The results clearly show that ESG factors have a significant relationship with company performance. These findings emphasise the importance of integrating ESG indicators into quantitative investment strategies using Machine Learning methodologies.

Keywords Sustainability · Machine learning · Portfolio management

JEL Classification G1 · G11 · G12

✉ Donatella Furia
donatella.furia@unich.it

Vittorio Carlei
vittorio.carlei@unich.it

Piera Cascioli
piera.cascioli@unich.it

Alessandro Ceccarelli
alessandro.ceccarelli@pharmadvice.it

[1] University of Chieti-Pescara, Viale Pindaro 42, 65127 Pescara, Italy

Published online: 30 April 2024

1 Introduction

The impact of ESG factors on corporate performance and stock returns has attracted significant attention in recent years. As ESG considerations influence an increasing number of publicly traded companies, extending beyond corporate boundaries to shape the wealth management industry landscape, understanding the relationship between ESG scores, corporate performance, and stock returns becomes crucial. This article provides a comprehensive review of the literature and an analysis of this relationship, examining the evolution of the role of ESG criteria in financial markets.

The article explores the components of ESG indices, highlighting their importance in measuring a company's impact on the environment, stakeholder relations, and governance practices. Despite the positive association between ESG scores and stock prices (Leung et al., 2018; Rossi & Harjoto, 2020), the study delves into nuances, such as the variable impact of individual ESG components on the accuracy of analyst forecasts.

Examining market reactions, the article discusses how companies with high ESG scores may exhibit less pronounced responses to positive news, suggesting that their positive impact has already been integrated into stock prices. Conversely, companies with notable ESG profiles are identified as less likely to experience future decreases in stock prices.

Empirical studies demonstrate that companies with sustainable objectives tend to show higher corporate value and improved capital and asset efficiency. However, conflicting research results present negative correlations between ESG indices and stock market premiums, highlighting the ongoing debate on the financial impact of ESG practices.

Despite divergent opinions, researchers conclude that companies with excellent ESG performance may experience reduced stock price volatility during crises, providing stability to premiums. The article also highlights the positive long-term returns associated with ESG investments, contributing to the sustainable development of financial markets and generating benefits for both the real economy and the environment.

Moving from theoretical discussions to practical applications, the article underscores the significant role of ESG scores in shaping investment decisions and influencing the construction and trading of stock portfolios. However, the reliability of ESG scores in asset purchase decisions remains a subject of investigation.

Generally, while existing literature bases analyses on synthetic indicators (La Torre et al., 2020), this work proposes an analytical deep dive based on a dataset containing the sub-indicators that give rise to the aforementioned synthetic indices. Since such dimensionality of variables requires specific processing, we deemed it necessary to use a machine learning algorithm, allowing us to study, with strong specificity, two types of relationships: the interaction between individual ESG variables and their effect on corporate performance.

To address gaps in recent literature, which mostly uses linear survey methodologies (Duque-Grisales & Aguilera-Caracuel, 2021; Alareeni & Hamdan, 2020; La

Torre et al., 2020; Ademi & Klungseth, 2022; Umar et al., 2022; Zhou & Zhou, 2022; Wu et al., 2022; Serafeim & Yoon, 2023; Li et al., 2022; Lapinskienė, 2023; Liu & Wan, 2023; Cohen, 2023; Nguyen et al., 2022), the article introduces a computational methodology based on machine learning. Leveraging the real Morningstar dataset "ESG Risk Ratings & Variables," which provides a robust range of indices for all major asset classes, the study focuses on a broad range of companies listed in the S&P 500. The methodology aims to clarify the interaction between ESG factors, stock price dynamics, and related variables within this diversified index.

The structure of the paper is as follows. Section 1 introduces the research question. In Sect. 2 a comprehensive examination of the literature is provided. Section 3 outlines the empirical methodology employed in this study. The findings of the analysis and their corresponding discussion are presented in Sect. 4. Finally, Sect. 5 contains the concluding remarks.

2 A Literature Review

Influencing now a large number of listed companies (Serafeim, 2014), the importance of ESG impact not only belongs to companies but also has become part of the asset management industry (Kim & Yoon, 2023). In this section of the article, we propose a brief literature review of studies that have analyzed the relationship among ESG scores, corporate performance, and the effect on stock returns.

The financial scandals of the twenty-first century and the global financial crisis of 2008 undermined the foundations of international markets, causing economic and social misalignments (Alareeni & Hamdan, 2020; Nicholson et al., 2011). As a result, these latter issues have directed companies toward more ethical behavior and risk-averse strategies (Galbreath, 2013). Indeed, ESG criteria have progressively gained prominence within financial markets due to innovative management strategies that are deeply influenced by factors such as the environment, health and safety, sustainability, and corporate social responsibility (CSR). Their popularity is mainly due to investors' risk protection, transparency, and compliance with stricter regulations on global risks, such as climate change (Lapinskienė et al., 2023; Rodionova et al., 2022). As an example, the ESG index includes three crucial factors. First, the environmental aspect measures the impact on the environment; in addition, the social aspect concerns the link between stakeholders. Finally, the governance criterion reveals how companies are managed.

The growing awareness of the environment, the search for greater ethics within the workplace, along with social responsibility, has brought new information to financial markets, where participating companies—to adapt to the new paradigm of implementing ESG criteria—have changed their activities and corporate composition (Beaver, 1968).

Although the literature has focused extensively on the relationship between sustainable corporate performance based on the evaluation of ESG criteria, this paper is related to the stream of studies that addresses the financial effect following the implementation of ESG indices on stock shares (Hong et al., 2019).

ESG factors tend to influence (often positively) the ESG score on share price predictions to the point that in companies with higher ESG scores, analysts are able to obtain a forecast error that is not high. However, each individual component does not always respond positively; in fact, Umar et al. (2022), performing a panel regression of analysts' forecast performance and ESG score, observe a significant relationship between environmental and governance scores and target price accuracy, while social factors show no influence.

Another study shows that the market reaction to positive news is less noticeable for companies with high ESG ratings, meaning that the positive impact has already been reflected in the stock price (Serafeim & Yoon, 2023). According to BlackRock, companies with noteworthy ESG profiles are less likely to have decreasing future stock prices (Jayachandran et al., 2013; Jiao, 2010; Konar & Cohen, 2001).

As a main consequence, for companies with sustainable goals, firm value appears higher and their capital and asset efficiency are improving. Eccles et al. (2014) conduct an analysis of several U.S. companies and report that highly sustainable companies outperform their competitors with lower ESG index valuations in the stock market. On the other hand, some researchers have outlined negative relationships, as the primary goal of a company is to increase the wealth of its shareholders (Friedman, 2009). All other purposes, even if considered positive, could hinder the primary objective and, thus, reduce the overall effectiveness. In this regard, Landi and Sciarelli (2018) find a negative correlation between ESG indices and stock market premiums. Ultimately, they state that some ESG-oriented companies never enjoy abnormal returns.

Nonetheless, the researchers have concluded that the excellent ESG performance of publicly traded companies reduced stock price volatility during the pandemic, stabilizing, among other things, the premiums (Zhou & Zhou, 2022). The latter study analyze the MSCI index in the Bloomberg database on environmental and social performance in relation to the corporate governance of listed companies. Ultimately, during COVID-19, green stocks with low ESG risks were found to have higher returns and lower volatility than those with higher ESG risk ratings (La Torre et al., 2020; Xiong, 2021).

Empirical studies investigating the correlation between ESG performance and corporate value (Duque-Grisales & Augilera-Caracuel, 2021; Fatemi et al., 2015; Huang, 2021, etc.) demonstrated that solid ESG performances are associated with better corporate evaluation. Moreover, when a company takes strategic measures to enhance its ESG performance, it simultaneously elevates its rating and reputation among various stakeholders (Kim et al., 2018).

Furthermore, Minor and Morgan (2011) provided a compelling example of how an improved Corporate Social Responsibility (CSR) reputation acts as a protective shield, effectively guarding companies against adverse shocks and thereby reinforcing their organizational legitimacy. In a similar vein, Marsat et al. (2021) presented robust evidence supporting the idea that strong adherence to environmental regulations enables companies to swiftly rebound from environmental controversies.

From a practical perspective, ESG ratings are used by investment professionals and influence the construction and trading of equity portfolios. However, their reliability in asset purchase decision making is still unclear (Berg et al., 2020; Chatterji et al., 2016). In this regard, one strand of the research has focused on the ability of ESG ratings to predict future ESG news by testing the usefulness of early scores (Serafeim & Yoon, 2023). The study shows that ratings indeed represent market expectations of future performance and are therefore useful in predicting stock returns.

ESG investments lead to higher long-term returns for investors, maintaining the healthy and sustainable development of financial markets and producing benefits for the real economy and the environment (Wu et al., 2022).

Wu et al. (2022) conduct a study of the impact of ESG certification on price efficiency in Chinese listed companies and find that, as a result of ESG certification, securities on ESG lists have better price efficiency than those removed from ESG lists. From this analysis, the authors also find two mechanisms through which ESG certification improves the evaluation of price efficiency, including improving stock liquidity and reducing information asymmetry. This positive effect leads investors, when making an investment decision to incorporate ESG certifications into their information set to assess the value of the company.

Based on a general review at the relevant literature, we have noted that there are numerous studies on quantitative models applied to a medium-small sample of data (Li et al., 2022); however, little research goes into testing the relationship between ESG indices and stock prices on a broad macrolevel range of companies, such as the S&P 500.

In this scenario, Ademi and Klungseth (2022), by implementing in STATA a fixed-effects regression and a weighted least squares model to analyze the panel data, investigate the relationship between a company's ESG performance and financial performance by using financial data and ESG scores of 150 publicly traded companies on the Standard and Poor's 500 index. Their results show that companies with superior ESG performance manage to perform better financially, achieving a higher market valuation than their industry peers. Cohen et al. (2023), use S&P 500 stocks and Altman's Z score to study the influence of ESG risks on companies' chances of survival, documenting a reduction in recent years in the ESG index on the stocks of this indicator, which highlights that companies focus on sustainability issues and invest resources to reduce them. In addition, Altman's Z score tends to be negatively affected by E and S but not by G, showing that for companies, the ESG score is important in that it can determine their financial performance, increasing the risk of default.

Nguyen et al. (2022), in their study, delved into the impact of ESG practices on the financial performance of non-financial firms in the United States from 2018 to 2020. Their analysis was based on a sample of 57 companies listed in the S&P 500. The findings revealed that enhanced ESG practices were associated with improved Return on Sales (ROS), Return on Equity (ROE), and TobinQ (a metric representing a firm's market value relative to the cost of its capital stock) indices. Notably, the study highlighted that the influence of ESG factors on TobinQ was significantly more pronounced than their impact on ROA and ROE. This suggests that ESG-related advantages may enhance a company's attractiveness to investors, potentially resulting in higher market valuations of a company's assets, as reflected in the TobinQ. However, it's worth noting that the TobinQ did not exhibit a substantial increase in the immediate term. Therefore, the substantial improvements in ROA and ROE may manifest over the long term rather than in the short term.

State-of-the-art research investigates the relationship between ESG indices and financial performance using various measurement techniques. In most cases, researchers utilized linear or multiple regression models (Duque-Grisales & Aguilera-Caracuel, 2021; Alareeni & Hamdan, 2020; La Torre et al., 2020; Ademi & Klungseth, 2022; Umar et al., 2022; Zhou & Zhou, 2022; Wu et al., 2022; Serafeim & Yoon, 2023; Li et al., 2022; Lapinskienė, 2023; Liu & Wan, 2023; Cohen, 2023; Nguyen et al., 2022), while some employed enterprise valuation models (Fatemi et al., 2015), and ultimately others Cox proportional hazard models (Marsat et al., 2021).

In their analyses, we observed that some authors tend to explore the impact of ESG factors on corporate performance using ESG scores (Nguyen et al., 2022), ESG ratings (Li et al., 2022; Liu & Wan, 2023; Serafeim & Yoon, 2023), ESG risk indices (Cohen, 2023), or comprehensive ESG indices (i.e. "ESG Overall Index") (La Torre et al., 2020). These metrics are provided from major databases such as MSCI, Sustainalytics, Thompson Reuters ASSET 4 (Ademi & Klungseth, 2022; Duque-Grisales & Aguilera-Caracuel, 2021; Nguyen et al., 2022; Serafeim & Yoon, 2023; Zhou & Zhou, 2022), CSRHub (La Torre et al., 2020), Bloomberg, WIND (Li et al., 2022; Liu & Wan, 2023). However, in all cases, analyses are based on aggregated summary indices, as opposed to fine-grain metrics such as the ones used in the present piece of research.

Therefore, with the aim to elucidate the interaction among ESG factors, stock price dynamics, and related variables within the S&P 500 index, we exported the custom dataset "ESG Risk Ratings & Variables". This piece of information was gently supplied by an Italian fintech (Qi4M S.R.L.) under permission of Sustainalytics, a world-wide provider owned by Morningstar, that thankfully granted the opportunity to perform research on a unique newly-released ESG dataset. Furthermore, to successfully accomplish the challenging task, we propose an innovative machine learning-driven computational methodology, designed to delve into the non-linear relationships between ESG risk factors and stock performance.

3 Unveiling ESG Effects With Machine Learning

To explain the relationship between ESG factors, stock price dynamics and correlated variables in the S&P 500 index (as a whole), we utilised data from the Morningstar dataset (*"ESG Risk Ratings & Variables"*), which provides a robust range of indexes for all major asset classes. As a positive remark, although Morningstar's variables are disclosed with varying frequencies (yearly or monthly), the present study grounds its findings on decomposed risk factors; thus, feeding the system with the whole spectrum of risk measures (12 in total) and not only a single aggregated figure. Conversely, the unified ESG score may be utilised for visualization purposes, preventing curse of dimensionality while visualising data points. Ultimately, as related to data pre-processing, a panel analysis containing the following datasets was conducted, to create a unified coherent view starting from:

(1) ESG.RiskRatings

- Dataset containing the final ESG Ratings for each Entity (company)

(2) ESG.Catalog

- Dictionary containing the variable name, description and PossibleValues (range of values)

(3) ESG.Variables

- Dataset containing the ESG variables

(4) SP500.Prices

- Dataset containing the ClosingPriceUsd on a daily basis for each company in the SP500 from 2010-01-01 to 2023-02-21

(5) SP500.Companies

- Dataset containing record information on companies in the S&P 500 like Name, Country or Ticker from 2010-01-01 to 2023-02-21

3.1 Data Engineering

A wide-ranging data processing pipeline is implemented in Python to merge stock price data with ESG information spanning more than 13 years. The procedure consists of several interconnected steps, as shown in Fig. 1, demonstrating a systematic approach to enhance the dataset. The resulting output represents a unified view of stock data, ESG scores and rating descriptions, facilitating further analysis and modelling throughout a time range from 2010-01-01 to 2023-02-01, with daily granularity. For improved clarity and readability, this paper will consistently use the term "ESG Risk Score" to denote a specific feature within the input space. For further reference, a comprehensive listing of all input variables can be found in Chapter 3. On the other hand, ESG scores or values encompass a broader spectrum of sustainability-related metrics.

To handle missing ESG values, across all risk measures, the pipeline applies forward and backward-fill techniques to each company's time series, filling 15% of missing data.

By employing these methods, missing values are substituted through a 2-steps procedure; respectively, with the most recent non-null value and then, if missing,

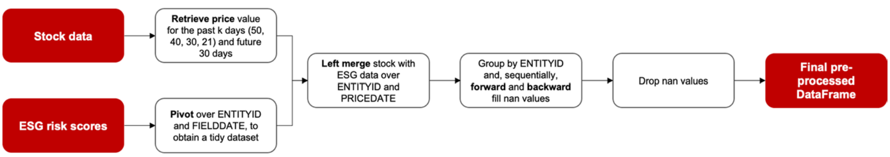


Fig. 1 Data pre-processing pipeline diagram

with the first non-missing one. In more detail, forward fill ("ffill") precedes backwards fill ("bfill") (Moahmed et al., 2014) ensuring that all gaps are properly deleted, while avoiding data leakage, as further explained in the following sections. Likewise, as predictors are slow-movers, the present research assumes that rating remains steady across years; thus, ensuring that long-term backwards fill reflects, for each company, real-world dynamics. In support of the latter statement, average standard deviation across all standardised ESG features reports a value of 0.022.

ESG risk rating data is then pivoted, resulting in a tidy transformed dataset with a panel representation. The latter reshaping operation allows the dataset to represent companies and dates as rows and, as columns, the whole range of detected risk measures. More in detail, the company's overall ESG risk rating applies the concept of risk decomposition to derive the degree of unmanaged risk. To this extent, each corporation is assigned a value between 0 and 100; the former indicates that risks have been fully managed (no unmanaged ESG risks) while 100 relates to the highest level of unmanaged risk. In short, to resemble a sustainable approach, lower values are preferred. More technically, this type of indicator is computed as the difference between a company's overall exposure score and its managed risk grade or, alternatively, by adding the Corporate Governance unmanaged risk figure to the sum of the company's unmanaged risk value.

As for trading data, to better enhance its predictive capabilities, past and future price variations, over a specified set of windows, are computed. This step involves calculating price percentage changes between current day and a predetermined number of trading days for both:

- Past (21, 30, 40 and 50 days)
- Future (30 days); the target variable that will be used within the supervised training framework

Finally, the pipeline merges augmented stock dataset with tidy ESG scores. In finer detail, the operation combines data based on company identifiers and detection date. To better highlight the uniqueness and relevance of the analysed dataset, it is crucial to clarify that we're training and testing our methodologies on one of the largest financial indexes in the world—The Standard & Poor's 500—over a period of over 13 years.

To improve diagram interpretability, an explanation of the pivoting procedure and pertinent field descriptions is reported below. In essence, a pivot table serves as an instrument for condensing and structuring data into a more concise and informative format. As depicted in the diagram, the ESG scores undergo a pivoting process to yield a final DataFrame where each row corresponds to a specific data collection date—denoted by "FIELDDATE"—and a single company (identified as "ENTITYID"). As columns, all available ESG scores are utilized. In this format, the dataset is tidy and ready for further processing.

To further exemplify, all relevant fields are listed as:

- ENTITYID

  - The unique identifier of a company
- FIELDDATE
  - The timestamp of the ESG score collection
- PRICEDATE
  - The closing price of a specified date

3.2 Machine Learning Architecture and Explainable AI

In this section, the methodology developed to train and evaluate the machine learning model used in this research is presented. As previously stated, the code implementation is written in Python and utilises various libraries and frameworks, such as CatBoost (Prokhorenkova et al., 2018), scikit-learn (Buitinck et al., 2013), and SHAP (Lundberg & Lee, 2017). The preprocessed data, in its tidy format and ordered by date, is split into training, validation and test sets using a custom function. To further illustrate, a baseline splitting date (2021-05-27) is derived from the nan analysis, ensuring that no missing data is leaked from training to test set due to backwards missing values fill. Ultimately, to avoid overfitting, 10% of training data is kept as a validation set.

The machine learning model used in this study is CatBoostRegressor, which is a gradient boosting algorithm (Prokhorenkova et al., 2018). Training and inference procedure as follows:

- Initialising a vanilla CatBoostRegressor with RMSE as loss function
- Fitting the model to the training data using the "fit" method, with the evaluation set provided for early stopping
- Adding the model predictions as a feature to the test set for evaluation

Predictive performances on both training and test set are evaluated using Root Mean Squared Error (RMSE). Mathematically, this metric represents the standard deviation of the residuals (the distance between the regression line and the data points).

From a technical standpoint, with the primary objective of validating the relationship between ESG data and stock market performance using machine learning methodologies, CatBoost was selected to cope with the variety of data-related challenges the research question poses. Namely, large size of available data, non-linearity of the objective function and curse of dimensionality. As proven by empirical studies (Prokhorenkova et al., 2018), the selected algorithm is thoroughly optimized for efficient model training, while keeping state-of-the-art performances on tabular data (Shwartz-Ziv & Armon, 2022). Thus, providing effective and efficient scaling capabilities during training and inference.

To further illustrate the rationale behind our algorithmic decision, prior research has underscored the absence of a linear relationship in financial data, as noted by Omran and Ragab (2004). Consequently, to effectively address these threats within our predictive pipeline, the utilisation of a robust non-linear model, such as CatBoost, becomes imperative. Besides, as we project the array of input ESG-related features into a multidimensional space, we encounter the phenomenon known as "curse of dimensionality" (COD). The idea was originally expounded by the mathematician Richard Bellman (1957) in the context of approximation theory. In the realm of data analysis, COD represents a significant hurdle in unveiling underlying patterns or structures within datasets characterized by a high number of variables. These challenges include increased data sparsity, computational complexity and a higher risk of overfitting models to the training data. Hence, the deployment of appropriate Machine Learning methodologies is essential to grapple with the inherent non-linear and COD-related complexities and challenges this use case presents.

Ultimately, extracting explainability-related insights, regarding feature relevance and Machine Learning output, is crucial to better inform the study. As a result, two distinct yet intercorrelated analyses are performed. Namely, boxplot prediction visualisation over K-clusters and SHAP assessment. To introduce more details, K-means is a widely used clustering algorithm that partitions data points into K distinct clusters by iteratively assigning points to the nearest centroid, while updating them until convergence. Likewise, SHapley Additive exPlanations values (SHAP), based on the game theoretically optimal "Shapley values", are computed to explain the impact of each feature on the model's predictions (Fig. 2).

4 Machine Learning-Driven Findings

The paper employs a comprehensive methodology to explore the relationship between ESG factors, stock price dynamics, and related variables in the S&P 500 index. Data from the Morningstar dataset ("ESG Risk Ratings & Variables") is utilized to provide a broad range of indices for major asset classes. Despite the variability in the disclosure frequencies of Morningstar variables (annual or monthly), the study relies on decomposed risk factors, incorporating the entire spectrum of risk measures (12 in total) rather than a single aggregated one.

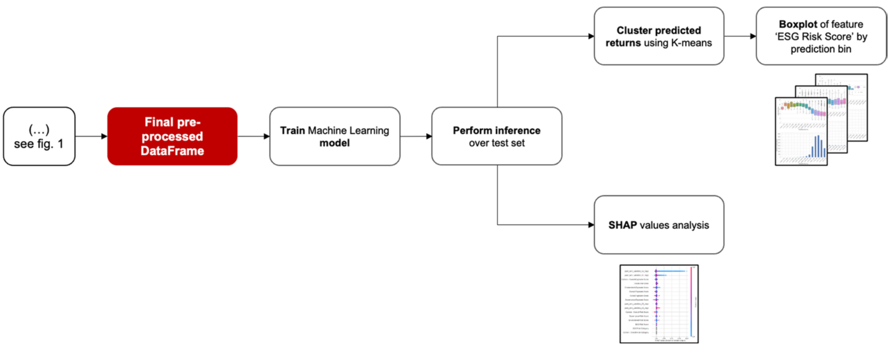


Fig. 2 Machine learning training and AI explainability pipeline diagram

From a technical standpoint, the methodology is fully implemented in Python, utilizing libraries such as CatBoost, scikit-learn, and SHAP. Preprocessed data, organized and sorted by date, is divided into training, validation, and test set, ensuring the absence of losses or leaks of missing data during the split. Furthermore, overfitting is prevented by retaining 10% of the training data as a validation set.

The selected machine learning model is CatBoostRegressor, a gradient boosting algorithm that has proven its efficacy in tabular data scenarios, like the one presented in this research (Shwartz-Ziv & Armon, 2022). As previously stated, the choice of CatBoost is motivated by the need to address specific challenges in financial data, such as large data size, non-linearity of the objective function and curse of dimensionality. Likewise, it is emphasized that the absence of a linear relationship in financial data justifies the use of a non-linear model like CatBoost.

Model training follows a standard procedure, initializing CatBoostRegressor with Root Mean Squared Error (RMSE) as loss function; evaluation set is kept to trigger early stopping and thus avoid overfitting. Ultimately, the model's predictive performance is estimated using RMSE while performing inference on both training and test set.

As a first result, the RMSE stands at 0.096 on the training set and 0.11 on the test set. These results demonstrate that the model, while potentially further optimizable through hyperparameter search, is indeed capable of retaining both predictive and generalization capabilities. Thus, the selected machine learning framework effectively addresses the aforementioned challenges; namely, non-linearity, curse of dimensionality and data complexity (as theoretically expected).

As shown in Fig. 2, running inference over the out-of-sample test set yields algorithmic results explained through two correlated analyses:

(1) Predictions visualization through boxplots on K-clusters (Fig. 3), which clarifies:

- The distribution of predicted performances
- The interaction between forecasted financial performances and actual ESG rating

(2) SHAP assessment (Figs. 4, 5, 6 and 7), employed to outline:

- Identification of key drivers influencing the variance in predicted financial performances and their correlation with actual ESG ratings
- Individual feature contributions to each prediction, enabling a deeper understanding of model decisions and consistent patterns

The objective of both analyses is to enhance comprehension of data and model interpretability, specifically focusing on elucidating the connection between ESG data and market performance.

Regarding the first analysis, Fig. 3 outlines variations in ESG scores across three clusters—split according to forecasted performances—and within different forecasted performance categories. To provide additional clarity, the figure also presents the distributions of ESG scores and the corresponding counts of data points for each group and prediction bin. Consistent with prior research

Cluster 0:	Cluster 1:	Cluster 2:
• Baseline market • Range of predicted returns (-2; 2.3]	• Mid-to-high performances • Range of predicted returns (2.3; 5.7]	• Best-in-class • Range of predicted returns (5.7; 24.9]

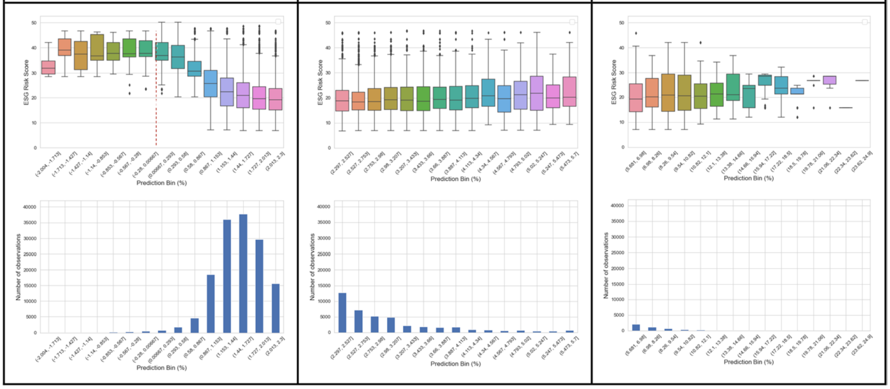


Fig. 3 For each forecasted performance cluster, boxplot of feature 'ESG Risk Score' by prediction bin (above) and Count of samples by prediction bin (below)

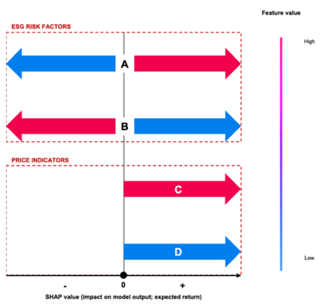
Domain	Hidden functional relationship	Symbol	
ESG Risk factors	Propensity (high risk and expected return)	A	
	Aversion (High risk and low expected return)	B	
Price indicators	Momentum	C	
	Mean-reverting	D	

Fig. 4 Summary table to interpret SHAP values and functional relationship through domain-specific knowledge

highlighting a positive relationship between ESG compliance and corporate performance (Carnini et al., 2022), as forecasted return increases, indicating an improved expected financial outlook, the ESG score steadily decreases and then stabilizes within cluster 0. More importantly, the latter cluster represents a baseline outlook of the market, encompassing 76% of all observations. Likewise, a clear threshold initiating a steady decrease of the ESG score curve is detected

SHAP relevance rank	Feature name	Functional relationship symbol
1	past_perc_variation_50_days	D
2	past_perc_variation_21_days	D
4	Social-Risk Score	A
5	Environment-Exposure Score	B
10	past_perc_variation_40_days	C

Fig. 5 Illustrative example over 5 features

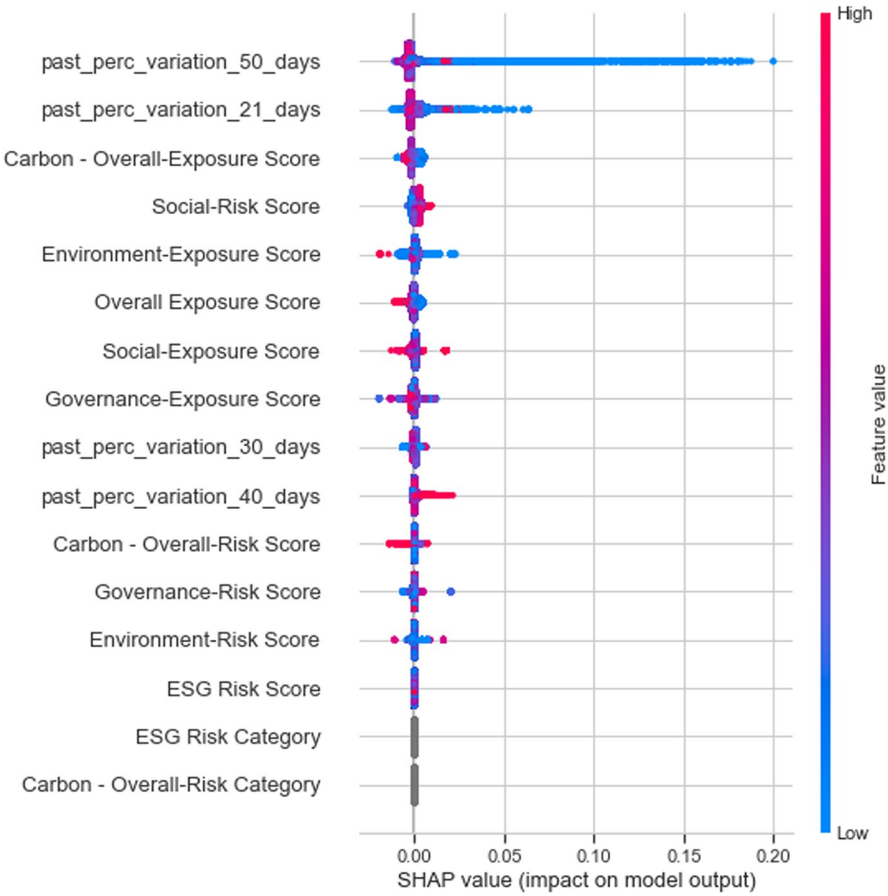


Fig. 6 Shap summary plot for whole input space

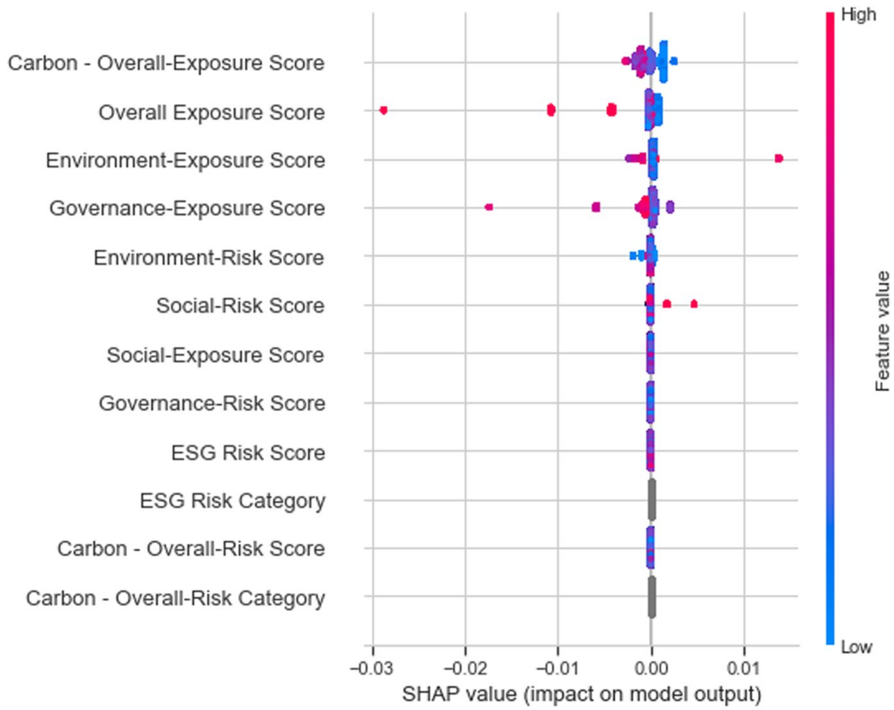


Fig. 7 Shap summary plot for ESG features only

around a 0% return (see dotted red line in Fig. 3, cluster 0). Additionally, clusters 1 and 2 reveal that, even considering abnormal returns, the ESG grade can either be maintained at the same level or improved. Consequently, this result appears in sharp contrast to the assumption of negative effects of ESG practices on companies' performance (Bolton & Kacperczyk, 2021; Nelling & Webb, 2008).

Concerning the second analysis, to further enhance the assessment of Shapley analysis results, it is crucial to incorporate machine learning-driven concealed functional connections for each individual feature with domain knowledge. To achieve this objective, Fig. 4 elucidates the association between functional interpretation and domain knowledge for both ESG factors and price indicators. For illustrative purposes, if an observation reports high ESG values (high unmanaged risk) and low SHAP (decreased expected returns)—as in the red case "B" from Fig. 4—the hidden functional relationship the model is exploiting is risk aversion: if ESG risks had been fully managed, it would have improved financial performance.

Additionally, an illustrative example involving five sampled features is presented below (Fig. 5). From a Machine Learning perspective, algorithms consist of a chain nested functions that map inputs to outputs. Therefore, unlike statistical approaches, there is no direct linear relationship between input and output; instead, a collection of step functions SHAP analysis aims to unveil. As a main consequence, we will solely refer to SHAP-driven functional relationships.

More in detail, SHAP summary plot (Figs. 4, 6 and 7) is a type of visualization designed to enhance the global interpretability of complex models. It achieves this by illustrating to which extent each feature in a model contributes to all predictions, quantified by SHAP values. These values, based on game theory, indicate the importance and direction of each feature's impact on model's output (Lundberg & Lee, 2017).

Key uses of the SHAP summary plot include:

- Identifying Influential Features, for interpretability and predictive purposes
- Clarifying Feature Interactions to explain influence over model's predictions
- Improving Transparency, for model accountability purposes

Based on these assumptions—SHAP values represent the relationship between a feature's value and model output—Fig. 6 introduces an overview over the whole input space (SHAP summary plot). Accordingly, the visualisation sorts features by the sum of SHAP value magnitudes, over all data points, and uses them to correlate features' impact distribution and algorithm's prediction. Ultimately, feature value is represented on a colour scale ranging from red (high) to blue (low).

Key insights include the notable mean-reverting behaviour of some lagged features, since low percentage variations from the past 50 and 21 days are positively related to an increase in the expected return. Conversely, percentage shifts from the last 40 days show a sustained momentum. In a similar manner, Environment-Risk score outlines a risk-averse behaviour (Kyaw et al., 2022); meanwhile, feature Social-Risk (Semet, 2020) suggests a positive relationship with higher grades of exposure. Although the latter result seems counterintuitive and only related to edge cases, it is indeed aligned with common financial dynamics; for instance, price momentum or risk management for expected future returns. Under these conditions, companies may induce investors to bear modest risks to obtain mid-to-long term performance improvements.

Equally important, since features are sorted in descending model relevance, there is a clear prevalence of risk measures as opposed to lagged indicators. To name a few, "Carbon", "Social", "Environment", "Overall" and "Governance" risk scores all appear before the second group of price variation KPIs. Ultimately, proving that, although a mixture of both types of features is preferred for forecasting purposes, predictive power is indeed retained within the ESG-related input space. Furthermore, the latter result is consistent with ESG-only SHAP visualization (Fig. 7).

Through high-granularity ESG and financial data, as well as and innovative non-linear ML methodologies, our results theoretically align and foster previous theories supporting the relationship between environmental practices and investor attraction (Hanley et al., 2016; Walley & Whitehead, 1994) and studies that identified a prior correlation between environmental practices, financial performance, and firm value (Semenova et al., 2019).

Moreover, the present research is coherent with the branch of literature asserting a positive relationship between ESG indices, corporate performance, and stock returns (Gillan et al., 2010; Borghesi et al., 2014; Serafeim, 2014; Bajic and Yurtoglu, 2016; Chelawat & Trivedi, 2016; Agyemang & Ansong, 2017; Garcia et al., 2017; Zhao et al., 2018; Aboud & Diab, 2019; Dalal & Thaker, 2019; Ikram et al., 2019; Alareeni & Hamdan, 2020; Shahbaz et al., 2020; Bansal et al., 2021; Hwang et al., 2021; Huang, 2021). Besides, the predictive power of ESG ratings may be framed within the efficient market hypothesis, particularly in its semi-strong form, suggesting that all publicly available information, including ESG ratings, is reflected in stock prices (Feng et al., 2021).

More generally, the study provides evidence in line with the stakeholder value orientation theory (Saini et al., 2023), which posits that companies considering the interests of all stakeholders, not just shareholders, ultimately achieve better long-term success. However, it is not guaranteed that ethical behavior is linked to positive returns on stock prices. Building upon this idea, other findings suggest that there may be different relationships between ESG and stock prices for specific sectors (Semenova & Hassel, 2019). For example, the analysis by Semenova & Hassel (2016) demonstrated an absence of correlation in various industries tested in their analysis, and one of the main differences that could explain the divergence in results is that these authors tested only environmental practices and not all ESG metrics, as in the present piece of research.

Ultimately, to gain a more comprehensive and detailed insight into the link between ESG indices and stock returns, this study diverges from previous approaches that relied on synthetic indices (like those in La Torre et al., 2020). Instead, it adopts an innovative methodology by utilizing the extensive and high-granularity dataset from Morningstar's "ESG Risk Ratings & Variables." This dataset provides a wide array of indices for all major asset classes among companies listed in the S&P 500, offering a more robust and varied foundation for analysis.

5 Conclusion

In this paper we reviewed, through Machine Learning lenses how ESG factors can lead a portfolio selection to create alpha. Various studies have examined the relationship between ESG factors and corporate performance, highlighting both positive and negative perspectives. Positive findings suggest that companies with strong ESG profiles tend to have higher stock prices, improved financial performance and greater market value. However, some researchers argue that pursuing ESG goals may hinder a company's primary goal of increasing shareholder wealth. However, as shown during the COVID-19 pandemic, ESG factors can serve as a defence mechanism during crises.

As sustainability becomes a paramount objective, companies are reorienting their strategies towards environmental preservation, social responsibility, and improved governance practices. Consequently, equity portfolio managers are confronted with the challenge of identifying assets that yield optimal returns, given ESG constraints.

In brief, this study leveraged domain specific, multivariate and unpublished data—as opposed to a single descriptive ESG KPI—as well as cutting-edge techniques. More precisely, the uniqueness of our approach stems from the application of high-performance non-linear machine learning techniques to a rich, multi-dimensional dataset, which includes both ESG ratings and a variety of lagged indicators.

This combination allows for a more nuanced understanding of the intricate dynamics between ESG factors and stock returns, thereby answering the central research question. In the latter regard, our findings not only contribute new insights to the existing body of knowledge but also challenge prevailing assumptions about the negative impact of ESG practices on financial performance. By leveraging advanced algorithms and a comprehensive dataset, our study bridges the gap in literature and provides empirical evidence that refutes the notion of a trade-off between ESG compliance and shareholder wealth maximization.

From a technical standpoint, the deployment of a newly-released ML algorithm was required to cope with the variety and complexity of the input space, as well as data volume. As a main consequence, greater specificity is ensured by both aspects—data and technique—allowing a deeper understanding of the intricate relationship between ESG risk factors and trading performances. To discern the interplay between these variables, we assessed the predictive capabilities of both ESG ratings and lagged indicators on stock returns. The findings suggest that ESG factors hold a pivotal role in portfolio selection and can potentially generate alpha. Moreover, the employed ML model reveals that ESG scores do not impede the achievement of superior market performance. Notably, ESG indicators exhibit a significant relevance in terms of feature importance, often reflecting a risk-averse behaviour. As a result, compelling evidence contradicts the assumption of detrimental effects of ESG practices on firm performance (Ruan and Liu, 2021).

In conclusion, this study offers valuable insights for practitioners in the investment industry, providing empirical evidence of the impact of ESG ratings on stock prices and their predictive power. The framework introduced in this research can guide the creation of ML-driven and ESG-driven portfolios that ensure positive performances, enhancing the integration of ESG factors into investment decision-making.

In our analysis, it is important to acknowledge limitations that may influence the scope and robustness of our findings. Firstly, as a direct consequence of the low frequency of ESG data publication, timeliness issues may be present. Additionally, the study may exhibit geographical and sectoral bias, limiting its generalizability (Semenova & Hassel, 2019). The dynamic nature of ESG factors, evolving over time, poses a challenge in capturing their ongoing relevance accurately. Likewise, external factors such as macroeconomic conditions and regulatory changes may not have been fully considered, impacting the external validity of our conclusions. While our study primarily focuses on the long-term implications of ESG factors, the short-term dynamics and causative relationships may not be fully elucidated. These acknowledged limitations underscore the need for ongoing research and refinement of methodologies to enhance the depth and reliability of our insights into the role of ESG factors in investment decision-making.

In our ongoing research endeavors, we aim to expand the scope and depth of our analysis. As data availability and the historical range of ESG integration in markets improve, we intend to conduct a more comprehensive examination of these relationships across diverse contexts. This future research could provide further insights into the long-term implications and broader applicability of ESG factors in investment decision-making.

Funding Open access funding provided by Università degli Studi G. D'Annunzio Chieti Pescara within the CRUI-CARE Agreement. The authors have not disclosed any funding.

Declarations

Conflict of interets The authors have not disclosed any competing interests.

References

Aboud, A., & Diab, A. (2019). The financial and market consequences of environmental, social and governance ratings. *Sustainability Accounting, Management and Policy Journal, 10*(3), 498–520.

Ademi, B., & Klungseth, N. J. (2022). Does it pay to deliver superior ESG performance? Evidence from US S&P 500 companies. *Journal of Global Responsibility, 13*(4), 421–449.

Agyemang, O. S., & Ansong, A. (2017). Corporate social responsibility and firm performance of Ghanaian SMEs mediating role of access to capital and firm reputation. *Journal of Global Responsibility, 8*(1), 47–62.

Alareeni, B. A., & Hamdan, A. (2020). ESG impact on performance of US S&P 500-listed firms. *Corporate Governance: THe International Journal of Business in Society, 20*(7), 1409–1428.

Bajic, S., & Yurtoglu, B. B. (2016). CSR, market value, and profitability: international evidence. In S. Boubaker, D. Cumming, & D. C. Nguyen (Eds.), *Handbook of Finance and Sustainability*. Edward Elgar.

Bansal, M., Samad, T. A., & Bashir, H. A. (2021). The sustainability reporting-firm performance nexus: Evidence from a threshold model. *Journal of Global Responsibility, 12*(4), 491–512.

Beaver, W. H. (1968). Market prices, financial ratios, and the prediction of failure. *Journal of Accounting Research, 6*(2), 179–192.

Bellman, R. E. (1957). *Dynamic programming*. Princeton University Press.

Berg, F., Fabisik, K., & Sautner, Z. (2020). Rewriting History II: The (Un)Predictable past of ESG ratings. *European Corporate Governance Institute–finance Working Paper, 708*, 10–2139.

Bolton, P., & Kacperczyk, M. (2021). Do investors care about carbon risk? *Journal of Financial Economics, 142*(2), 517–549.

Borghesi, R., Houston, J., & Naranjo, A. (2014). Corporate socially responsible investments: CEO altruism, reputation, and shareholder interests. *Journal of Corporate Finance, 26*(3), 164–181.

Carnini Pulino, S., Ciaburri, M., Magnanelli, B. S., & Nasta, L. (2022). Does ESG disclosure influence firm performance? *Sustainability, 14*(13), 7595.

Chatterji, A. K., Durand, R., Levine, D. I., & Touboul, S. (2016). Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal, 37*(8), 1597–1614.

Chelawat, H., & Trivedi, I. V. (2016). The business value of ESG performance: The Indian context. *Asian Journal of Business Ethics, 5*(1/2), 195–210.

Cohen, G. (2023). ESG risks and corporate survival. *Environment Systems and Decisions, 43*(1), 16–21.

Dalal, K. K., & Thaker, N. (2019). ESG and corporate financial performance: A panel study of Indian companies. *IUP Journal of Corporate Governance, 18*(1), 44–59.

Duque-Grisales, E., & Aguilera-Caracuel, J. (2021). Environmental, social and governance (ESG) scores and financial performance of Multilatinas: Moderating effects of geographic international diversification and financial slack. *Journal of Business Ethics, 168*(2), 315–334.
Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science, 60*(11), 2835–2857.
Fatemi, A., Fooladi, I., & Tehranian, H. (2015). Valuation effects of corporate social responsibility. *Journal of Banking & Finance, 59*, 182–192.
Feng, J., Goodell, J. W., & Shen, D. (2021). ESG rating and stock price crash risk: Evidence from China. *Finance Research Letters, 46*, 102476.
Friedman, M. (2009). *Capitalismand Freedom*. University of Chicago Press.
Galbreath, J. (2013). ESG in focus: The Australian evidence. *Journal of Business Ethics, 118*, 529–541.
Garcia, A. S., Mendes-Da-Silva, W., & Orsato, R. J. (2017). Sensitive industries produce better ESG performance: Evidence from emerging markets. *Journal of Cleaner Production, 150*, 135–147.
Hanley, N., Shogren, J. F., & White, B. (2016). *Environmental economics: In theory and in practice*. Macmillan.
Huang, D. Z. X. (2021). Environmental, social and governance (ESG) activity and firm performance: A review and consolidation. *Accounting and Finance, 61*(1), 335–360.
Hwang, J., Kim, H., & Jung, D. (2021). The effect of ESG activities on financial performance during the COVID-19 pandemic - evidence from Korea. *Sustainability, 13*(20), 11362.
Ikram, M., Sroufe, R., Mohsin, M., Solangi, Y. A., Shah, S. Z. A., & Shahzad, F. (2019). Does CSR influence firm performance? A longitudinal study of SME sectors of Pakistan. *Journal of Global Responsibility, 11*(1), 27–53.
Jayachandran, S., Kalaignanam, K., & Eilert, M. (2013). Product and environmental social performance: Varying effect on firm performance. *Strategic Management Journal, 34*(10), 1255–1264.
Jiao, Y. (2010). Stakeholder welfare and firm value. *Journal of Banking and Finance, 34*(10), 2549–2561.
Kim, K. H., Kim, M., & Qian, C. (2018). Effects of corporate social responsibility on corporate financial performance: A competitive-action perspective. *Journal of Management, 44*(3), 1097–1118.
Kim, S., & Yoon, A. (2023). Analyzing active fund Managers' commitment to ESG: Vidence from the United Nations principles for responsible investment. *Management Science, 69*(2), 741–758.
Konar, S., & Cohen, M. A. (2001). Does the market value environmental performance? *Review of Economics and Statistics, 83*(2), 281–289.
Kyaw, K., Olugbode, M., & Petracci, B. (2022). Stakeholder engagement: Investors' environmental risk aversion and corporate earnings. *Business Strategy and the Environment, 31*(3), 1220–1231.
La Torre, M., Mango, F., Cafaro, A., & Leo, S. (2020). Does the ESG index affect stock return? Evidence from the Eurostoxx50. *Sustainability, 12*, 6387.
Landi, G., & Sciarelli, M. (2018). Towards a more ethical Market: The impact of ESG rating on corporate financial performance. *Social Responsibility Journal, 15*(1), 11–27.
Lapinskienė, G., Gedvilaitė, D., Liučvaitienė, A., & Peleckis, K. (2023). How does environmental data from ESG concept affect stock returns: Case of the European Union and US capital markets. *Emerging Science Journal, 7*(2), 410–427.
Leung, W. S., Mazouz, K., Chen, J., & Wood, G. (2018). Organization capital, labor market flexibility, and stock returns around the world. *Journal of Banking and Finance, 89*, 150–168.
Li, S., Yin, P., & Liu, S. (2022). Evaluation of ESG ratings for chinese listed companies from the perspective of stock price crash risk. *Frontiers in Environmental Science, 10*, 933639.
Liu, X., & Wan, D. (2023). Retail investor trading and ESG pricing in China. *Research in International Business and Finance, 65*, 101911.
Marsat, S., Pijourlet, G., & Ullah, M. (2021). Does environmental performance help firms to be more resilient against environmental controversies? *International Evidence. Finance Research Letters, 44*, 102028.
Minor, D., & Morgan, J. (2011). CSR as reputation insurance: Primum non nocere. *California Management Reviev, 53*(3), 40–59.
Nelling, E., & Webb, E. (2008). Corporate social responsibility and financial performance: The "virtuous circle" revisited. *Review of Quantitative Finance and Accounting, 32*(2), 197–209.
Nguyen, D. T., Hoang, T. G., & Tran, H. G. (2022). Help or hurt? The impact of ESG on firm performance in S&P 500 non-financial firms. *Australasian Accounting, Business and Finance Journal, 16*(2), 91–102.

Nicholson, G., Kiel, G., & Kiel-Chisholm, S. (2011). The contribution of social norms to the global financial crisis: A systemic actor focused model and proposal for regulatory change. *Corporate Governance: An International Review, 19*(5), 471–488.

Omran, M., & Ragab, A. (2004). Linear versus non-linear relationships between financial ratios and stock returns: Empirical evidence from Egyptian firms. *Review of Accounting and Finance, 3*(2), 84–102.

Rodionova, M., Skhvediani, A., & Kudryavtseva, T. (2022). ESG as a booster for logistics stock returns—evidence from the us stock market. *Sustainability, 14*(19), 12356.

Rossi, F., & Harjoto, M. A. (2020). Corporate non-financial disclosure, firm value, risk, and agency costs: Evidence from Italian listed companies. *Review of Managerial Science, 14*(5), 1149–1181.

Ruan, L., & Liu, H. (2021). Environmental, social, governance activities and firm performance: Evidence from China. *Sustainability, 13*(2), 767.

Saini, M., Aggarwal, V., Dhingra, B., Kumar, P., & Yadav, M. (2023). ESG and financial variables: A systematic review. *International Journal of Law and Management, 65*(6), 663–682.

Semenova, N., & Hassel, L.G. (2016). The moderating effects of environmental risk of the industry on the relationship between corporate environmental and financial performance. *Journal of Applied Accounting Research, 17*(1), 97–114.

Semenova, N., & Hassel, L. (2019). The performance of investor engagement dialogues to manage sustainability risks. *The Nordic Journal of Business, 68*(2), 5–22.

Serafeim, G. (2014). Turning a profit while doing good: Aligning sustainability with corporate performance. *Brookings, 32*(2), 122–126.

Serafeim, G., & Yoon, A. (2023). Stock price reactions to ESG news: The role of ESG ratings and disagreement. *Review of Accounting Studies, 28*(3), 1500–1530.

Shahbaz, M., Karaman, A. S., Kilic, M., & Uyar, A. (2020). Board attributes, CSR engagement, and corporate performance: What is the nexus in the energy sector? *Energy Policy, 143*, 111582.

Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion, 81*, 84–90.

Umar, M., Mirza, N., Rizvi, S. K. A., & Naqvi, B. (2022). ESG scores and target price accuracy: Evidence from sell-side recommendations in BRICS. *International Review of Financial Analysis, 84*, 102389.

Walley, N., & Whitehead, B. (1994). It's not Easy Being Green. *Reader in Business and the Environment, 36*(81), 4.

Wu, C., Xiong, X., & Gao, Y. (2022). Does ESG certification improve price efficiency in the Chinese stock market? *Asia-Pacific Financial Markets, 29*(1), 97–122.

Xiong, J. X. (2021). The Impact of ESG Risk on Stocks. *Journal of Impact and ESG Investing, 2*(1), 7–18.

Zhao, C., Guo, Y., Yuan, J., Wu, M., Li, D., Zhou, Y., & Kang, J. (2018). ESG and corporate financial performance: Empirical evidence from China's listed power generation companies. *Sustainability, 10*(8), 2607.

Zhou, D., & Zhou, R. (2022). ESG performance and stock price volatility in public health crisis: Evidence from COVID-19 pandemic. *International Journal of Environmental Research and Public Health, 19*(1), 202.

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., & Varoquaux, G. (2013). API design for machine learning software: experiences from the scikit-learn project. *arXiv preprint* arXiv:1309.0238.

Gillan, S., Hartzell, J. C., Koch, A., & Starks, L. T. (2010). Firms' environmental, social and governance (ESG) choices, performance and managerial motivation. *Unpublished Working Paper, 10*.

Hong, H. G., Kubik, J. D., Liskovich, I., & Scheinkman, J. A. (2019). Crime, punishment and the value of corporate social responsibility. *Available at SSRN 2492202*.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Moahmed, T. A., El Gayar, N., & Atiya, A. F. (2014). Forward and backward forecasting ensembles for the estimation of time series missing data. In *Artificial Neural Networks in Pattern Recognition: 6th IAPR TC 3 International Workshop*, ANNPR 2014, Montreal, QC, Canada, October 6–8, 2014. Proceedings 6 (93–104). Springer International Publishing.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in Neural Information Processing Systems, 31*.

Semet, R. (2020). The Social Issue of ESG Analysis. *Available at SSRN 3838372*.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.