North American Journal of Economics and Finance 80 (2025) 102517

Contents lists available at ScienceDirect

North American Journal of Economics and Finance

journal homepage: www.elsevier.com/locate/najef

# Outperforming ESG stocks portfolio: A machine learning ranking model with catboots regressor

Vittorio Carlei, Donatella Furia *, Alessandro Ceccarelli, Piera Cascioli

*University of Chieti-Pescara, Pescara 65127, Italy*

ARTICLE INFO

*Jel classification:*
G11
G17
Q56

*Keywords:*
ESG investing
Alpha generation
Machine learning
Cat boots regressor

ABSTRACT

The paper investigates whether outperforming ESG (Environmental, Social, and Governance) stocks can generate alpha over the S&P 500 through a machine learning (ML)-driven simulation. Leveraging advanced ML techniques, particularly the CatBoostRegressor model, the study explores the relationship between ESG factors and financial performance to construct a high-performing, ESG-compliant portfolio.

The findings suggest that incorporating ESG criteria not only supports sustainable investing but also enhances return predictability. However, backtesting reveals that portfolios more tolerant of lower ESG scores tend to yield higher short-term returns. A customizable threshold for ESG criteria allows the strategy to accommodate varying risk appetites, demonstrating that strategic flexibility can align ESG values with financial goals. This work highlights the transformative role of machine learning in sustainable investing, illustrating how ethical considerations and profitability can effectively coexist within a robust, ML-powered framework.

## 1. Introduction

The Efficient Market Hypothesis (EMH) posits that asset prices fully reflect all available information, implying that outperforming the market systematically without assuming additional risk is theoretically impossible. However, in recent decades, the integration of Environmental, Social, and Governance (ESG) factors into investment decision-making has introduced new dimensions to traditional asset pricing models (Fama and French 2015), suggesting that non-financial information could enhance the predictability of returns (Friede et al., 2015; Serafeim & Yoon, 2023). ESG integration not only embodies ethical and sustainable corporate practices but also represents a potential source of alpha (a measure of the excess return of an investment relative to the market or a benchmark, adjusted for risk) generation through the identification of underpriced sustainable firms (Krüger, 2015; Huang, 2021).

Recent empirical studies have shown that firms with high ESG ratings tend to display greater resilience to financial crises, better operational efficiency, and lower costs of capital (Fatemi et al., 2018; Broadstock et al., 2021). Consequently, sustainable investing strategies have attracted growing attention from both academic researchers and financial practitioners, leading to the development of ESG-focused portfolio construction techniques aimed at combining profitability with responsibility.

At the same time, the rise of Artificial Intelligence (AI) and Machine Learning (ML) has profoundly reshaped the landscape of quantitative finance. Recent contributions, such as Gu et al. (2020) and Kelly & Xiu (2023), highlight how machine learning models —

* Corresponding author.
*E-mail addresses:* vittorio.carlei@unich.it (V. Carlei), donatella.furia@unich.it (D. Furia), alessandro.ceccarelli@pharmadvice.it (A. Ceccarelli), piera.cascioli@unich.it (P. Cascioli).

https://doi.org/10.1016/j.najef.2025.102517
Received 8 November 2024; Received in revised form 8 May 2025; Accepted 28 July 2025
Available online 6 August 2025

thanks to their ability to capture complex, non-linear relationships — outperform traditional econometric models in asset pricing tasks. In particular, the use of ensemble methods, attention-based architectures, and autoencoders has allowed researchers to identify latent factors and cross-sectional patterns that traditional factor models fail to capture (Chen et al., 2023; Tepelyan & Gopal, 2023).

However, despite these advances, most existing studies applying Machine Learning to ESG investing have relied on aggregated ESG scores, potentially missing relevant information hidden in the sub-components of ESG evaluations (La Torre et al., 2020; Umar et al., 2022). Moreover, while some works have attempted to forecast returns using ESG data, few have developed systematic, ESG-compliant portfolio construction frameworks powered by machine learning models, especially using backtesting procedures that flexibly accommodate varying sustainability thresholds.

This study addresses these gaps by proposing an innovative machine learning-driven approach for ESG investing, characterized by three main contributions: first, unlike most previous studies that rely on aggregate ESG scores, this research utilizes disaggregated ESG risk variables from the Morningstar "ESG Risk Ratings & Variables" dataset, enhancing the granularity and predictive power of the input features; second, we deploy a CatBoostRegressor — a gradient boosting model optimized for tabular and categorical data — to predict stock-specific future returns based on ESG sub-indicators and past price variations, overcoming the limitations of linear and traditional models, and third, a custom backtesting methodology is developed, enabling dynamic portfolio construction based on predicted returns and ESG compliance thresholds. The framework allows flexible ESG constraint adjustment, reflecting different investor risk appetites, while systematically evaluating alpha generation relative to the S&P 500 benchmark.

In particular, this study proposes an innovative ESG-driven stock picking framework (Carlei et al., 2021; Wolff & Echterling, 2024), where the machine learning algorithm predicts the future returns of individual stocks and dynamically selects those with the best return prospects, while ensuring compliance with sustainability criteria. This approach differs from traditional ESG portfolio construction methods, which typically rely on simple inclusion or exclusion based on aggregate thresholds, by offering an active stock selection strategy optimized along both financial and ESG dimensions. In fact, the objective is to demonstrate that active stock management, guided by granular ESG signals and supported by advanced machine learning techniques, can generate alpha relative to the benchmark while maintaining an adequate sustainable risk profile.

Thus, the novelty of this research lies in the integration of fine-grained ESG data with a robust machine learning architecture and a tailored backtesting procedure, illustrating how sustainable investing can be enhanced through advanced predictive analytics.

Unlike prior ESG studies focused on performance analysis or static filtering, this work introduces an adaptive, ML-based portfolio construction framework that explicitly models return predictability conditioned on ESG risk decomposition, conditioned on ESG risk decomposition, enabling a more dynamic and responsive investment process.

The remainder of the paper is structured as follows: Section 2 reviews the relevant literature on ESG investing and Machine Learning applications in asset pricing. Section 3 describes the empirical methodology and data processing pipeline. Section 4 presents the backtesting results and comparative performance analyses. Section 5 concludes with implications and future research directions.

## 2. Literature review

The theory of financial market efficiency, introduced by Eugene in Fama (1970), posits that asset prices always reflect all available information. In an efficient market, investors cannot achieve returns above the average without taking on greater risk, as prices quickly adjust to new information. This concept has significantly influenced investment strategies and the analysis of financial performance.

In recent years, the introduction of ESG (Environmental, Social, Governance) indices has transformed how financial performance is evaluated.

This shift has also prompted the development of data-driven models that combine ESG scoring with portfolio optimization, including machine learning approaches focused on explainability and alpha prediction (Carlei et al., 2024; Sorathiya et al., 2024). In particular, Mohanty et al. (2021) demonstrate that applying an ESG overlay to multifactor investment strategies, especially those based on value, low volatility, and quality factors, enhances alpha while reducing both systematic and idiosyncratic tail risks.

Recent empirical findings also emphasize that ESG stocks respond asymmetrically to market sentiment, geopolitical tensions, and crypto-related uncertainty, demonstrating both hedging and diversification properties under specific quantile regimes (Yang et al., 2024).

Empirical results also show that firms with higher ESG scores exhibit more accurate valuation forecasts by financial analysts, reinforcing the predictive value of ESG dimensions (Umar et al., 2022).

Investors are paying increasing attention to the environmental, social, and governance impacts of companies, recognizing that these factors can affect long-term sustainability and stable returns. The inclusion of ESG criteria in investment decisions reflects a growing awareness of the importance of managing non-financial risks and promoting ethical corporate behaviour, with positive effects on reputation and economic outcomes.

Consequently, participating companies have adjusted their activities and corporate structures to align with the new paradigm of ESG criteria implementation (Beaver, 1968). The literature on asset pricing models, stock selection, and portfolio construction (Wolff and Echterling, 2024) has thus expanded to include new strategies aimed at increasing the profitability of an equity portfolio, exploring the relationship between sustainable corporate performance and the evaluation of ESG criteria (Hong et al., 2019; Carlei et al., 2024).

Some recent studies go beyond classical models, applying Bayesian optimization and reinforcement learning to maximize risk-adjusted returns in ESG-constrained portfolios (Garrido-Merchán et al., 2023a; Acero et al., 2024).

Additionally, Steuer and Utz (2023) propose a tri-objective optimization framework where ESG metrics are explicitly integrated into the efficient frontier, transforming it into a three-dimensional efficient surface to capture the trade-off between financial and sustainability objectives.

The advent of technologies such as Artificial Intelligence (AI) and advanced machine learning techniques, such as autoencoders (Tepelyan & Gopal, 2023) and attention layers (Chen et al., 2023), is transforming the traditional paradigm (Gu et al., 2020). Nonetheless, as highlighted by Israel, Kelly, and Moskowitz (2020), machine learning applications in finance face distinct challenges due to small datasets, low signal-to-noise ratios, and evolving market dynamics, which necessitate careful model design and validation.

In the financial context, as Gu et al. (2020) state, Machine Learning, compared to traditional empirical methods used for asset pricing, offers the possibility to consider a much broader set of predictive variables and to adopt more sophisticated functional specifications, allowing for greater flexibility to expand the boundaries in measuring risk premiums.

The analysis of the relationships among different securities reveals a deeply different market structure, enabling specific securities to be treated distinctly from others. "Stock picking," in fact, unlike the classic forecast of expected returns (a key theme in the theory of efficient markets), is a process focused on identifying specific opportunities within the market rather than adhering to a uniform view of expected returns (Carlei et al., 2021; Kelly, 2023; Li et al., 2023; Wolff & Echterling, 2024).

At the level of empirical analysis, before the introduction of Machine Learning, traditional models like Sharpe (1964) and Fama and French (2015) were used to identify excess returns in portfolios. Today, advanced Machine Learning models surpass the statistical insignificance limitations present in linear models, such as the q-factor (Hou et al., 2018), capturing both linear and non-linear relationships between factors and returns. Similarly, tools like support vector machines (SVM) and neural networks have improved the accuracy of financial forecasts and portfolio optimization (Novak & Veluščcek, 2016; Chen & Ge, 2021 while more recent approaches, such as GANs, autoencoders (Gu et al., 2020), deep reinforcement learning (Acero et al., 2024) and Bayesian optimization under ESG constraints (Garrido-Merchán et al., 2023b), have demonstrated the potential to generate superior risk-adjusted returns in ESG-focused portfolio strategies, albeit with challenges related to interpretability and implementation complexity, while recent studies also emphasize the importance of ESG signal deconstruction and model transparency in AI-based investing (Ehlers et al., 2024; Enders et al., 2025).

This work falls within the research stream exploring innovative ways to generate extra profits from a portfolio through the use of advanced Machine Learning techniques (Gu et al., 2020; Carlei et al., 2024; Ahmed et al., 2022; Ghosh et al., 2022; Kelly & Xiu, 2023), while simultaneously integrating sustainability principles. The objective is to leverage unconventional methodologies that optimize financial performance, considering environmental, social, and governance impacts, thereby opening new perspectives in sustainable wealth management. The integration of ESG criteria into quantitative investment models, particularly through rating aggregation and multi-source consensus approaches, has also shown measurable performance gains (Berg et al., 2024).

Additionally, rating dispersion and disagreement across ESG providers have been shown to influence portfolio risk and return asymmetrically, raising challenges and opportunities in portfolio design (Horn & Oehler, 2024).

## 3. Data and methodology

To elucidate the interplay between ESG factors, stock price dynamics, and correlated variables within the S&P 500 index, we employed data from the Morningstar dataset ("ESG Risk Ratings & Variables"). This dataset offers a comprehensive array of indices for all major asset classes. Notably, while Morningstar's variables are disclosed at varying frequencies (either yearly or monthly), our study leverages decomposed risk factors, incorporating the full spectrum of 12 risk measures rather than relying on a single aggregated figure. For visualization purposes, however, the unified ESG score can be used to avoid the curse of dimensionality when representing data points.

In the data preprocessing phase, we conducted a panel analysis to create a coherent and unified view from the following datasets:

- ESG.RiskRatings:
  - o Contains the final ESG Ratings for each company
- ESG.Catalog:
  - o A dictionary detailing variable names, descriptions, and possible values
- ESG.Variables:
  - o Contains the ESG variables
- SP500.Prices:
  - o Includes the daily closing prices in USD for each company in the S&P 500 from January 1, 2010, to February 21, 2023
- SP500.Companies:
  - o Contains records of S&P 500 companies, including their names, countries, and tickers from January 1, 2010, to February 21, 2023

### 3.1. Data engineering

We implemented a comprehensive data processing pipeline in Python to merge stock price data with ESG information over more than 13 years. To facilitate the backtesting pipeline, the process involves several interconnected steps designed to enhance the dataset. The outcome is a unified view of stock data, ESG scores, and rating descriptions, enabling further analysis and modeling from January 1, 2010, to January 1, 2023, with daily granularity. For clarity, we will use "ESG Risk Score" consistently to refer to a specific feature within the input space. Likewise, a detailed list of all input variables can be found in Chapter 3. However, ESG scores or values, in contrast, encompass a broader range of sustainability-related metrics.

To address missing ESG values across all risk measures, the pipeline uses forward and backward-fill techniques on each company's time series. This two-step method replaces missing values first with the most recent non-null value (forward fill) and then with the first non-missing value (backward fill). This approach ensures that all gaps are appropriately filled while avoiding data leakage, as detailed in subsequent sections. Given that ESG ratings are slow-moving predictors, the present research assumes that ratings remain steady over the years, reflecting real-world dynamics. To further support the assumption, the average standard deviation across all standardized ESG features is 0.022, indicating the stationarity of the variables.

The ESG risk rating data is then pivoted, resulting in a tidy dataset with a panel representation. This transformation allows the dataset to represent companies and dates as rows, with columns for the full range of detected risk measures. More in detail, the overall ESG risk rating applies the concept of risk decomposition to determine the degree of unmanaged risk for each corporation, assigning values between 0 and 100—where 0 indicates fully managed risks and 100 indicates the highest level of unmanaged risk. Lower values are preferable for sustainability. This indicator is computed as the difference between a company's overall exposure score and its managed risk grade, or by summing the Corporate Governance unmanaged risk figure with the company's total unmanaged risk value.

For trading data, to enhance its predictive capabilities, we calculate past and future price variations over specific windows. This involves computing price percentage changes between the current day and a set number of trading days for both past (21, 30, 40, and 50 days) and future (30 days) periods, with the future change serving as the target variable for supervised training.

Finally, the pipeline merges the augmented stock dataset with the tidy ESG scores, combining data based on company identifiers and detection dates. This process highlights the uniqueness and relevance of the analyzed dataset, which includes one of the largest financial indexes in the world – the S&P 500 – over more than 13 years (Fig. 1).

To enhance the interpretability of the diagram, below is an explanation of the pivoting procedure and pertinent field descriptions. A pivot table consolidates and structures data into a concise and informative format. As shown in the diagram, ESG scores are pivoted to produce a final DataFrame where each row corresponds to a specific data collection date ("FIELDDATE") for a given company ("ENTITYID"). The columns include all available ESG scores.

Key fields include:

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

### 3.2. Machine learning architecture and backtesting design

This section details the methodology employed to train and evaluate the machine learning model used in this research. The code implementation is in Python and leverages various libraries and frameworks, including CatBoost (Prokhorenkova et al., 2018), scikit-learn (Buitinck et al., 2013), and SHAP (Lundberg, 2017). The preprocessed data, organized in a tidy format and sorted by date, is divided into training, validation, and test sets using a custom function. Specifically, a baseline splitting date (2021–05-27) was established based on nan analysis to ensure no missing data from the training set leaks into the test set due to the backward filling of missing values. To further prevent overfitting, 10 % of the training data is reserved as a validation set and no shuffling is performed (Table 1).

The machine learning model utilized in this study is the CatBoostRegressor, a gradient-boosting algorithm (Prokhorenkova et al., 2018). The training and inference procedure includes:

1. Initializing a vanilla CatBoostRegressor with:
   a. Depth equal to 4 to avoid overfitting
   b. Training epochs capped at 500
   c. RMSE as the loss function
   d. Early stopping set to 50 iterations
2. Fitting the model to the training data using the "fit" method, with the evaluation set provided for early stopping
3. Adding the model predictions as a feature to the test set for evaluation

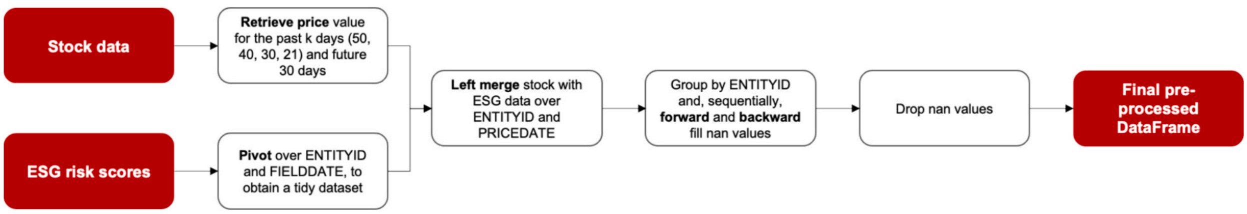


**Fig. 1.** Data pre-processing pipeline diagram.

**Table 1**
Data chronology and experimental split.

| Sub-set | Calendar window | Share of rows |
|---|---|---|
| Training | 2014–01-01 → 2021–05-26 | 90 % of rows before split |
| Validation | 2014–01-01 → 2021–05-26 | last 10 %, in chronological order |
| Test simulation | 2021–05-27 → 2023–12-31 | 100 % |

4. Computing predictive performance on the test set is also evaluated using Root Mean Squared Error (RMSE)
5. Saving the best-iteration model, freezing weights for all downstream tests

The primary objective is to validate the learnable relationship between ESG data and stock market performance using machine learning methodologies. CatBoost was chosen to address various data-related challenges, including the large size of available data, the non-linearity of the objective function, and the curse of dimensionality. This choice aligns with recent work by Ren et al. (2024), who show that adapting ensemble models with time- and class-weighted adjustments significantly enhances prediction robustness under concept drift and data imbalance—challenges that are pervasive in financial time-series applications. These modeling challenges reflect the structural complexities of financial data highlighted by Israel, Kelly, and Moskowitz (2020), who emphasize the importance of model robustness and out-of-sample validation in low signal-to-noise environments. Empirical studies (Prokhorenkova et al., 2018) have demonstrated that CatBoost is optimized for efficient model training while maintaining state-of-the-art performance on tabular data (Shwartz-Ziv & Armon, 2022). This makes it effective and efficient for scaling during training and inference. Likewise, as shown by prior research, financial data does not exhibit linear relationships (Omran and Ragab, 2004). Consequently, addressing these challenges within our predictive pipeline necessitates a robust non-linear model like CatBoost. More in detail, as we project the array of input ESG-related features into a multidimensional space, we encounter the "curse of dimensionality" (COD), a concept introduced by Richard Bellman in the context of approximation theory. In data analysis, COD represents a significant obstacle in uncovering underlying patterns or structures within datasets characterized by a high number of variables. Therefore, deploying appropriate machine learning methodologies is essential to manage the inherent non-linear complexities and challenges presented by this use case.

A central part of the present research is the introduction of a custom backtesting methodology (reported in Table 2), used to evaluate the performance of ESG stocks within the S&P 500 index. The primary objective is to determine whether top-performing ESG stocks can generate alpha over a specified period, employing a systematic and robust ML-driven approach.

The selected stocks are held for exactly 30 days. After this holding period, all stocks are sold and return R is computed using the formula presented in Fig. 3, where:

- $P_0$ is the traded price on the initial purchase day
- $P_{30}$ represents the stock price at the end of the holding period

To clarify further, 30 portfolios were constructed using a 30-day rolling window. At the close of each trading day, the entire portfolio formed at $t_0$ was liquidated and replaced by a newly formed portfolio for the subsequent day. Thus, this straightforward yet effective calculation measures the investment's performance over the analyzed horizon. Ultimately, the resulting equity curve therefore reflects continuous capital deployment, free of idle days, and is directly comparable with a 30-day rolling S&P 500 benchmark constructed under identical rebalancing mechanics.

Consequently, as a benchmark to beat, the performance of the ESG-based investment strategy is evaluated by comparing the returns against the broader market performance of the S&P 500, applying the same 30-day rolling rationale. Key metrics include average return, average extra returns, and hit rate over the whole testing period.

Visual representations of ESG scores and corresponding stock returns are employed to facilitate a comprehensive analysis. These plots help identify trends and patterns that might not be immediately apparent from the raw data. The sampling universe and the distribution of market returns are also plotted to contextualize the ESG scores within the broader market environment.

## 4. Results

This section provides an exhaustive analysis of the backtesting results from the investment strategy as well as listing all input

**Table 2**
Back-testing engine diagram.

| Step | Action at trade-date t |
|---|---|
| ① | Predict $\hat{r}_{i,t+30}$ for every S&P 500 constituent |
| ② | Define "A-class" universe:<br>a) $\hat{r} > 0.10$<br>b) ESG Score $\leq$ daily K-quantile |
| ③ | Pick top-M = 15 names by $\hat{r}$; fall back to ESG-benchmark top-M (as shown in Fig. 2) if ② empty |
| ④ | Buy equal-weighted and hold exactly 30 days (30 overlapping portfolios live in parallel) |
| ⑤ | At $t_{+30}$, compute realised return and reinvest in new portfolio at $t_{+1}$ |

variables for the ML pipeline (as shown in Table 3). On the whole, the focus is on evaluating key performance metrics, providing insights into the strategy's efficacy.

In order to evaluate the performance of the proposed ESG portfolio strategy based on CatBoostRegressor, we present Table 4, which compares our model against a selection of recent and representative approaches from the literature. This comparison includes both traditional econometric models such as OLS, FF5, and VAR, as well as more advanced machine learning and optimization techniques-including Random Forests, Bayesian Optimization, and Deep Reinforcement Learning-all incorporating ESG-related constraints or signals.

The evaluation relies on four key annualized performance metrics commonly used in financial literature: Alpha, Portfolio Return, Volatility, and the Sharpe Ratio. Within this framework, our proposed model achieves an annualized alpha of 0.132 and a Sharpe Ratio of 1.05, indicating a strong ESG-aware risk-adjusted return profile. For instance, the FF5-optimized ESG portfolio developed by Berg et al. (2024) attains a higher Sharpe Ratio (1.22), but with a considerably lower alpha (0.064). Similarly, the Random Forest model examined by Wolff & Echterling (2024) underperforms in terms of alpha (0.09), exhibits lower returns, and suffers from higher volatility (0.227).

In addition, the proposed model significantly outperforms the Sharpe-maximizing Bayesian optimization approach of Garrido-Merchán et al. (2023a), delivering a substantially higher annualized return of 0.596 while maintaining a competitive volatility level of 0.121. It is worth noting that several benchmark models report incomplete performance metrics, frequently omitting alpha or volatility, which points to a lack of methodological coherence that the present framework explicitly addresses.

The results indicate that combining high-frequency ESG sub-indicators with a robust gradient-boosted decision tree model (CatBoost) enhances alpha generation and improves overall risk-adjusted performance. This evidence supports the effectiveness of machine learning-based ESG integration in portfolio construction and highlights the suitability of the proposed strategy for institutional investors seeking to reconcile sustainability considerations with strong financial performance.

As a preliminary step, a sensitivity analysis is conducted. Specifically, the simulation incrementally increases the ESG cutoff (refer to Fig. 2). From a conceptual standpoint, as the threshold approaches 1, the ESG constraint weakens, broadening the sustainability-related constraint. This analysis systematically evaluates how variations in the value of K impact the average performance of the ML-driven portfolio.

As shown in Fig. 4, portfolio returns show a complex relationship between ESG cutoffs and financial returns, while predominantly beating market performances. The results highlight that higher ESG threshold values, which may include companies with lower ESG performance, seem to yield the highest financial return. To illustrate, the portfolio return starts at 3.10 % for the 0.1 quantile threshold (companies with more sustainable ESG profiles), peaks at 4.97 % at 0.2, and then experiences a sharp decline to −1.63 % at 0.3. However, after the dip, the portfolio return increases steadily from 0.2 % at the 0.6 quantile to 6.26 % at the 0.9 quantile, the latter representing companies with the least ESG-friendly practices. Conversely, market return remains constant and it is represented in the plot for benchmarking purposes.

The sensitivity analysis suggests that portfolios that are more tolerant of anti-ESG practices can potentially generate higher financial returns. However, this idea does not necessarily imply a straightforward trade-off between ESG performance and profitability. Specifically, the dip in returns at the 0.3 quantile highlights the complexity of this relationship. Specifically, while the ESG cutoff decreases linearly, the number of companies in the investable space does not. As previously shown in Carlei et al. (2024), ESG scoring reports a highly skewed distribution; this results in substantially fewer companies being highly compliant with environmental, social, and governance practices. To this extent, lower ESG quantiles (representing companies with better ESG performance) still demonstrate positive returns.

Additionally, the U-curve presented in Fig. 4 can be further divided into 3 categories, under the assumption that the market reports 2 predominant forces: ESG and momentum. Intuitively, if a company does not have specific merits in terms of compliance or expected performances, it falls in the middle, underperforming against the broader market. Additionally, even though we're quite admissive at

**Table 3**
Strategies input space overview and classification.

| Feature Name | Type |
|---|---|
| Carbon − Overall-Exposure Score | ESG |
| Carbon − Overall-Risk Category | ESG |
| Carbon − Overall-Risk Score | ESG |
| ESG Risk Category | ESG |
| ESG Risk Score | ESG |
| Environment-Exposure Score | ESG |
| Environment-Risk Score | ESG |
| Governance-Exposure Score | ESG |
| Governance-Risk Score | ESG |
| Overall Exposure Score | ESG |
| Social-Exposure Score | ESG |
| Social-Risk Score | ESG |
| past_perc_variation_50_days | Quantitative |
| past_perc_variation_40_days | Quantitative |
| past_perc_variation_30_days | Quantitative |
| past_perc_variation_21_days | Quantitative |

**Table 4**
Comparison of ESG investment strategies based on annualized performance metrics.

| Model/Method | Alpha | Return | Volatility | Sharpe Ratio | Source |
|---|---|---|---|---|---|
| CatBoostRegressor + ESG | 0.132 | 0.596 | 0.121 | 1.05 | Current study |
| Random Forest + ESG | 0.09 | 0.169 | 0.227 | 0.75 | Wolff & Echterling (2024) |
| OLS, CAPM, VAR on S&P500 with ESG Risk Scor | 4.37 | n.a | n.a. | n.a | Sorathiya et al. (2024) |
| Bayesian Optimization with ESG penalties (Sharpe target) | n.a. | 0.0463 | 0.16 | 2.44 | Garrido-Merchán et al. (2023a) |
| Deep RL vs. MVO with ESG + financial objectives (Sharpe/Sortino) | n.a. | 0.1203 | n.a | n.a | Acero et al. (2024) |
| FF5 alpha, ESG-optimized portfolios (OPT, AVGvote, MAHA) + Treynor–Black weighting | 0.064 | 0.065–0.084 | n.a. | 1.22 | Berg et al. (2024) |
| Greenest-quintile ESG portfolio (Q1), US 4-factor alpha | 0.0797 | 0.247 | 0.1396 | 0.46 | Enders et al. (2025)[1] |
| EW/MinVar portfolios, FF3 alpha by region (NA, JP, EU) | 0.0047 | 0.0124 | n.a. | 0.14 | Horn & Oehler (2024) |
| ESG overlay on Quality; 7-factor robust regression | 0.031 | 0.0734 | 0.1562 | 0.41 | Mohanty et al. (2021) |
| ESG Screening (CSR Strategy, Community Score), MSCI ACWI benchmark | n.a. | n.a. | 0.009–0.016 | 0.59–0.69 | Ehlers et al. (2024)[2] |

[1]Enders et al. (2025) analyze ESG quintile portfolios using monthly data from 2009 to 2019. For the greenest quintile (Q1), they report a monthly return of 1.87%, volatility of 4.03%, and a 4-factor alpha of 0.64%. Annualized values are estimated through standard financial transformation.
[2]Ehlers et al. (2024) report Sharpe ratios between 0.59 and 0.69 for ESG-screened portfolios, based on Community and CSR Strategy scores. Tracking error ranges from 0.009 to 0.016 relative to the MSCI ACWI. Alpha values are described as positive but not statistically significant, and annual return is not disclosed.

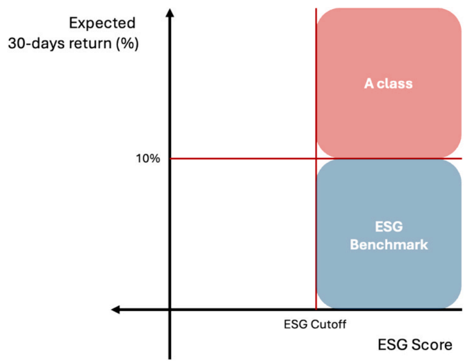


**Fig. 2.** Illustrative representation of sampling universe.

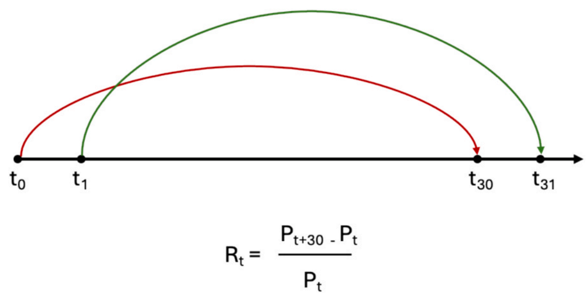


**Fig. 3.** Holding period and return computation, for each stock and trading day.

the 0.9 ESG cutoff, selected stocks still report predictions above the defined threshold for both sustainability and expected performances (denoted by the ML forecast).

To further deepen the portfolio ESG risk profile corresponding to each ESG threshold, a series of additional graphs in Fig. 5 provides a comprehensive analysis. The latter plot reveals, for individual ESG scores, their magnitude as the ESG threshold increases (the portfolio becomes less "sustainable" and thus more tolerant). As expected, the score increases with the cutoff across various dimensions—environmental, social, governance, and overall exposure— and the strategy shows a consistent upward trend in portfolio performances too. However, the latter point illustrates the adaptability of the strategy to different investor risk profiles, highlighting the flexibility of the present approach. Briefly, this methodology allows for customizable ESG thresholds, where higher cutoffs (such as 0.9) represent a selection for relatively less risk-averse investors. From a practical standpoint, these types of investors are willing to

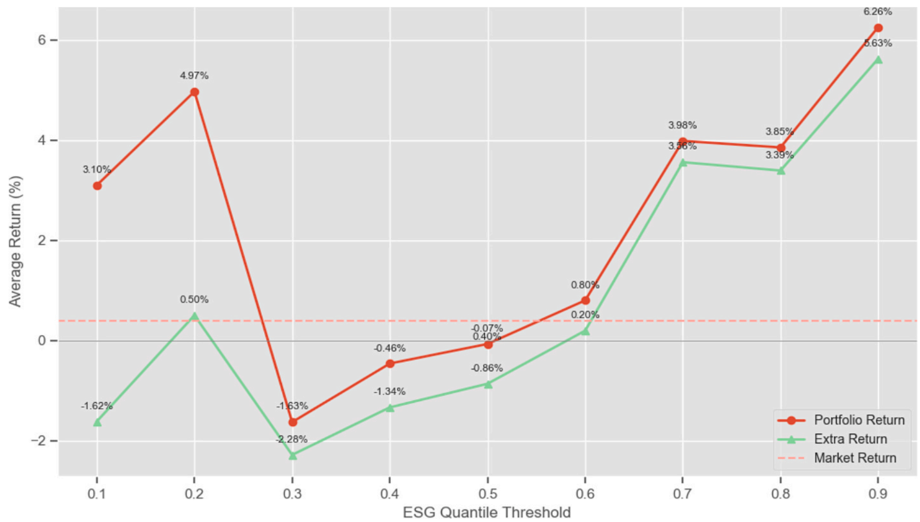


**Fig. 4.** Sensitivity Analysis: average 30-day return of portfolio vs. market across ESG quantile thresholds (2021–05-27——2023-02–21).

broaden ESG constraints in exchange for potentially higher returns.

More importantly, the strategy demonstrates, across the whole spectrum, that risk levels can be effectively managed and are often below the benchmark (Fig. 5). This proves a key trade-off: while higher returns are associated with increased ESG risk exposure, the overall risk remains controlled and competitive relative to standard benchmarks. Notably, environmental and governance risk scores show a stronger correlation with returns as the ESG quantile increases, indicating that the strategy effectively manages risk in these domains, even in portfolios with stricter ESG compliance.

Fig. 5 Introduces a comprehensive analysis of risk and exposure scores, segmented by morningstar's criteria, across various simulations driven by different esg cutoffs. specifically, within each figure, a box represents the risk category the average value of the variable falls into, for a given run and across all ml-driven portfolios, with k set to a specific threshold(e.g.*, Environment-Risk Score* across all the ML portfolios created within the test set reports an average value that is *Negligible*, for K equal to 0.5). Overall, each plot illustrates the distribution of a range of distinct ESG variables across all potential risk categories (Negligible, Low, Medium, High, and Severe) and at varying ESG cutoff levels. More in detail, each variable corresponds to a particular risk measure, with the ESG quantile thresholds displayed through a color gradient: lighter shades indicate lower ESG quantiles (e.g., 0.1, 0.2) and darker shades denote higher ones (e.g., 0.9). This breakdown underscores, from a theoretical perspective, how different ESG tolerance levels – represented by the ESG Quantile value – affect the sustainability profile of a portfolio.

As the ESG cutoff increases (allowing less sustainable companies into the portfolio), the distribution does not report variations across categories, in areas such as "Environment-Risk Score," "Social-Risk Score," and "Governance-Risk Score", predominantly allowing for "negligible" risk profiles across all portfolios. This visualization confirms that portfolios with broader ESG constraints (higher ESG thresholds) indeed do not generally show an increasing concentration in higher-risk categories. This pattern is in spark contrast with the hypothesis that higher returns observed at higher thresholds come at the cost of increased ESG risk. However, in some categories such as "Overall Exposure Score", reported behaviour displays a particularly pronounced shift into the "Medium" and "High" risk categories, as ESG cutoff increases.

The detailed breakdown in Fig. 6 therefore highlights the lack of relevant trade-off faced by investors, in terms of ESG constraints relaxation and financial returns, with the only exception of Exposure Scores. The flexibility to adjust ESG thresholds in the strategy allows investors to not only choose portfolios that align with their risk tolerance and sustainability objectives but also trace expected financial gains without sacrificing ESG risk profile. Furthermore, from a machine learning perspective, the model effectively learns and exploits policies, from the observed data, that positively correlate financial returns with compliant ESG profiles. Ultimately, proving that the algorithm favours the creation of portfolio strategies that manage – at different levels – ESG performances and nonetheless create alpha.

In conclusion, this ML-driven investment strategy skillfully integrates ESG risk assessments with robust portfolio performance, proving that relative mitigation of ESG constraints may improve overall financial performance while retaining a low overall risk profile. This capability to strategically leverage high-risk, low-ESG assets allows for optimal returns within controlled risk thresholds. By providing precise ESG risk management without compromising financial outcomes, the strategy underscores the transformative role of machine learning in sustainable investing, making it a valuable framework for investors seeking to balance ESG values with strong financial performance.

## 5. Conclusion

In conclusion, this study demonstrates that the proposed strategy is not only flexible but also effective in generating alpha across a

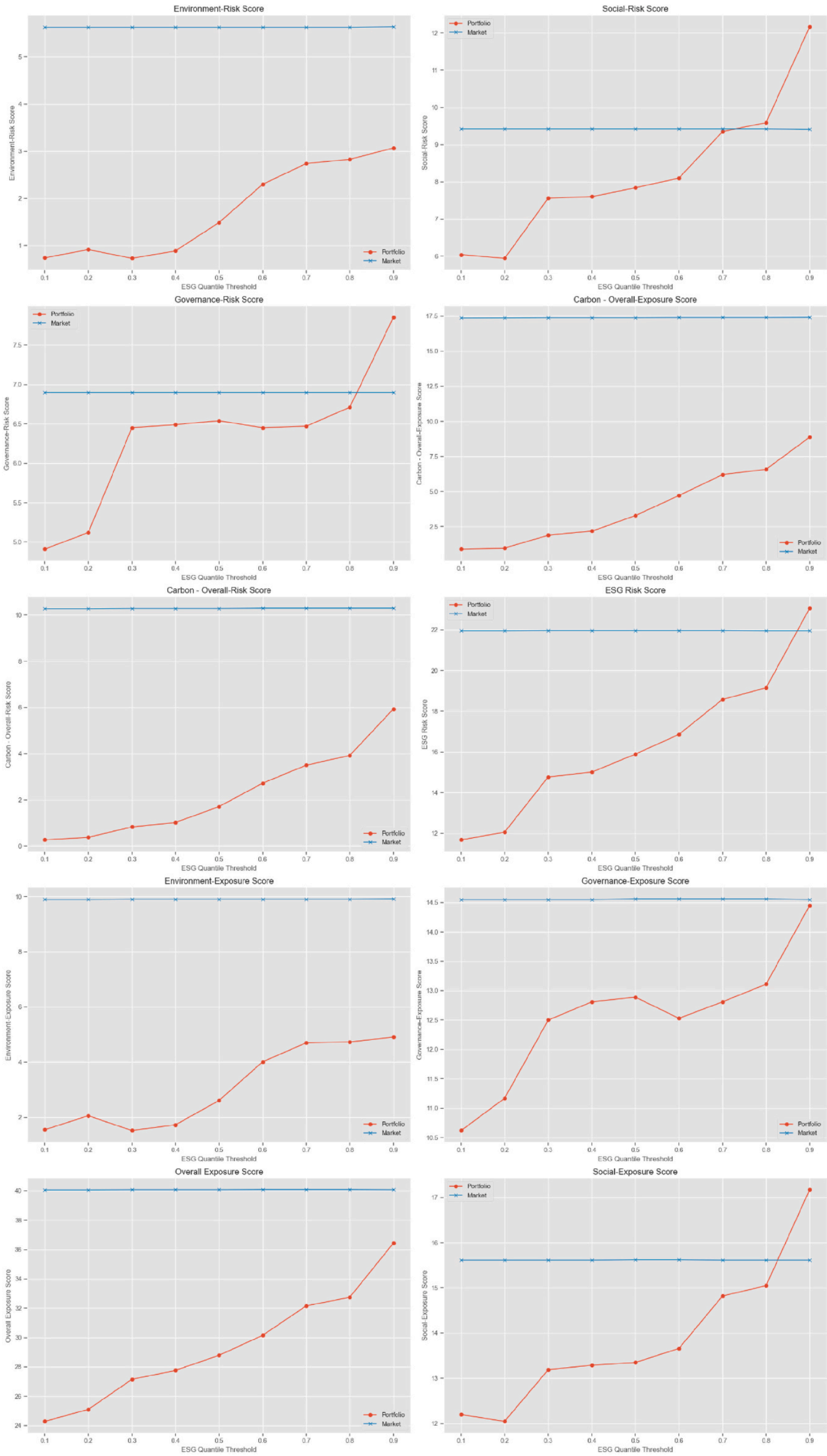


**Fig. 5.** Comprehensive sensitivity analysis of full input space.

broad spectrum of ESG risk profiles, including more permissive ones. Remarkably, even when portfolios are constructed with lower ESG restrictions—often aimed at achieving market-beating returns—the algorithm maintains its capacity to deliver alpha. This adaptability underlines the strength of the model in accommodating diverse investment goals, allowing investors to benefit from

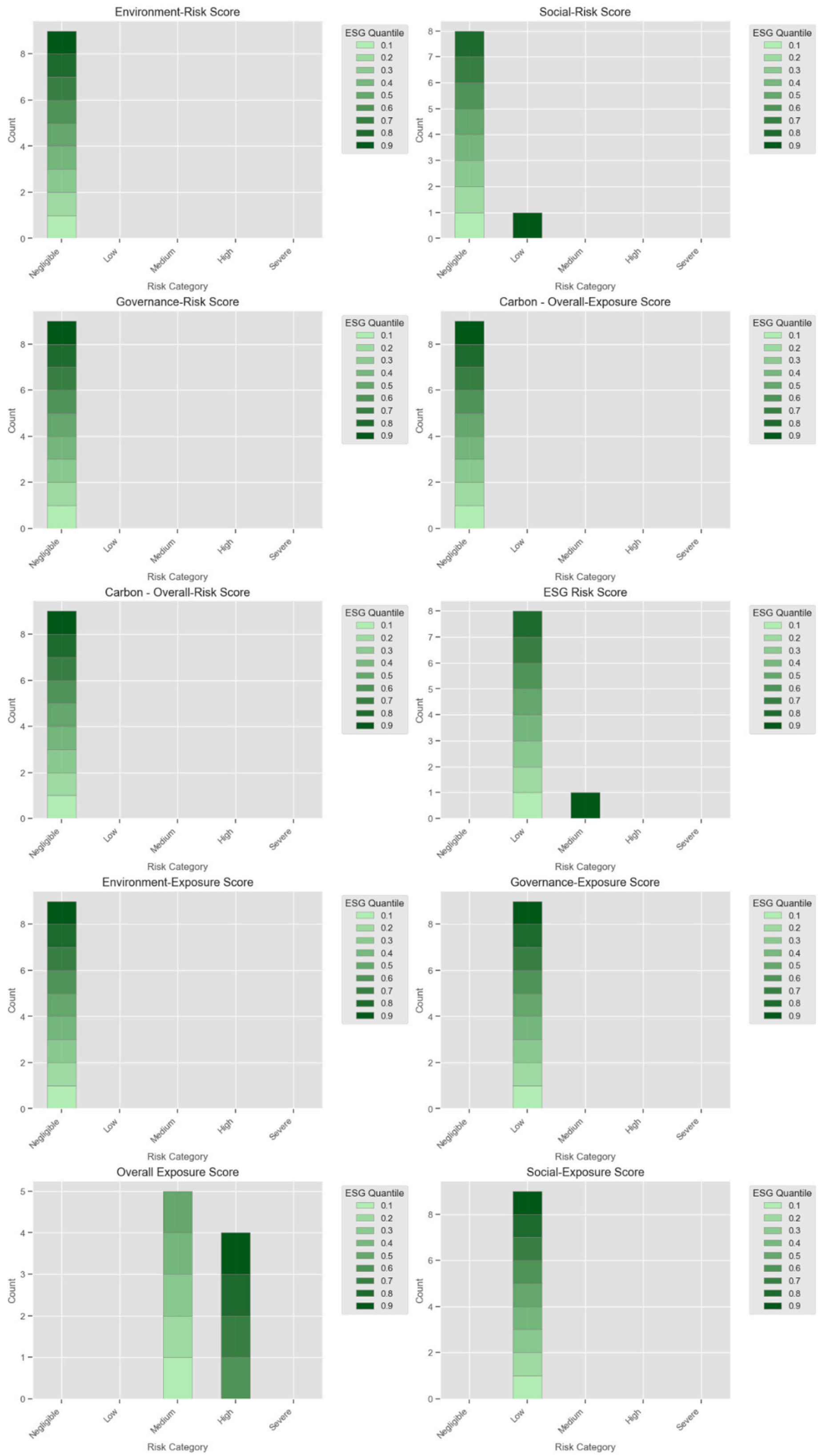


**Fig. 6.** Sensitivity analysis by category of full input space.

optimized stock selection aligned with their chosen ESG tolerance.

The analysis showcases a tailored approach, enabling the creation of bespoke portfolios that reflect individual ESG risk preferences. Investors can thus balance their ethical commitments with financial objectives, crafting portfolios that range from conservative, high-ESG compositions to more permissive, performance-oriented configurations. Ultimately, this study affirms that ESG-focused investing is not a trade-off against profitability but a versatile, resilient strategy that can be customized to harmonize both ethical priorities and financial gains, regardless of the ESG rigor applied.

## CRediT authorship contribution statement

**Vittorio Carlei:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Donatella Furia:** Writing – review & editing, Writing – original draft, Validation, Supervision, Conceptualization. **Alessandro Ceccarelli:** Writing – original draft, Visualization, Methodology, Investigation, Data curation. **Piera Cascioli:** Writing – review & editing, Writing – original draft, Investigation, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References


Acero, F., Zehtabi, P., Marchesotti, N., Cashmore, M., Magazzeni, D., & Veloso, M. (2024). Deep reinforcement learning and mean-variance strategies for responsible portfolio optimization. arXiv preprint arXiv:2403.16667.

Ahmed, S., Alshater, M. M., Ammari, A. E., & Hammami, H. (2022). Artificial intelligence and machine learning in finance: A bibliometric review. *Research in International Business and Finance, 61*, Article 101646.

Beaver, W. (1968). The information content of annual earnings announcements. *Journal of Accounting Research 6 (Selected Studies)*, 67–92.

Berg, F., Lo, A. W., Rigobon, R., Singh, M., & Zhang, R. (2024). Quantifying the returns of ESG investing: An empirical analysis with six ESG metrics. *The Journal of Portfolio Management, Special Issue Dedicated to Harry Markowitz, 50*(8), 216–238.

Broadstock, D. C., Chan, K., Cheng, L. T., & Wang, X. (2021). The role of ESG performance during times of financial crisis: Evidence from COVID-19 in China. *Finance Research Letters, 38*, Article 101716.

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., ... & Varoquaux, G. (2013). API design for machine learning software: experiences from the scikit-learn project. arXiv preprint arXiv:1309.0238.

Carlei, V., Cascioli, P., Ceccarelli, A., & Furia, D. (2024). can machine learning explain alpha generated by ESG factors? *Computational Economics*, 1–21.

Carlei, V., Terzi, S., Giordani, F., & Adamo, G. (2021). Portfolio management via empirical asset pricing powered by machine learning. In *Decision Economics: Minds, Machines, and their Society* (pp. 121–129). Springer International Publishing.

Chen, S., & Ge, L. (2021). A learning-based strategy for portfolio selection. *International Review of Economics and Finance, 71*, 936–942.

Chen, Z., Xiao, F., Guo, F., & Yan, J. (2023). Interpretable machine learning for building energy management: A state-of-the-art review. *Advances in Applied Energy, 9*, Article 100123.

Ehlers, T., Elsenhuber, U., Jegarasasingam, A., & Jondeau, E. (2024). Deconstructing ESG scores: Investing at the category score level. *Journal of Asset Management, 25*(3), 222–244.

Enders, A., Lontzek, T., Schmedders, K., & Thalhammer, M. (2025). Carbon risk and equity prices. *Financial Review, 60*(1), 13–32.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*, 383–417.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics, 116*, 1–22.

Fatemi, A., Glaum, M., & Kaiser, S. (2018). ESG performance and firm value: The moderating role of disclosure. *Global Finance Journal, 38*, 45–64.

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. *Journal of sustainable finance & investment, 5*(4), 210–233.

Garrido-Merchán, E.C., Mora-Figueroa-Cruz-Guzmán, S., & Coronado-Vaca, M. (2023b). Deep Reinforcement Learning for ESG financial portfolio management. arXiv preprint arXiv:2307.09631.

Garrido-Merchán, E. C., Piris, G. G., & Vaca, M. C. (2023a). Bayesian optimization of ESG (Environmental Social Governance) financial investments. *Environmental Research Communications, 5*(5), Article 055003.

Ghosh, P., Neufeld, A., & Sahoo, J. K. (2022). Forecasting directional movements of stock prices for intraday trading using LSTM and random forests. *Finance Research Letters, 46*, Article 102280.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies, 33*(5), 2223–2273.

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.

Horn, M., & Oehler, A. (2024). Constructing stock portfolios by sorting on ESG ratings: Does the rating provider matter? *International Review of Financial Analysis, 96*, Article 103568.

Hou, K., Xue, C., & Zhang, L. (2018). Replicating anomalies. *Review of Financial Studies, 33*, 2019–2133.

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.

Israel, R., Kelly, B. T., & Moskowitz, T. J. (2020). Can Machines 'Learn' Finance? *Journal of Investment Management, 18*(2).

Kelly, B., & Xiu, D. (2023). Financial machine learning. Foundations and Trends®. *Finance, 13*(3–4), 205–363.

Kelly, C. (2023). *Machine Learning in Empirical Asset pricing. a thesis submitted in fulfilment of the requirements for the degree of Doctor of philosophy in Finance*. Queen's University Belgast.

Krüger, P. (2015). Corporate goodness and shareholder wealth. *Journal of Financial Economics, 115*(2), 304–329.

La Torre, M., Mango, F., Cafaro, A., & Leo, S. (2020). Does the ESG index affect stock return? evidence from the Eurostoxx50. *Sustainability, 12*, 6387.

Li, H., Li, L., Liang, J., & Dai, W. (2023). A quantitative fusion strategy of stock picking and timing based on Particle Swarm Optimized-Back Propagation Neural Network and Multivariate Gaussian-Hidden Markov Model. In *In Proceedings of the 2023 4th International Conference on Big Data Economy and Information Management* (pp. 185–192).

Lundberg, S. & Lee S.I. (2017). A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874.

Mohanty, S. S., Mohanty, O., & Ivanof, M. (2021). Alpha enhancement in global equity markets with ESG overlay on factor-based investment strategies. *Risk Management, 23*(3), 213.

Novak, M. G., & Velušček, D. (2016). Prediction of stock price movement based on daily high prices. *Quantitative Finance, 16*, 793–826.

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.

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*.

Ren, T., Li, S., & Zhang, S. (2024). Stock market extreme risk prediction based on machine learning: Evidence from the American market. *The North American Journal of Economics and Finance, 74*, Article 102241.

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.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance, 19*(3), 425–444.

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

Sorathiya, A., Saval, P., & Sorathiya, M. (2024). Data-driven sustainable investment strategies: Integrating ESG, financial data science, and time series analysis for alpha generation. *International Journal of Financial Studies, 12*(2), 36.

Steuer, R. E., & Utz, S. (2023). Non-contour efficient fronts for identifying most preferred portfolios in sustainability investing. *European Journal of Operational Research, 306*(2), 742–753.

Tepelyan, R., & Gopal, A. (2023). Generative Machine Learning for Multivariate Equity returns. In *Proceedings of the Fourth ACM International Conference on AI in Finance* (pp. 159–166).

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*, Article 102389.

Wolff, D., & Echterling, F. (2024). Stock picking with machine learning. *Journal of Forecasting, 43*(1), 81–102.

Yang, J., Agyei, S. K., Bossman, A., Gubareva, M., & Marfo-Yiadom, E. (2024). Energy, metals, market uncertainties, and ESG stocks: Analysing predictability and safe havens. *The North American Journal of Economics and Finance, 69*, Article 102030.