European Financial Management

EUROPEAN FINANCIAL MANAGEMENT

ORIGINAL ARTICLE OPEN ACCESS

Sustainable Portfolio Construction via Machine Learning: ESG, SDG and Sentiment

Xin Feng[1] | Hans-Jörg von Mettenheim[2] | Georgios Sermpinis[1] | Charalampos Stasinakis[1]

[1]University of Glasgow, Adam Smith Business School, University of Glasgow, Glasgow, UK | [2]IPAG Business School, Paris, France

Correspondence: Charalampos Stasinakis (Charalampos.Stasinakis@glasgow.ac.uk)

Received: 18 April 2024 | **Revised:** 4 October 2024 | **Accepted:** 24 October 2024

Funding: The authors received no specific funding for this work.

Keywords: ESG | machine learning | portfolio construction | SDG | sentiment indicators

ABSTRACT

This study proposes portfolio construction strategies based on novel sentiment, ESG and SDG scores. We utilize natural language processing to establish a novel daily score system that mitigates concerns of different rating standards. The portfolios constructed are optimized via machine learning algorithms on a monthly basis using daily historical returns. Utilizing the equal-weighted portfolios as benchmarks, we empirically show that our optimized portfolios exhibit better trading performance in both the SPX500 and STOXX600 indices. The findings demonstrate that nonlinear models such as random forests, neural networks, and genetic algorithms can perform better than other machine learning models in portfolio management.

JEL Classification: F3, G11

1 | Introduction

The concept of sustainability has gained significant attraction globally, prompting various stakeholders, such as businesses, governments, and generally society, to implement diverse innovative approaches. The prominence of sustainability and technological advancement has resulted in the emergence of Environmental, Social, and corporate Governance (ESG) as a crucial metric for assessing the sustainable profile of stocks in the capital markets[1]. Corporate Social Responsibility (CSR) is commonly perceived as a criterion for firms to attain sustainability and numerous academic studies have demonstrated its significant role in financial management (Edmans 2011; Lins, Servaes, and Tamayo 2017; Feng, Chen, and Tseng 2018). A significant portion of the prior research has examined the correlation between CSR and financial performance[2]. Our study focuses on the development of portfolios utilizing daily sustainability indicators, as well as the incorporation of sentiment index and machine learning methodologies to address the demands of processing large data volumes.

This paper is motivated by three dimensions of the literature. First, the application of the sentiment index in financial investment is experiencing an upward trend. The Efficient Market Hypothesis (EMH) proposed by Fama (1970) suggests that stock price in the market incorporates all pertinent information in a timely, accurate, and comprehensive manner. However, several studies explain that in the stock market, investors' behaviour has greater heterogeneity, resulting in abnormal returns that cannot be explained by traditional financial theories (Kumar, Page, and Spalt 2013). Therefore, most scholars consider investors' sentiment in financial investment as an important factor. Several researchers have attempted empirical frameworks to investigate this topic, and many of them find a positive relationship between current investor sentiment and future stock returns (Gao, Gu, and Koedijk 2021). Our study focuses on trading performance of portfolios constructed through equal-weighted stock selection based on the top sentiment scores (Gillam, Guerard, and Cahan 2015).

Second, corporate sustainability is also extensive and multidimensional depending on the measures used from researchers, namely ESG, CSR or Sustainable Development Goals (SDG). Although ESG, CSR and SDG are related concepts, they differ in focus, implementation and purpose. ESG is a quantitative, externally regulated approach that focuses on measuring a company's sustainability. ESG is implemented through measurable goals and audits and is often used to influence investor decisions. ESG reporting focuses on monitoring and measuring a company's impact on the environment, society, and governance. On the other hand, CSR is a qualitative, self-regulated approach that focuses on a company's broader social impact. CSR is implemented through a company's culture, values, and brand management and it is often associated with the idea that businesses have responsibilities to society that go beyond making a profit. Finally, SDGs corresponds to a set of 17 global goals set by the United Nations in 2015, including also 169 sub-targets that aim to address the most pressing sustainability challenges faced by the global community (UNGA 2015). The SDGs are intended to be achieved through a combination of efforts from various parties such as governments, local authorities, corporations, civil society groups, and individuals worldwide, unlike ESG and CSR that are firm oriented. Finally, SDG does have a timeline (end year is 2030), while ESG and CSR are long-term and continuous commitments for firms. All three are considered to overlap but also reinforce each other. Companies can use ESG to quantify the impact of their CSR efforts and align their strategies to contribute to the SDGs. This can help companies improve their reputation and address global challenges. However, companies must also navigate the different cultural and market contexts in which they operate. Hence, although there might be a contextual and regulatory overlap, there is high diversity in terms of generated environmental culture, sustainable innovation and efforts towards societal impact.

Third, machine learning has gained significant prominence in the realm of quantitative finance and shows great predictive ability (Krauss, Do, and Huck 2017; Giannone, Lenza, and Primiceri 2021; Akyildirim et al. 2023), particularly in the current era of big data. Machine learning models are more flexible because of their ability to cope with the dimensionality issue compared to more traditional approaches, making them appropriate for processing large-volume data and fitting nonlinear correlations. In terms of the machine learning techniques, in this study, we use the Least Absolute Shrinkage and Selection Operator (LASSO), Genetic Algorithms (GA), Support Vector Regression (SVR), Classification and Regression Tree (CART), Random Forests (RF) and Neural Networks (NN).

The benefits of ESG are documented by several studies associating improved financial performance with high-ESG firms. Numerous prior studies have demonstrated the strong risk resistance (Albuquerque, Koskinen, and Zhang 2019; Hoepner et al. 2024), profitability (Cornett, Erhemjamts, and Tehranian 2016) and financial cost reduction (Ng and Rezaee 2015; Apergis, Poufinas, and Antonopoulos 2022) of ESG investment. On the other hand, the increasing trend of sustainable investing using ESG factors has been accompanied by heated debates regarding the reliability of ESG research in recent times. These concerns are mostly related to the heterogeneity and inconsistency of ESG ratings provided by different agencies, such as MSCI ESG KLD, Refinitiv, Moody's ESG, Sustainalytics, and S&P Global. Berg, Kölbel, and Rigobon (2022) observe that dissimilarities in measurement are the main cause of variations in ESG ratings, and different agencies hold divergent perspectives on which categories are most crucial in ESG assessment. ESG uncertainty appears to affect the equity premium, reduce investors' demand for risky assets, and decrease economic welfare (Avramov et al. 2022). Gibson Brandon, Krueger and Schmidt (2021) demonstrate that disagreement is more common among large firms and firms without credit ratings. It is worth noting that the firms included in our study belong to the SPX500 and STOXX600, both of which are classified as large firms, hence the incorporation of these scores into the construction of a portfolio can have a significant impact.

This study is attempting to alleviate the above concerns by applying Natural Language Processing (NPL) to construct new and daily ESG-related and SDG-related sentiment scores in this study. Some researchers employ NLP to create sentiment indices for their subsequent research (Loughran and McDonald 2011; Heston and Sinha 2017). NLP employs computer technology to comprehend, process, analyze, and express natural language. The method allows us to study quantified linguistic data which enables the calculation of daily sentiment scores. This approach enables the monitoring of the daily sentiment pertaining to listed firms, as well as the sentiment directed towards the firms' ESG and SDG. Furthermore, the sentiment indicators that have been acquired are utilized as selection criteria when constructing portfolios.

This study contributes to the literature by proposing a novel empirical framework for sustainable portfolio investment. First, we construct daily ESG-related, SDG-related, and overall sentiment indicators for further study. Our empirical setup begins collecting online news for all the firms of the SPX500 and STOXX600 from February 2015 to February 2022. The creation of these three indicators serves as a means to mitigate the presence of heterogeneity and inconsistency in the ESG scores that are issued by various agencies. Second, we demonstrate the effectiveness of utilizing sentiment indicators in the process of stock selection. The best-in-class (positive) approach is employed in our portfolio trading strategy, which is also used in the research of Kempf and Osthoff (2007) and Statman and Glushkov (2009).

Finally, we construct portfolios with machine learning models and show their superior performance. The Mean-Variance (M-V) methodology is employed to construct portfolios using daily historical returns, which are then optimized through models involving machine learning algorithms on a monthly basis. We apply linear models, namely simple linear regression and LASSO, as well as nonlinear algorithms, such as GA, SVR, CART, RF and NN. Using the equal-weighted portfolios as benchmarks, we empirically show that our optimized portfolios exhibit better trading performance in both the SPX500 and STOXX600 indices. The findings also demonstrate that many nonlinear machine learning models beat the linear models, while GA, RF and NN generally beat other machine learning models even when accounting for transaction costs. Our results hold even when the Mean-Semivariance (M-SV) approach and the long-term test are utilized.

The remainder of the paper is organized as follows. Section 2 reviews the literature on sentiment analysis using ESG and SDG, as well as the implementation of machine learning techniques in finance. Section 3 describes our data sample and Section 4 describes our modelling framework. Section 5 presents our main empirical results, while Section 6 summarizes the robustness tests. Finally, Section 7 provides the concluding remarks.

2 | Related Literature

2.1 | Sentiment and Score Analysis in Finance

With the development of big data and computational technology, investor sentiment has become a useful tool in financial research, especially in asset pricing and stock investment. The sentiment index that has gained widespread acceptance is the BW index proposed by Baker and Wurgler (2006). They form a comprehensive index that captures the common components of the six agents, including the equity share in new issues, the closed-end fund discount, the number and average first-day returns on IPOs, the dividend premium, and NYSE share turnover. They find that this index captures opinions better than any other component in explaining stock returns. Other studies also employ the BW index and find its significant predictive power with respect to stock returns (Dergiades 2012; Stambaugh, Yu, and Yuan 2012; Huang et al. 2015; Sibley et al. 2016; Mbanga, Darrat, and Park 2019). The other popular sentiment index is constructed by lexicon-based approaches focusing on the textual tone. Jiang et al. (2019) use textual analysis to obtain monthly sentiment indices and test their forecasting ability on stock returns. Recent studies focus more on daily investor sentiment analysis due to the success of NPL (Heston and Sinha 2017; Renault 2017). Sun, Najand, and Shen (2016) even use the high-frequency (half-hour) investor sentiment indices and show that robust results of predictability of the intraday stock returns can be achieved. Da, Engelberg, and Gao (2011) observe that high market sentiment contributes to high first-day returns and long-term reversals of IPO stocks.

Generally, the literature reveals a positive association between investor sentiment and stock returns within the same period, but a negative association exists between investor sentiment and expected stock returns. According to Fang et al. (2021), firms with optimistic investor sentiment experience a notable increase in their stock returns during the present month, and this effect subsequently reverses. Da, Engelberg, and Gao (2015) construct a Financial and Economic Attitudes Revealed by Search (FEARS) index by aggregating the volume of queries related to market concerns. The index is correlated with low returns on the day but predicts high returns for the next one. Bathia and Bredin (2013) target a G7 countries sample and find the inverse relationship between investor sentiment and stock returns. These results are supplemented by Wang, Su, and Duxbury (2021), who focus on emerging markets and their results exert the negative predictability of sentiment from the subsequent 2 to 12 months. If investor sentiment causes the stock price to be higher (lower) than its intrinsic value, the future stock return will be lower (higher) due to mean reversion. This indicates the presence of an inverse relationship between investor sentiment and expected stock returns.

ESG is equally increasingly popular in financial research. Many studies investigate the relationship between ESG scores and financial performance. Friede, Busch, and Bassen (2015) in literature meta-analysis explain that 90% of the studies present a nonnegative relationship between corporate financial performance and ESG, making a clear business case for sustainability. In a similar approach, Atz et al. (2023) explaining that ESG investment can provide arbitrage opportunities especially in periods of economic or climate crisis. Clark, Feiner, and Viehs (2015) conduct a literature review of over 200 research studies coming to the conclusion that 88% of those studies find a positive relationship between financial performance and ESG scores. ESG score is an important indicator for measuring the sustainability profile of current and future stocks and it is associated with risk resistance and profitability in the capital market. Zhang, Zhao, and Qu (2021) consider Chinese stocks and show that stocks with high ESG scores achieve substantially higher abnormal returns. Khan (2019) constructs a long-only cap-weighted portfolio of the top quartile of ESG score, which can be found to outperform the overall index of all firms. Furthermore, Madhavan, Sobczyk, and Ang (2021) discover a correlation between elevated ESG scores and increased returns. However, the authors posit that this correlation is associated with ESG components, that are correlated with style factors such as size, value, momentum, volatility, and quality. More specifically, Pástor, Stambaugh, and Taylor (2022) conclude that green stocks usually have better performance than brown stocks, especially when considering those ones that are related to greater climate concerns. These findings are consistent with the empirical findings of Ardia et al. (2020) and Engle et al. (2020). Cheng, Ioannou, and Serafeim (2014) study the returns of target companies involved in promoting ESG improvement, indicating that firms implementing ESG policies get higher returns than peer companies that are less ESG-active. This effect is found to be significant mainly within 6 to 12 months.

Given the above, there is a strand in the literature that attempts to find appropriate ways to integrate ESG into investment. Screening is a relatively common way to implement ESG, which can be separated into positive (best-in-class) and negative (exclusion) methods. Statman and Glushkov (2009) discover that portfolios constructed with the best-in-class approach have higher risk-adjusted returns than portfolios with the negative screening method. This argument is complemented by Capelle-Blancard and Monjon (2014), who claim that sectoral screens (including the exclusion of sin stocks) have a negative effect on portfolio performance. That in turn indicates that the best-in-class approach is favourable. This is logical, as positive screening is more appropriate when a large proportion of investors want portfolios in accordance with their beliefs about green stocks (Dimson, Karakaş, and Li 2015; Krueger, Sautner, and Starks 2020). The previously mentioned positive relationship, although more common, is not clear-cut in the literature. Some studies bring forward evidence in support of the negative relationship between ESG and stock returns (Luo 2022). Luo and Balvers (2017) build a two-factor CAPM model (market and boycott factor) to test the investor boycott risk and claim that UK ESG-unfriendly firms (sin stocks) are anticipated to yield higher returns (even after controlling for various risks). This is also consistent with the findings of Hong and Kacperczyk (2009) on US stocks, however, sin stocks can suffer by higher litigation risk and lack of institutional investors' interest. Such empirical findings are supported by the notion of reputation importance and cost. In other words, firms with higher ESG scores that possess greater reputation and market value are indicative of higher future returns; sin firms face higher expenses related to legal and market standards in the form of regulatory corporate governance and potential penalties for sustainability offenses, which in turn can decrease the returns.

Several researchers propose methods for sustainable investing by gradually constructing ESG news scores rather than using typical 'static' institutional data. This is due to the fact that ESG ratings have been heavily scrutinized due to data consistencies, comparability, and quality (Berg, Kölbel, and Rigobon 2022). Tang and Zhang (2020) find that the stock market responds positively to the firms that announced the issuance of green bonds, leading to an increase of the stock liquidity and potential positive abnormal returns post-announcement. Similarly, Flammer (2021) argues that the stock market responds positively to the announcement of green bond issuance in the short-term period and that this may result in positive stock market outcomes. Capelle-Blancard and Petit (2019) study the stock market's response to ESG news and investigate their effects on the financial performance of the target firms. Serafeim (2020) brings up views of the public sentiment of companies' sustainable activities, which may affect the performance of firms. Overall, the literature suggests that the ESG sentiment index can be an attractive proxy used in financial investing.

Investor sentiment related to ESG, constructed through NLP, presents additional opportunities for high-frequency analysis. Krüger (2015) obtain investor responses to the ESG events with NLP from the KLD newsletter[3]. Recently, Serafeim and Yoon (2022) generate daily ESG news sentiment scores from TVL data and discover the potential predictive capability of ESG performance scores regarding forthcoming ESG sentiment scores[4]. These scores are daily and mitigate the possibility of encountering inconsistent data in our study.

2.2 | Machine Learning in Finance

With the explosive increase of financial data, the improvement of computing ability, and the continuous development of new optimization algorithms, the application of machine learning has become a crucial tool for contemporary financial investment. Aziz et al. (2022) present a diverse literature that explores the various applications of machine learning in finance, while Duarte, Duarte, and Silva (2024) propose a machine learning framework to cope with high-dimensional and nonlinear dynamic problems in asset pricing and portfolio selection. Many studies in the field of equity markets apply machine learning models to optimize portfolios (Ban, El Karoui, and Lim 2018), predict the future returns of stocks (Chen, Pelger, and Zhu 2024) and portfolios (Fang and Taylor 2021). Gu, Kelly, and Xiu (2020) compare the differences between a wide range of machine learning methods and explore the monthly behaviour of risk premium, finding that NNs can be the best-performing method. Akyildirim et al. (2023) also test the prediction ability of various machine learning methods at the 1-min frequency in different sliding windows, and they empirically conclude that XGBoost performs better in predicting excess returns among the used machine learning algorithms. However, Ban, El Karoui, and Lim (2018) present different empirical results and conclude that regularized linear models generally have increased predicting performance, while the more complex algorithms do not. Recently, Bagnara (2024) survey the use of ML in asset pricing and explain that their dynamic abilities can capture new patterns that static approximations would not. One such example of pattern is that high-frequency stock returns are driven by industry factors instead of traditional characteristics-based factors.

Many studies conclude that the portfolios constructed and optimized through machine learning methods outperform in most cases the benchmark, which is always set as an equal-weighted portfolio (Huck 2019; D'Hondt et al. 2020; Gu, Kelly, and Xiu 2020; Kynigakis and Panopoulou 2022). Wu et al. (2021) discover that portfolios selected via machine learning methods show significant increases in alphas and annualized returns compared to the benchmark index. DeMiguel et al. (2023) demonstrates that machine-learning techniques may effectively detect and utilize non-linearities and interactions in the connection between fund attributes and performance. This enables investors to select funds that provide positive returns after accounting for fees and transaction costs. In addition, Lin and Taamouti (2024) find that machine learning for portfolio construction yields statistically gains in terms of both profitability and minimizing systemic risk.

3 | Data and Summary Statistics

3.1 | Sustainability Score Data

We collect daily news data from thousands of online news sources covering the firms selected in the SPX500 and STOXX600 from February 2015 to February 2022. The majority of the relevant articles, totalling 17,270,585 in our research, are written in English, while 2,240,259 articles published in other languages are translated into English before further processing to maintain consistency in subsequent procedures. In addition, all the articles are tagged with firms and topics selected based on a NoSQL database software, ElasticSearch, which is built on the Apache Lucene search library. Then, we choose articles that are relevant to ESG/SDG of certain firms for further processing.

We have three main sentiment indicators used in this paper, namely the overall sentiment score, ESG score, and SDG score. The overall sentiment score is the average sentiment score of the general news that mentions the given firm. ESG score refers to the average sentiment score of all the ESG-related news of the firm. SDG score is the average sentiment score of all the given firm's news related to the 17 different sustainable development goals defined by the United Nations.

The lexicon-based approach with the pre-existing dictionaries of tonal words is used to test the sentiment of each word from the main text of each article. We calculate the daily sentiment indicators of each stock through NLP by obtaining the average sentiment variables of all news articles referring to the firm every day. The word can be classified as positive, negative, or neutral after the preprocessing progress, including removing stop words and lemmatization. The positive tone is the percentage of positive words in the article while the negative tone is the percentage of negative words. The scores range from −100 to +100, indicating the tones from totally negative to totally positive, where 0 reflects the neutral tone. We use 0 to substitute the null data in the score for further analysis. Figure 1 below illustrates the schematic diagram of the sentiment indicators' extraction procedure.

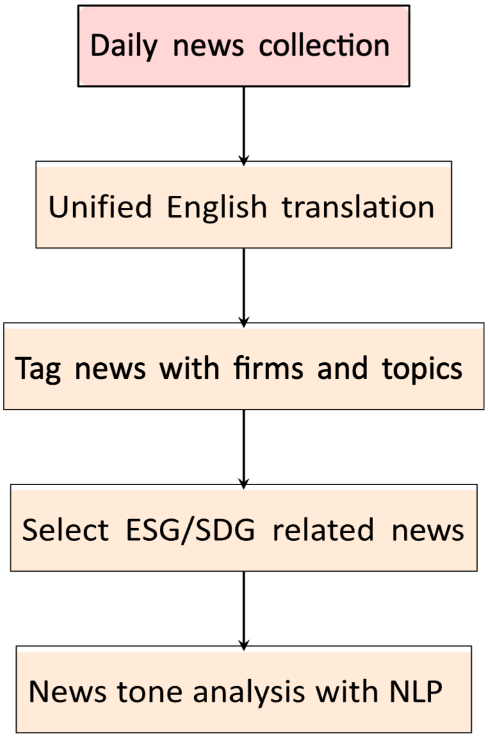

FIGURE 1 | Sentiment indicators' extraction process.

If the tone is close to zero, there may be two possibilities: either the emotional response of the text is generally low, or the positive and negative sentiment scores are at a similar level, and they offset each other. These two cases can be distinguished by using polarity variables, which is the sum of absolute values of positive and negative tones representing the degree of polarization of the article. The polarity variables range from 0 to +200.

For the news sentiment score data, we present two pieces of information. First, Table 1 provides the summary statistics for the valid score data we obtained in the different sample periods. Second, we provide the descriptive statistics of the obtained daily scores used for further analysis in Table 2. More than 0.8 million daily sentiment scores for stocks in both indices are created using NLP following data cleaning and processing. Table 2 demonstrates that stocks in SPX500 generally have higher market sentiment than those in the STOXX600, which is similar to the results of Schmeling (2009).

3.2 | Financial Data

The required financial data are collected from the Bloomberg database, including the daily returns of the stocks from the SPX500 and STOXX600 indices, risk-free rate, and index returns from February 2015 to February 2022. The returns are collected on a daily basis to align with the frequency of our generated sentiment scores. The risk-free rate here is the 1-Month Treasury Rate, which is the yield received for investing in a US government-issued treasury bill that has a maturity of 1 month. We screen out stocks that cannot be detected by our lexicon-base process, although included in the index. We exclude stocks that have more than 30% missing daily return data during the in-

TABLE 1 | Sustainability score availability of stocks.

		ESG Score			SDG Score			Overall Score		
IS period	**Index**	**20%**	**50%**	**80%**	**20%**	**50%**	**80%**	**20%**	**50%**	**80%**
3 m	SPX500	431	354	190	430	335	199	451	385	253
	STOXX600	237	150	93	237	154	95	285	183	117
6 m	SPX500	434	356	194	431	337	202	452	387	256
	STOXX600	237	152	93	237	155	95	285	183	117

Note: The table displays the number of stocks in SPX500 and STOXX600 indices, separately, that contain a greater than certain percentage of available scores through the aforementioned NLP technique in different sampling periods. "IS Period" is the in-sample period used for model construction, which are 3- and 6-month, respectively. This data set is analyzed before pre-processing, especially before the process of filling null values with zeros.

TABLE 2 | The descriptive statistics of daily sustainability scores.

Index	Variable	Observations	Mean	SD	Min	Max
SPX500	ESG Score	809,599	0.736	5.271	−100.000	99.701
	SDG Score	809,599	0.676	5.651	−100.000	99.701
	Overall Score	809,599	0.928	5.930	−100.000	99.713
STOXX600	ESG Score	887,507	0.214	4.520	−99.422	99.950
	SDG Score	887,507	0.268	4.778	−99.957	99.593
	Overall Score	887,507	0.441	5.123	−99.422	99.922

Note: This table shows the descriptive statistics of the three scores in SPX500 and STOXX600 indices, respectively. The table reports the sample size (Observations), standard deviation (SD), average (Mean), minimum (Min) and maximum (Max) of the daily scores.

sample (IS) period. Finally, the total average number of stocks in our data set is 489 and 542 in SPX500 and STOXX600 respectively.

3.3 | Ordinary Least Squares Regression

A necessary condition for our sentiment indicator scores to be useful in portfolio construction is that there are some statistical relationships between the scores and stock returns. To test this, we use a simple regression analysis based on the CAPM model to test whether significant linear relationships are identified with the subsequent return of stocks, as shown below:

$$R_{w,t} = \beta_{w,0} R_{I,t} + \sum_{w=1}^{W} \beta_{w,n} \text{Sentiment\_Score}_{w,t-1} + \epsilon_{w,t}, \quad (1)$$

where $R_{I,t}$ is the log return of the given stock index I at time$t$, $\text{Sentiment\_Score}_{w,t-1}$ is the value of each sentiment indicator at time $t-1$ and $\epsilon_{w,t}$ is the error term.

We use a Pooled Ordinary Least Squares (OLS) framework for both SPX500 and STOXX600 to verify the potential utility of these indicators. The panel regression results for the SPX500 and STOXX600 datasets are presented from Tables 3–8. All tables below show the significant and positive relationships between the Tone and Polarity variables and next month's return of the stock. Word count is found not to have a significant effect. For example, from Table 3 we see that one unit increase in the Tone leads to an increase between 0.08% and 0.20% in the next month's stock return (for Polarity equivalently the increase ranges between 0.07% and 0.20%). All these findings motivate the use of the three scores in our subsequent empirical setup.

4 | Methodology

The basic idea of our portfolio construction is to pre-select stocks first based on their sentiment scores. Many financial studies use

TABLE 3 | ESG sentiment indicators regression (SPX500).

	Log return $(t+1)$			
Dependent variable	**(1)**	**(2)**	**(3)**	**(4)**
Tone	0.0011***		0.002***	0.0008***
Polarity		0.0003***	0.0007***	0.002***
Word Count				−5.34e−08
Index Return	1.0941***	1.0948***	1.0941***	1.0941***

Note: The table shows the regression results of 4 different combinations of the ESG sentiment indicators on the SPX500 data set. 'Tone' is the average negative and positive tone of the document as a whole. 'Polarity' represents the indicator of how emotionally polarized the text of the article is. 'Word Count' is the total number of words contained in the document. *** represents significance at 1% level.

TABLE 4 | SDG sentiment indicators regression (SPX500).

	Log return $(t+1)$			
Dependent variable	**(1)**	**(2)**	**(3)**	**(4)**
Tone	0.0011***		0.0012***	0.0008***
Polarity		0.0004***	0.0007***	0.003***
Word Count				−6.12e−07
Index Return	1.0855***	1.0853***	1.0865***	1.0855***

Note: The table shows the regression results of 4 different combinations of the SDG sentiment indicators on the SPX500 data set. 'Tone' is the average negative and positive tone of the document as a whole. 'Polarity' represents the indicator of how emotionally polarized the text of the article is. 'Word Count' is the total number of words contained in the document. *** represents significance at 1% level.

TABLE 5 | Overall sentiment indicators regression (SPX500).

	Log return $(t+1)$			
Dependent variable	**(1)**	**(2)**	**(3)**	**(4)**
Tone	0.0012***		0.0012***	0.0008***
Polarity		0.0004***	0.0005***	0.002***
Word Count				−3.12e−08
Index Return	1.0752***	1.0754***	1.0781***	1.0759***

Note: The table shows the regression results of 4 different combinations of the Overall sentiment indicators on the SPX500 data set. 'Tone' is the average negative and positive tone of the document as a whole. 'Polarity' represents the indicator of how emotionally polarized the text of the article is. 'Word Count' is the total number of words contained in the document. *** represents significance at 1% level.

TABLE 6 | ESG sentiment indicators regression (STOXX600).

	Log return $(t+1)$			
Dependent variable	**(1)**	**(2)**	**(3)**	**(4)**
Tone	0.0011***		0.0009***	0.0008***
Polarity		0.0005***	0.0007***	0.0008***
Word Count				−6.01e−07
Index Return	1.0745***	1.0758***	1.0751***	1.0751***

Note: The table shows the regression results of 4 different combinations of the ESG sentiment indicators on the STOXX600 data set. 'Tone' is the average negative and positive tone of the document as a whole. 'Polarity' represents the indicator of how emotionally polarized the text of the article is. 'Word Count' is the total number of words contained in the document. *** represents significance at 1% level.

TABLE 7 | SDG sentiment indicators regression (STOXX600).

	Log return $(t+1)$			
Dependent variable	**(1)**	**(2)**	**(3)**	**(4)**
Tone	0.0010***		0.001***	0.0008***
Polarity		0.0003***	0.0008***	0.0002***
Word Count				−4.18e−07
Index Return	1.0811***	1.0823***	1.0813***	1.0810***

Note: The table shows the regression results of 4 different combinations of the SDG sentiment indicators on the STOXX600 data set. 'Tone' is the average negative and positive tone of the document as a whole. 'Polarity' represents the indicator of how emotionally polarized the text of the article is. 'Word Count' is the total number of words contained in the document. *** represents significance at 1% level.

TABLE 8 | Overall sentiment indicators regression (STOXX600).

	Log return $(t+1)$			
Dependent variable	**(1)**	**(2)**	**(3)**	**(4)**
Tone	0.0011***		0.0008***	0.0009***
Polarity		0.0006***	0.0005***	0.0008***
Word Count				−2.37E−07
Index Return	1.0792***	1.0794***	1.0797***	1.0797***

Note: The table shows the regression results of four different combinations of the Overall sentiment indicators on the STOXX600 data set. 'Tone' is the average negative and positive tone of the document as a whole. 'Polarity' represents the indicator of how emotionally polarized the text of the article is. 'Word Count' is the total number of words contained in the document. *** represents significance at 1% level.

screening strategies to generate high-performance portfolios (Humphrey and Lee 2011; Humphrey and Tan 2014; Bender and Wang 2016; Trinks and Scholtens 2017; Alessandrini and Jondeau 2021; Görgen, Jacob, and Nerlinger 2021).

The dynamic rolling window approaches are utilized in many financial studies (Forni et al. 2018; Zhang, Ma, and Liao 2020; Feng, Zhang, and Wang 2024) due to improvements in statistical and trading out-of-sample (OOS) performance. The analysis of Giamouridis and Vrontos (2007) shows that dynamic models can construct portfolios with lower risk and higher OOS returns than static models. Zhao et al. (2023) argue that while approaches like sample division can help mitigate the issue of parameter non-constancy caused by structural changes, there is still a possibility of persistent biases if the change points are not identified. Zografopoulos et al. (2024) also compare ML methods to forecast and trade U.S. industry portfolio returns. Their findings show that linear and nonlinear benchmarks are outperformed by advanced ML techniques in terms of forecasting accuracy, profitability and economically meaningful predictions.

To address the issue of parameter inconsistency caused by ambiguous change points in structural alterations, we employ the bootstrap rolling window subsample method. Here we choose the top M (in our study, $M = 10$ and 30) sentiment scores stocks in the rolling window T, which is 3-month and 6-month.

Then, we obtain the weight of each selected stock with the portfolio construction models listed in 4.2. The IS data is the daily returns of score-selected stocks in the rolling window T; the OOS data is the lagged returns in the subsequent month. We compare the annualized return and Sharpe ratio of each portfolio to test the success of the machine learning-constructed portfolio strategy. As a benchmark, we use the naive equal-weighted portfolio ('$1/M$'). For this benchmark, we assign an equal weight for each of the M stocks selected according to the scores. This strategy does not involve any optimization or estimation.

Regarding the portfolio construction models we use in this paper, input variables are each stock's historical daily returns in the IS period. The M-V portfolio model is one of the most important approaches in portfolio selection and optimization proposed by Markowitz and Markowitz (1952). The aim is to optimize the asset weights to achieve the highest expected returns with lower volatility. In our case, the outputs are the rolling differences between two mean-variance portfolio returns: $MV\,(t - T_1 + 1\colon t + 1) - MV\,(t - T_1\colon t)$, where T_1 is the number of days in the first month of the sample period. Then, we fit the models described below to find the optimal weight according to the Shapley value of each stock, which indicates the significance of each stock for the models. We use grid search and fivefold cross-validation to tune the best parameters for each algorithm. The GridSearchCV function in Python is used in our case.

Feature importance is an essential criterion in machine learning algorithms. In our case, we use feature selection to define the weight of each stock in our pre-selected portfolio. SHapley Additive exPlanation (SHAP) proposed by Lundberg and Lee (2017) is an additive interpretation model inspired by the Shapley value, which can interpret the output of any machine learning model. Shapley value comes from the cooperative game theory, which is a fair metric that equitably distributes total rewards to each feature based on the contributions of different features (Shapley 1953). The key advantage of SHAP is that it can reflect the influence of features in each sample and show the positive and negative effects. This is calculated as follows:

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left[f_{S \cup \{j\}}(x_{S \cup \{j\}}) - f_S(x_S) \right], \tag{2}$$

where $|F|!$ is the overall permutation of all features, $|S|$ is the number of elements in the set S, $[f_{S \cup \{j\}}(x_{S \cup \{j\}}) - f_S(x_S)]$ is feature j's marginal contribution in participating in cooperation S, here is the feature j's marginal contribution to the prediction of the machine learning model.

Under the above framework, we employ a set of portfolio construction models. Our first approach is using a simple linear regression model. Then, we apply a shrinkage approach, in particular the LASSO model (Tibshirani 1996). We also use a GA model as proposed by Holland (1975) and a traditional SVR approach by Cortes and Vapnik (1995). Additionally, we experiment with decision tree methods such as CART and Random Forest (Breiman et al. 1984; Breiman 2001). Finally, we employ a back-propagation NN based on the premises of Rumelhart, Hinton and McClelland (1986). These are well-known approaches in the literature, but we provide their short descriptions and relevant specifications in Appendix A.[5]

5 | Empirical Results

As explained before, we sample our data at every 3- and 6-month intervals and then apply the models to evaluate the next 1-month portfolio performance. Tables 9 and 10 present the annualized Sharpe ratios in the SPX500 and STOXX600 accounting for 1 basis point (bp) transaction costs for each traded stock.

From the two tables, several interesting findings are brought forward. For both indices, equal-weighted sentiment-selected portfolios show consistently good performance, and screening using the SDG sentiment score indicates a generally better performance than the other two scores in the SPX500. Although a benchmark strategy, this trading performance shows that the sustainability screening based on the three sentiment indicators can be useful for stock trading. Generally, RF, NN, and GA present overall better results than other models. In the case of SPX500 (STOXX600), the best performance after 1 bp transaction costs is achieved with the selection of 10 stocks based on SDG scores and 6-month IS period with a Sharpe ratio of 1.42 (1.1468). The second best results in SPX500 are obtained by RF

TABLE 9 | Trading performance—Annualized Sharpe ratios (SPX500 −1 bp).

Model									
Num	**Score type**	**L**	**Linear**	**LASSO**	**GA**	**SVR**	**CART**	**RF**	**NN**
10	ESG	1.0966	1.0053	1.0037	1.2741	1.1335	1.2024	**1.3030**	1.2703
	SDG	1.1619	1.1208	1.1279	1.1785	1.1367	1.1344	1.2084	**1.2403**
	ALL	1.1547	0.9585	0.9597	1.2395	0.9341	0.9987	1.1800	**1.2540**
30	ESG	0.9282	1.1261	1.1246	1.0077	1.0144	0.6138	**1.1857**	1.1405
	SDG	0.9457	1.1616	1.1613	1.0049	0.7968	0.5645	1.1795	**1.1822**
	ALL	0.9285	1.0879	1.0915	0.9942	1.0674	0.8884	**1.1205**	1.1026
10	ESG	0.9594	0.9745	0.9754	0.9445	0.7616	0.9497	1.1387	**1.2947**
	SDG	1.1187	0.1772	0.1764	1.1069	1.0506	0.9708	1.1951	**1.4200**
	ALL	0.8994	0.8859	0.8836	0.9510	0.6226	0.7768	0.9430	**1.2357**
30	ESG	0.9355	0.8154	0.8164	0.9793	0.9129	0.7190	1.0260	**1.2576**
	SDG	1.0017	0.4038	0.4030	1.0045	0.7576	0.7460	**1.2241**	1.1145
	ALL	1.0119	0.6631	0.6641	1.0014	0.7106	0.7947	1.0249	**1.1472**

Note: The table presents the annualized Sharpe ratios for the constructed M-V portfolios out of the score-selected stocks in SPX500 after 1 bp transaction costs. 'IS Period' is the in-sample period used for model construction, which are 3-and 6-month, respectively. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the benchmark equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.

TABLE 10 | Trading performance—Annualized Sharpe ratios (STOXX600 −1 bp).

Model IS period	Num	Score type	Equal	Linear	LASSO	GA	SVR	CART	RF	NN
3 m	10	ESG	0.6539	0.5781	0.5767	0.6533	0.8183	0.5792	0.6577	**0.8956**
		SDG	0.6483	0.8312	0.8304	0.7160	0.4701	0.6847	0.8563	**0.9714**
		ALL	0.6693	0.4554	0.4533	**0.8220**	0.5072	0.5342	0.6956	0.7763
	30	ESG	0.6603	0.5375	0.5343	0.7671	0.7801	0.4679	0.7092	**1.0144**
		SDG	0.6089	0.2024	0.2025	0.7344	0.3312	0.3184	**0.8493**	0.6942
		ALL	0.6215	0.1648	0.1670	**0.7539**	0.5123	0.4185	0.6717	0.7127
6 m	10	ESG	0.6732	0.4365	0.4375	0.8901	0.5803	0.4959	0.7745	**0.9899**
		SDG	0.7661	0.3775	0.3756	1.0287	0.8899	0.7472	0.8357	**1.1468**
		ALL	0.9249	0.3965	0.3963	1.1114	0.8095	0.6430	0.9823	**1.1158**
	30	ESG	0.6379	0.3184	0.3163	**0.8331**	0.4760	0.4778	0.6677	0.6769
		SDG	0.8501	0.3828	0.3825	**0.9828**	0.7326	0.5751	0.9019	0.9722
		ALL	0.7366	0.3815	0.3798	**0.9098**	0.4283	0.5089	0.8201	0.8377

Note: The table presents the annualized Sharpe ratios for the constructed M-V portfolios out of the score-selected stocks in STOXX600 after 1 bp transaction costs. 'IS Period' is the in-sample period used for model construction, which are 3- and 6-month, respectively. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the benchmark equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.

using 10 stocks based on ESG scores and a 3-month IS period with a Sharpe ratio of 1.303. Equivalently in the case of STOXX600, the second best Sharpe ratio of 1.1158 is achieved by NN in the 6-month IS and using 10 stocks selected by the overall sentiment score. In general, machine learning techniques such as GA, RF and NN consistently provide better results than the benchmark model, the linear regression, and the LASSO. This finding is crucial as we prove that sustainability sentiment scores' investing based on machine learning methods can improve substantially portfolio performance.

For further insights, we present the equivalent results for Tables 9 and 10 of annualized returns in Appendix B. These results paint a similar picture to the above. Overall, machine learning models tend to have higher performance in terms of annualized Sharpe ratios. Although the results show that machine learning models such as RF and NN cannot totally outperform the equal-weighted portfolios or linear approaches in each subset (as with returns), the average Sharpe ratios still perform well. In addition, GA seems to perform well when applied to the STOXX500.

When evaluating the results across the two indices, we observe that nonlinear models perform better when applied to the STOXX600 index. When 1 bp transaction costs are taken into account, in the SPX500, the average Sharpe ratio for linear models is 0.8653, which is 0.1736 greater than the average Sharpe ratio for nonlinear models (1.0389); in the STOXX600, the difference is 0.3155, with 0.4215 for linear models and 0.7369 for nonlinear models. Among the three scores, portfolios with ESG-related scores present the best performance in SPX500, while portfolios with SDG-related scores present the best performance in STOXX600. This finding is consistent with Brooks and Oikonomou (2018) and Chen and Xie (2022) that explore the beneficial outcomes of incorporating sustainability analysis into financial investment.

When considering transaction costs in reality, the number of stocks selected by sentiment scores becomes more important. The application of 1bs transaction costs is realistic, however, we also present performances without transaction costs for annualized returns and Sharpe ratios in Appendix C. When transaction costs are growing, 10-stock portfolios tend to have better performances than 30-stock portfolios, as more transaction costs are needed to adjust the weights in portfolios. As expected, the significant advantages of RF and NN are alleviated when transaction costs are growing. In general, machine learning models are more efficient in capturing nonlinear relationships in the data and fitting the model better. This requires more frequent adjustment of stocks' weights, which can squeeze the expected returns.

6 | Robustness

6.1 | Mean-Semivariance Model

In addition to the M-V model, Markovitz (1959) also introduces the concept of the semivariance of returns. He proposes that investors exhibit a greater degree of sensitivity towards downside risk, which is indicative of returns that fall below a specific threshold, as opposed to overall volatility. Therefore, the adoption of semivariance, instead of variance, results in the creation of enhanced portfolios.

$$\min SV(\omega_j) = \frac{1}{N}\sum_{i=1}^{N}\left(\min\left\{0, \sum_{j=1}^{M}\omega_j(r_{ij} - R_j)\right\}\right)^2$$
$$\text{s.t.}\begin{cases}\sum_{j=1}^{M}\omega_j R_j \geqslant \mu\\ \sum_{j=1}^{M}\omega_j = 1,\\ \omega_j \geqslant 0 j = 1, 2, \cdots, M.\end{cases} \quad (3)$$

TABLE 11 | Trading performance—Annualized Sharpe ratios (SPX500 −1 bp).

Model IS period	Num	Score type	Equal	Linear	LASSO	GA	SVR	CART	RF	NN
3 m	10	ESG	1.0966	0.2854	0.2839	**1.2147**	1.1267	0.6136	1.1360	1.1144
		SDG	1.1619	1.1457	1.1462	1.1460	1.0716	0.9847	**1.3249**	1.3101
		ALL	1.1547	0.9367	0.9374	1.1316	0.6011	0.7618	1.2446	**1.2887**
	30	ESG	0.9282	0.6831	0.6884	0.9884	0.8439	0.8249	**0.9997**	0.9415
		SDG	0.9457	1.1766	1.1838	1.0145	1.1001	0.7715	1.2152	**1.2282**
		ALL	0.9285	0.9075	0.9047	0.9866	0.8011	0.6122	**1.0744**	1.0082
6 m	10	ESG	0.9594	0.7272	0.7267	0.9106	0.8830	**1.0333**	0.9697	1.0139
		SDG	1.1187	0.6728	0.6728	1.1243	0.6944	0.9006	1.2329	**1.2416**
		ALL	0.8994	1.0454	1.0438	0.9570	1.0166	0.9039	**1.0773**	1.0733
	30	ESG	0.9355	1.0625	1.0625	1.0131	0.6482	0.9228	**1.1174**	1.0631
		SDG	1.0017	1.0921	1.0853	0.9866	0.7824	1.0654	1.0979	**1.1286**
		ALL	1.0119	1.3011	1.3015	1.0037	0.8017	0.8008	1.3171	**1.4436**

Note: The table presents the annualized Sharpe ratios for the constructed M-SV portfolios out of the score-selected stocks in SPX500 after 1 bp transaction costs. 'IS Period' is the in-sample period used for model construction, which are 3- and 6-month, respectively. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the benchmark equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.

TABLE 12 | Trading performance—Annualized Sharpe ratios (STOXX600 −1 bp).

Model IS period	Num	Score type	Equal	Linear	LASSO	GA	SVR	CART	RF	NN
3 m	10	ESG	0.6539	0.4492	0.4462	0.6299	0.5910	0.6423	0.7329	**0.8668**
		SDG	0.6483	0.9202	0.9203	0.7124	0.7799	0.8863	0.9364	**0.9468**
		ALL	0.6693	0.5530	0.5527	**0.7643**	0.4829	0.3044	0.6719	0.7280
	30	ESG	0.6603	0.6155	0.6159	0.7483	0.6450	0.5652	0.6815	**0.7583**
		SDG	0.6089	0.2017	0.1996	0.7421	0.6317	0.6678	**0.8138**	0.6236
		ALL	0.6215	0.3000	0.2986	**0.7077**	0.5240	0.4646	0.6279	0.6317
6 m	10	ESG	0.6732	0.5981	0.5935	**1.0463**	0.8310	0.6141	0.9289	0.9626
		SDG	0.7661	0.6847	0.6828	**1.1297**	0.4940	0.9075	0.9679	1.0535
		ALL	0.9249	0.7761	0.7742	**1.1948**	1.1477	1.1553	1.0564	0.9829
	30	ESG	0.6379	0.2698	0.2717	**0.8496**	0.4912	0.5837	0.6826	0.6527
		SDG	0.8501	0.3344	0.3384	**1.0294**	0.3980	0.8317	0.8539	0.8764
		ALL	0.7366	0.7145	0.7138	**0.8636**	0.4655	0.4810	0.8030	0.8103

Note: The table presents the annualized Sharpe ratios for the constructed M-SV portfolios out of the score-selected stocks in STTOXX600 after 1 bp transaction costs. 'IS Period' is the in-sample period used for model construction, which are 3- and 6-month, respectively. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the benchmark equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.

where R_j is the return of stock j, and μ is the minimum expected return of investors, and we use the risk-free rate here.

Tables 11 and 12 below show the annualized Sharpe ratio with M-SV models in SPX500 and STOXX600 respectively.

These results provide more evidence that machine learning models, especially RF, GA and NN, have better fitting ability than other models, and our strategy with these three algorithms is successful as it outperforms the equal-weighted portfolios and linear models in most cases. When it comes to GA in the M-SV approach, the performance seems more prominent than in the M-V method. The highest Sharpe ratio is obtained by NN (1.4436) in a 6-month IS period with 30 stocks selected based on the overall sentiment indicator for the case of SPX500. In STOXX600, the best Sharpe ratio is generated by GA (1.1948) in a 6-month IS period with 10 stocks selected based on the overall sentiment indicator. Moreover, GA shows superior performance in STOXX600 than in SPX500 in both M-V and M-SV methods. As with our main empirical results, we present in Appendix D also annualized returns with 1 bp transaction costs under the M-SV method, as well as the annualized performance without transaction costs in Appendix E.

TABLE 13 | Trading performance—Annualized Sharpe ratios (alternative IS/OOS).

Model										
Index	**Num**	**Score type**	**Equal**	**Linear**	**LASSO**	**GA**	**SVR**	**CART**	**RF**	**NN**
SPX500	10	ESG	0.6124	0.5497	0.5436	0.6478	0.6318	0.6130	**0.6354**	0.6989
		SDG	0.6148	0.5708	0.6312	0.6467	0.6287	**0.9938**	0.5751	0.6545
		ALL	0.5334	0.5810	0.5999	**0.8178**	0.6529	0.6164	0.8127	0.7020
	30	ESG	0.7110	0.7139	0.7222	0.8920	0.7684	0.7304	**0.9164**	0.8470
		SDG	0.7286	0.6460	0.7154	0.9661	0.8318	0.9136	0.8652	**1.0330**
		ALL	0.7200	1.0435	1.0871	1.1152	0.9213	0.9938	0.9042	**1.1313**
STOXX600	10	ESG	0.8142	0.5435	0.7475	0.8657	0.9574	0.7351	0.8183	**0.9897**
		SDG	0.7283	0.5991	0.5960	**0.8973**	0.7310	0.5559	0.8429	0.8857
		ALL	0.8873	0.6823	0.7164	**0.9816**	0.8785	0.6556	0.9676	0.9597
	30	ESG	0.7499	1.0065	1.0594	0.9365	1.0150	0.8526	1.0072	**1.1334**
		SDG	0.8279	0.8285	0.8590	0.9670	0.8972	**1.2175**	1.0089	1.0080
		ALL	0.8991	0.8916	0.9144	**1.0190**	0.9103	0.9065	0.9259	1.0047

Note: The table presents the annualized Sharpe ratios for the constructed M-V portfolios out of the score-selected stocks in SPX500 and STOXX600 based on a 56 and 26 months IS and OOS period respectively. 'Index' refers to the indices to which the selected stocks belong, which are SPX500 and STOXX600. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the benchmark equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.

6.2 | Alternative in-Sample/Out-of-Sample Split

As a final robustness check, we utilize two-thirds (56 months) of the total daily frequency data as the training set and the remaining one-third (26 months) as the test set to evaluate the consistency of our results. Table 13 below shows the annualized Sharpe ratio for SPX500 and STOXX600. The results further substantiate that machine learning models, particularly GA and NN, generally exhibit more effective fitting capabilities compared to alternative models. In addition, the tree models including CART and RF can also obtain high Sharpe ratios in some cases. The highest Sharpe ration in SPX500 is generated by NN(1.1313) with 30 stocks selected by the overall sentient score, while in STOXX600 is CART(1.2175) with 30 stocks selected by the SDG score. The empirical performance with the annualized returns also presented in Appendix F. This set of results indicates that portfolios constructed with machine learning models based on the sentiment scores generally yield better outcomes in terms of both returns and Sharpe ratios. These empirical findings are consistent with those of the previous section.

7 | Conclusions

The motivation behind this study stems from the growing significance of sentiment factors involving ESG and SDG, and machine learning techniques in the realm of financial investment. To cope with the potential heterogeneity of sustainability scores provided by various agents, we initially construct daily ESG, SDG, and overall sentiment indicators. To do so, we collect thousands of online news sources covering the firms selected in the SPX500 and STOXX600. After data cleaning and processing, more than 0.8 million daily sentiment scores are established through NPL for stocks in SPX500 and STOXX600 indices. Next, we investigate the efficacy of machine learning algorithms in constructing profitable portfolios, utilizing sustainability sentiment scores as a basis. In that effort, we utilize both linear models, such as OLS and LASSO, and nonlinear models, including GA, SVR, RF, CART and NN. The performances of the competing portfolios are compared with equal-weighted sentiment-selected portfolios.

Our analysis reveals several key findings. First, our results indicate that daily sentiment indicators developed through NLP have tangible value when used in financial investment. The use of ESG, SDG, and overall sentiment scores in portfolio screening has yielded positive trading performances for both SPX500 and STOX600, while the results are more prominent when checking portfolios with stocks from SPX500 over STOXX600. Looking at specific scores, portfolios featuring ESG-related scores exhibit higher performances in the SPX500, whereas portfolios featuring SDG-related scores demonstrate the most favourable performance in the STOXX600. Second, we show that portfolios constructed from nonlinear machine learning models outperform those constructed using equal-weighted and linear models in both indices. Notably, the use of NN, RF, and GA generally led to better performances compared to the other competing methods. The findings mentioned above highlight the significance of incorporating machine learning techniques in financial investment decision-making protocols. Finally, our results are consistent under the M-V and the M-SV framework, as well as the long-term test, indicating the robustness of our strategy.

Our study demonstrates the increasing significance of NLP in financial big data analytics, particularly in relation to the implementation of daily sentiment analysis. We also highlight the potential benefits of utilizing machine learning techniques with ESG and SDG-related sentiment screening in portfolio construction and investment decision-making processes. By incorporating advanced models and techniques, investors can potentially enhance their trading performances and manage effectively the risk of their investment portfolios. Our study further motivates the need for the design of adaptive sentiment systems leading to fast decisions accounting for rapid changes in news and the mood of investors. In that process, advanced machine learning and sophisticated algorithms can provide further arbitrage opportunities, which traditional strategies and models cannot capture. Finally, our paper shows that sustainability scores do have value, despite the recent uncertainty and scrutiny that rating agencies face. This is particularly important for ESG and SDG practitioners who continue to promote the benefits of sustainable investment and corporate governance.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Research data are not shared.

NOTES

[1]Matos (2020) reports that the collective assets under management of PRI's exceeded $80 trillion globally as of the end of 2019.

[2]See among others, Cochran and Wood (1984), Tang, Hull, and Rothenberg (2012), Torugsa, O'Donohue, and Hecker (2012), Nofsinger, Sulaeman, and Varma (2019), Chen, Dong, and Lin (2020), Bae, Choi, and Lim (2020), Doukas and Zhang (2023) and Gao, He, and Wu (2024).

[3]KLD formerly disseminated newsletters at inconsistent intervals, featuring the most notable CSR-related news articles curated by the analysts.

[4]TVL employs NLP to discern articles that are relevant to ESG concerns for particular companies, which are subsequently categorized based on distinct ESG matters.

[5]In this study, we do not focus on intraday data. Higher frequency forecasting and trading with our approach can lead to higher noise, impeding the ML models under study. Even if we cope with the higher noise, such an approach would yield computational demands that would make the trading application unrealistic, while it would lead to higher transaction costs.

References

Akyildirim, E., D. K. Nguyen, A. Sensoy, and M. Šikić. 2023. "Forecasting High-Frequency Excess Stock Returns via Data Analytics and Machine Learning." *European Financial Management* 29, no. 1: 22–75.

Albuquerque, R., Y. Koskinen, and C. Zhang. 2019. "Corporate Social Responsibility and Firm Risk: Theory and Empirical Evidence." *Management Science* 65, no. 10: 4451–4469.

Alessandrini, F., and E. Jondeau. 2021. "Optimal Strategies for ESG Portfolios." *The Journal of Portfolio Management* 47, no. 6: 114–138.

Apergis, N., T. Poufinas, and A. Antonopoulos. 2022. "ESG Scores and Cost of Debt." *Energy Economics* 112: 106186.

Ardia, D., K. Bluteau, K. Boudt, and K. Inghelbrecht. 2020. "Climate Change Concerns and the Performance of Green Versus Brown Stocks." *Management Science Articles in Advance* 69, no. 12: 7607–7632. https://pubsonline.informs.org/doi/10.1287/mnsc.2022.4636.

Atz, U., T. Van Holt, Z. Z. Liu, and C. C. Bruno. 2023. "Does Sustainability Generate Better Financial Performance? Review, Meta-Analysis, and Propositions." *Journal of Sustainable Finance & Investment* 13, no. 1: 802–825.

Avramov, D., S. Cheng, A. Lioui, and A. Tarelli. 2022. "Sustainable Investing with ESG Rating Uncertainty." *Journal of Financial Economics* 145, no. 2: 642–664.

Aziz, S., M. Dowling, H. Hammami, and A. Piepenbrink. 2022. "Machine Learning in Finance: A Topic Modeling Approach." *European Financial Management* 28, no. 3: 744–770.

Bae, J., W. Choi, and J. Lim. 2020. "Corporate Social Responsibility: An Umbrella or a Puddle on a Rainy Day? Evidence Surrounding Corporate Financial Misconduct." *European Financial Management* 26, no. 1: 77–117.

Bagnara, M. 2024. "Asset Pricing and Machine Learning: A Critical Review." *Journal of Economic Surveys* 38, no. 1: 27–56.

Baker, M., and J. Wurgler. 2006. "Investor Sentiment and the Cross-Section of Stock Returns." *The Journal of Finance* 61, no. 4: 1645–1680.

Ban, G. Y., N. El Karoui, and A. E. B. Lim. 2018. "Machine Learning and Portfolio Optimization." *Management Science* 64, no. 3: 1136–1154.

Bathia, D., and D. Bredin. 2013. "An Examination of Investor Sentiment Effect on G7 Stock Market Returns." *The European Journal of Finance* 19, no. 9: 909–937.

Bender, J., and T. Wang. 2016. "Can the Whole Be More Than the Sum of the Parts? Bottom-Up Versus Top-Down Multifactor Portfolio Construction." *The Journal of Portfolio Management* 42, no. 5: 39–50.

Berg, F., J. F. Kölbel, and R. Rigobon. 2022. "Aggregate Confusion: The Divergence of ESG Ratings." *Review of Finance* 26, no. 6: 1315–1344.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45: 5–32.

Breiman, L., J. Friedman, R. Olshen, and C. Stone. 1984. Classification and Regression Trees. Wadsworth Int. Taylor & Francis Group.

Brooks, C., and I. Oikonomou. 2018. "The Effects of Environmental, Social and Governance Disclosures and Performance on Firm Value: A Review of the Literature in Accounting and Finance." *The British Accounting Review* 50, no. 1: 1–15.

Capelle-Blancard, G., and A. Petit. 2019. "Every Little Helps? ESG News and Stock Market Reaction." *Journal of Business Ethics* 157: 543–565.

Capelle-Blancard, G., and S. Monjon. 2014. "The Performance of Socially Responsible Funds: Does the Screening Process Matter?," *European Financial Management* 20, no. 3: 494–520.

Chen, L., M. Pelger, and J. Zhu. 2024. "Deep Learning in Asset Pricing." *Management Science* 70, no. 2: 714–750.

Chen, T., H. Dong, and C. Lin. 2020. "Institutional Shareholders and Corporate Social Responsibility." *Journal of Financial Economics* 135, no. 2: 483–504.

Chen, Z., and G. Xie. 2022. "ESG Disclosure and Financial Performance: Moderating Role of ESG Investors." *International Review of Financial Analysis* 83: 102291.

Cheng, B., I. Ioannou, and G. Serafeim. 2014. "Corporate Social Responsibility and Access to Finance." *Strategic Management Journal* 35, no. 1: 1–23.

Clark, G. L., A. Feiner, and M. Viehs. 2015. From the Stockholder to the Stakeholder: How Sustainability Can Drive Financial Outperformance. Working Paper.

Cochran, P. L., and R. A. Wood. 1984. "Corporate Social Responsibility and Financial Performance." *Academy of Management Journal* 27, no. 1: 42–56.

Cornett, M. M., O. Erhemjamts, and H. Tehranian. 2016. "Greed or Good Deeds: An Examination of the Relation between Corporate Social Responsibility and the Financial Performance of US Commercial Banks Around the Financial Crisis." *Journal of Banking & Finance* 70: 137–159.

Cortes, C., and V. Vapnik. 1995. "Support-Vector Networks." *Machine Learning* 20: 273–297.

Da, Z., J. Engelberg, and P. Gao. 2011. "In Search of Attention." *The Journal of Finance* 66, no. 5: 1461–1499.

Da, Z., J. Engelberg, and P. Gao. 2015. "The Sum of all FEARS Investor Sentiment and Asset Prices." *Review of Financial Studies* 28, no. 1: 1–32.

DeMiguel, V., J. Gil-Bazo, F. J. Nogales, and A. A. P. Santos. 2023. "Machine Learning and Fund Characteristics Help to Select Mutual Funds with Positive Alpha." *Journal of Financial Economics* 150, no. 3: 103737.

Dergiades, T. 2012. "Do Investors' Sentiment Dynamics Affect Stock Returns? Evidence from the US economy." *Economics Letters* 116, no. 3: 404–407.

D'Hondt, C., R. De Winne, E. Ghysels, and S. Raymond. 2020. "Artificial Intelligence Alter Egos: Who Might Benefit from Robo-Investing?," *Journal of Empirical Finance* 59: 278–299.

Dimson, E., O. Karakaş, and X. Li. 2015. "Active Ownership." *Review of Financial Studies* 28, no. 12: 3225–3268.

Doukas, J. A., and R. Zhang. 2023. When Does CSR Payoff? *European Financial Management* 30, no. 4: 2242–2304.

Duarte, V., D. Duarte, and D. H. Silva. 2024. Machine Learning for Continuous-Time Finance. *The Review of Financial Studies*, forthcoming.

Edmans, A. 2011. "Does the Stock Market Fully Value Intangibles? Employee Satisfaction and Equity Prices." *Journal of Financial Economics* 101, no. 3: 621–640.

Engle, R. F., S. Giglio, B. Kelly, H. Lee, and J. Stroebel. 2020. "Hedging Climate Change News." *The Review of Financial Studies* 33, no. 3: 1184–1216.

Fama, E. F. 1970. "Efficient Market Hypothesis: A Review of Theory and Empirical Work." *Journal of Finance* 25, no. 2: 28–30.

Fang, H., C. P. Chung, Y. C. Lu, Y. H. Lee, and W. H. Wang. 2021. "The Impacts of Investors' Sentiments on Stock Returns Using Fintech Approaches." *International Review of Financial Analysis* 77: 101858.

Fang, M., and S. Taylor. 2021. "A Machine Learning Based Asset Pricing Factor Model Comparison on Anomaly Portfolios." *Economics Letters* 204: 109919.

Feng, Y., Y. Zhang, and Y. Wang. 2024. "Out-of-Sample Volatility Prediction: Rolling Window, Expanding Window, or Both?" *Journal of Forecasting* 43, no. 3: 567–582.

Feng, Z. Y., C. R. Chen, and Y. J. Tseng. 2018. "Do Capital Markets Value Corporate Social Responsibility? Evidence from Seasoned Equity Offerings." *Journal of Banking & Finance* 94: 54–74.

Flammer, C. 2021. "Corporate Green Bonds." *Journal of Financial Economics* 142, no. 2: 499–516.

Forni, M., A. Giovannelli, M. Lippi, and S. Soccorsi. 2018. "Dynamic Factor Model with Infinite-Dimensional Factor Space: Forecasting." *Journal of Applied Econometrics* 33, no. 5: 625–642.

Friede, G., T. Busch, and A. Bassen. 2015. "ESG and Financial Performance: Aggregated Evidence from More Than 2000 Empirical Studies." *Journal of Sustainable Finance & Investment* 5, no. 4: 210–233.

Gao, L., J. He, and J. Wu. 2024. "Standing Out from the Crowd via CSR Engagement: Evidence from Non-Fundamental-Driven Price Pressure." *Journal of Financial and Quantitative Analysis* 59, no. 1: 39–67.

Gao, X., C. Gu, and K. Koedijk. 2021. "Institutional Investor Sentiment and Aggregate Stock Returns." *European Financial Management* 27, no. 5: 899–924.

Giamouridis, D., and I. D. Vrontos. 2007. "Hedge Fund Portfolio Construction: A Comparison of Static and Dynamic Approaches." *Journal of Banking & Finance* 31, no. 1: 199–217.

Giannone, D., M. Lenza, and G. E. Primiceri. 2021. "Economic Predictions with Big Data: The Illusion of Sparsity." *Econometrica* 89, no. 5: 2409–2437.

Gibson Brandon, R., P. Krueger, and P. S. Schmidt. 2021. "ESG Rating Disagreement and Stock Returns." *Financial Analysts Journal* 77, no. 4: 104–127.

Gillam, R. A., J. B. Guerard, Jr., and R. Cahan. 2015. "News Volume Information: Beyond Earnings Forecasting in a Global Stock Selection Model." *International Journal of Forecasting* 31, no. 2: 575–581.

Görgen, M., A. Jacob, and M. Nerlinger. 2021. "Get Green or Die Trying? Carbon Risk Integration into Portfolio Management." *The Journal of Portfolio Management* 47, no. 3: 77–93.

Gu, S., B. Kelly, and D. Xiu. 2020. "Empirical Asset Pricing Via Machine Learning." *The Review of Financial Studies* 33, no. 5: 2223–2273.

Heston, S. L., and N. R. Sinha. 2017. "News vs. Sentiment: Predicting Stock Returns from News Stories." *Financial Analysts Journal* 73, no. 3: 67–83.

Hoepner, A. G. F., I. Oikonomou, Z. Sautner, L. T. Starks, and X. Y. Zhou. 2024. "ESG Shareholder Engagement and Downside Risk." *Review of Finance* 28, no. 2: 483–510.

Holland, J. H. 1975. *Adaptation in Natural and Artificial Systems*. Ann Arbor, Michigan: The University of Michigan Press.

Hong, H., and M. Kacperczyk. 2009. "The Price of Sin: The Effects of Social Norms on Markets." *Journal of Financial Economics* 93, no. 1: 15–36.

Huang, D., F. Jiang, J. Tu, and G. Zhou. 2015. "Investor Sentiment Aligned: A Powerful Predictor of Stock Returns." *Review of Financial Studies* 28, no. 3: 791–837.

Huck, N. 2019. "Large Data Sets and Machine Learning: Applications to Statistical Arbitrage." *European Journal of Operational Research* 278, no. 1: 330–342.

Humphrey, J. E., and D. D. Lee. 2011. "Australian Socially Responsible Funds: Performance, Risk and Screening Intensity." *Journal of Business Ethics* 102: 519–535.

Humphrey, J. E., and D. T. Tan. 2014. "Does It Really Hurt to Be Responsible?," *Journal of Business Ethics* 122: 375–386.

Jiang, F., J. Lee, X. Martin, and G. Zhou. 2019. "Manager Sentiment and Stock Returns." *Journal of Financial Economics* 132, no. 1: 126–149.

Kempf, A., and P. Osthoff. 2007. "The Effect of Socially Responsible Investing on Portfolio Performance." *European Financial Management* 13, no. 5: 908–922.

Khan, M. 2019. "Corporate Governance, ESG, and Stock Returns Around the World." *Financial Analysts Journal* 75, no. 4: 103–123.

Krauss, C., X. A. Do, and N. Huck. 2017. "Deep Neural Networks, Gradient-Boosted Trees, Random Forests: Statistical Arbitrage on the S&P 500." *European Journal of Operational Research* 259, no. 2: 689–702.

Krueger, P., Z. Sautner, and L. T. Starks. 2020. "The Importance of Climate Risks for Institutional Investors." *The Review of Financial Studies* 33, no. 3: 1067–1111.

Krüger, P. 2015. "Corporate Goodness and Shareholder Wealth." *Journal of Financial Economics* 115, no. 2: 304–329.

Kumar, A., J. K. Page, and O. G. Spalt. 2013. "Investor Sentiment and Return Comovements: Evidence from Stock Splits and Headquarters Changes." *Review of Finance* 17, no. 3: 921–953.

Kynigakis, I., and E. Panopoulou. 2022. "Does Model Complexity Add Value to Asset Allocation? Evidence from Machine Learning Forecasting Models." *Journal of Applied Econometrics* 37, no. 3: 603–639.

Lin, W., and A. Taamouti. 2024. "Portfolio Selection under Non-Gaussianity and Systemic Risk: A Machine Learning Based Forecasting Approach." *International Journal of Forecasting* 40, no. 3: 1179–1188.

Lins, K. V., H. Servaes, and A. Tamayo. 2017. "Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis." *The Journal of Finance* 72, no. 4: 1785–1824.

Loughran, T., and B. McDonald. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *The Journal of Finance* 66, no. 1: 35–65.

Lundberg, S. M., and S. I. Lee. 2017. "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems* 30: 4768–4777.

Luo, D. 2022. "ESG, Liquidity, and Stock Returns." *Journal of International Financial Markets, Institutions and Money* 78: 101526.

Luo, H. A., and R. J. Balvers. 2017. "Social Screens and Systematic Investor Boycott Risk." *Journal of Financial and Quantitative Analysis* 52, no. 1: 365–399.

Madhavan, A., A. Sobczyk, and A. Ang. 2021. "Toward ESG Alpha: Analyzing ESG Exposures through a Factor Lens." *Financial Analysts Journal* 77, no. 1: 69–88.

Markovitz, H. M. 1959. *Portfolio Selection: Efficient Diversification of Investments*. New Haven, Connecticut: Yale University Press. http://www.jstor.org/stable/j.ctt1bh4c8h.

Markowitz, H., and H. Markowitz. 1952. "Portfolio Selection." *The Journal of Finance* 7: 77–91.

Matos, P. 2020. ESG and Responsible Institutional Investing around the World: A Critical Review. CFA Institute Research Foundation.

Mbanga, C., A. F. Darrat, and J. C. Park. 2019. "Investor Sentiment and Aggregate Stock Returns: the Role of Investor Attention." *Review of Quantitative Finance and Accounting* 53: 397–428.

McCulloch, W. S., and W. Pitts. 1943. "A Logical Calculus of the Ideas Immanent in Nervous Activity." *Bulletin of Mathematical Biology* 5: 115–133.

Ng, A. C., and Z. Rezaee. 2015. "Business Sustainability Performance and Cost of Equity Capital." *Journal of Corporate Finance* 34: 128–149.

Nofsinger, J. R., J. Sulaeman, and A. Varma. 2019. "Institutional Investors and Corporate Social Responsibility." *Journal of Corporate Finance* 58: 700–725.

Pástor, Ľ., R. F. Stambaugh, and L. A. Taylor. 2022. "Dissecting Green Returns." *Journal of Financial Economics* 146, no. 2: 403–424.

Renault, T. 2017. "Intraday Online Investor Sentiment and Return Patterns in the US Stock Market." *Journal of Banking & Finance* 84: 25–40.

Rumelhart, D. E., G. E. Hinton, and J. L. McClelland. 1986. "A General Framework for Parallel Distributed Processing." *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* 1: 45–76.

Schmeling, M. 2009. "Investor Sentiment and Stock Returns: Some International Evidence." *Journal of Empirical Finance* 16, no. 3: 394–408.

Serafeim, G. 2020. "Public Sentiment and the Price of Corporate Sustainability." *Financial Analysts Journal* 76, no. 2: 26–46.

Serafeim, G., and A. Yoon. 2022. "Which Corporate ESG News Does the Market React To?," *Financial Analysts Journal* 78, no. 1: 59–78.

Shapley, L. S. 1953. "Stochastic Games." *Proceedings of the National Academy of Sciences* 39, no. 10: 1095–1100.

Sibley, S. E., Y. Wang, Y. Xing, and X. Zhang. 2016. "The Information Content of the Sentiment Index." *Journal of Banking & Finance* 62: 164–179.

Stambaugh, R. F., J. Yu, and Y. Yuan. 2012. "The Short of It: Investor Sentiment and Anomalies." *Journal of Financial Economics* 104, no. 2: 288–302.

Statman, M., and D. Glushkov. 2009. "The Wages of Social Responsibility." *Financial Analysts Journal* 65, no. 4: 33–46.

Sun, L., M. Najand, and J. Shen. 2016. "Stock Return Predictability and Investor Sentiment: A High-Frequency Perspective." *Journal of Banking & Finance* 73: 147–164.

Tang, D. Y., and Y. Zhang. 2020. "Do Shareholders Benefit from Green Bonds?," *Journal of Corporate Finance* 61: 101427.

Tang, Z., C. E. Hull, and S. Rothenberg. 2012. "How Corporate Social Responsibility Engagement Strategy Moderates the CSR–Financial Performance Relationship." *Journal of Management Studies* 49, no. 7: 1274–1303.

Tibshirani, R. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58, no. 1: 267–288.

Torugsa, N. A., W. O'Donohue, and R. Hecker. 2012. "Capabilities, Proactive CSR and Financial Performance in SMEs: Empirical Evidence from an Australian Manufacturing Industry Sector." *Journal of business ethics* 109: 483–500.

Trinks, P. J., and B. Scholtens. 2017. "The Opportunity Cost of Negative Screening in Socially Responsible Investing." *Journal of Business Ethics* 140: 193–208.

UNGA. 2015. Transforming our world: The 2030 agenda for sustainable development.

Wang, W., C. Su, and D. Duxbury. 2021. "Investor Sentiment and Stock Returns: Global Evidence." *Journal of Empirical Finance* 63: 365–391.

Wu, W., J. Chen, Z. Yang, and M. L. Tindall. 2021. "A Cross-Sectional Machine Learning Approach for Hedge Fund Return Prediction and Selection." *Management Science* 67, no. 7: 4577–4601.

Zhang, X., X. Zhao, and L. Qu. 2021. "Do Green Policies Catalyze Green Investment? Evidence from ESG Investing Developments in China." *Economics Letters* 207: 110028.

Zhang, Y., F. Ma, and Y. Liao. 2020. "Forecasting Global Equity Market Volatilities." *International Journal of Forecasting* 36, no. 4: 1454–1475.

Zhao, Q., C. Qin, L. Ding, Y. Y. Cheng, and S. Vătavu. 2023. "Can Green Bond Improve the Investment Efficiency of Renewable Energy?," *Energy Economics* 127: 107084.

Zografopoulos, L., M. C. Iannino, I. Psaradellis, and G. Sermpinis. 2025. Industry return prediction via interpretable deep learning. *European Journal of Operational Research* 321, no. 1: 257–268.

Appendix
Appendix A. Portfolio Construction Models

8.1.1 | A.1 Simple Linear Model

We start our analysis from the method with the simplest model, the simple linear regression model. Linear regression trains a set of model parameters **w** by minimizing the residual sum of squares between the observed values and the predicted values of the linear model. The objective function can then be expressed as:

$$\hat{\omega}_j = \arg\min_{\omega} \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{M} \omega_j x_{ij} \right)^2 \tag{A.1}$$

where ω_j is the weight.

We expect relatively poor results from this method, but we use this as a benchmark approach for the following more sophisticated algorithms. The parameter estimation of the OLS method depends on the linear independence between the features.

8.1.2 | A.2 Least Absolute Shrinkage and Selection Operator

The LASSO regression proposed by Tibshirani (1996) seeks to construct a penalty function and minimize the residual sum of squares under the constraint that the sum of absolute values of regression coefficients is less than a constant. This method makes some coefficients smaller and some coefficients with smaller absolute values are even directly shrank to zero. The minimization function is shown below:

$$\hat{\omega}_j = \arg\min_{\omega}\left(\sum_{i=1}^{N}\left(y_i - \sum_{j=1}^{M}\omega_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{M}|\omega_j|\right) \tag{A.2}$$

where ω_j is the weight, λ is the penalty coefficient, and $|\omega_j|$ represents the L1-norm.

LASSO can screen variables and selectively put variables into the model to obtain better performance parameters. It can also reduce the complexity of the model to avoid overfitting. The complexity of LASSO is determined by λ, which is the penalty coefficient. As the value of λ increases, the associated penalty for the regression model with a higher number of variables also increases. This enables us to derive a model with a reduced number of variables.

8.1.3 | A.3 Genetic Algorithm

GA was first designed and proposed by Holland (1975) according to the evolutionary law of organisms in nature. It is a computational model simulating the natural selection and genetic mechanism of Darwin's biological evolution theory, and a method of searching for the optimal solution by simulating the natural evolution process. The genetic algorithm starts the search process for the optimal solution from the initial population composed of many individuals. It continuously conducts crossover, mutation, selection, and other operations on the initial population gradually eliminating the solution with low fitness value and producing a new generation of population. To be more precise, our approach involves heuristic crossover, which leverages the fitness values (Sharpe ratio) of two parent chromosomes to determine the search direction. The ranking of parental performance progresses from the worst to the best. The offspring are created according to the equation:

$$\begin{cases}\text{Offspring}_A = \text{Best Parent} + \beta'(\text{Best Parent} - \text{Worst Parent})\\ \text{Offspring}_B = \text{Worst Parent} - \beta'(\text{Best Parent} - \text{Worst Parent})\end{cases}, \tag{A.3}$$

where β' is a random number between 0 and 1. $Offspring_A$ and $Offspring_B$

are two offspring chromosomes which are fractions of total capital assigned to each stock after crossover. Best Parent and Worst Parent are two parent chromosomes that have the highest and lowest fitness values separately.

GA has inherent implicit parallelism and better global optimization abilities. Additionally, they can adaptively adjust the search direction. Given that GA is typically employed for search-based optimization problems, the fitness function is iteratively computed to obtain an optimal solution, rather than constructing a particular model. As such, the parameters in GA are manually adjusted.

8.1.4 | A.4 Support Vector Regression

Support Vector Machines (SVM) introduced by Cortes and Vapnik (1995) is a binary classification model defined in the feature space with the largest interval. SVM utilizes kernel functions, which makes it a virtually nonlinear classifier. SVR is a regression application of SVM, which can tolerate the error between model output and real value up to the margin of tolerance. Only when the error exceeds the margin can the loss value be calculated. The SVR specification attempts to solve the following minimization problem:

$$\min_{\omega,\zeta}\frac{1}{2}\|\omega_j\|_2 + C\sum_{i=1}^{N}|\zeta_j| \\ \text{s.t.}|y_i - \omega_j x_{ij}| \le \epsilon' + |\zeta_j|, \forall\, i, \tag{A.4}$$

where $\|\omega_j\|$ is the magnitude of the normal vector to the surface. ϵ' denotes the margin of tolerance, which is defined as the maximum error that has been determined to be less than or equal to a predetermined magnitude. C serves as the penalty coefficient, whereby an increase in its value results in a corresponding increase in the tolerance level for points that fall outside the pre-specified range of ϵ'. ζ is the slack variable, denoting the deviation of any value that falls outside ϵ' from the margin to slack the model.

SVR can model nonlinear decision boundaries and also has strong robustness in overfitting. For this method, we use Gaussian kernel (RBF), and three basic parameters need to be tuned, which are penalty coefficient (C), coefficient of RBF (γ), and the margin of tolerance (ϵ'). One of the benefits of Support Vector Regression (SVR) is its capability to address nonlinear regression problems while exhibiting strong generalization ability.

8.1.5 | A.5 Classification and Regression Tree

According to Breiman et al. (1984), CART is a decision tree construction algorithm. CART is a learning method using the binary tree to output the conditional probability distribution of the target variable (y) under the given input variable (x). The regression tree uses the mean square error as the loss function. When the tree is generated, it will recursively divide the space according to the optimal feature and the optimal value under the optimal feature until the stop criterion is met. The minimization function is as follows:

$$\min_{v,s}\left[\min_{c_1}\sum_{x_i\in\mathbb{R}_1(v,s)}(y_i - c_1)^2 + \min_{c_2}\sum_{x_i\in\mathbb{R}_2(v,s)}(y_i - c_2)^2\right], \tag{A.5}$$

where c_1 and c_2 represent the mean of the outcomes for regions $\mathbb{R}_1$ and $\mathbb{R}_2$, respectively. These regions are determined by the splitting variable v and the splitting point s. The complexity of the CART model is managed through a pruning algorithm, which simplifies the fully grown decision tree by removing certain subtrees.

The objective function for pruning is given by:

$$\sum_{q=1}^{|K|}\sum_{x_i\in\mathbb{R}_q}(y_i - c_q)^2 + \lambda'|K|, \tag{A.6}$$

where c_q denotes the mean outcome within region $\mathbb{R}_q$ associated with the q-th leaf node. The term $|K|$ is the number of leaf nodes in the subtree K, and λ' is a penalty coefficient that controls the trade-off between the tree's complexity and its fit to the training data.

The CART model has good interpretability and high efficiency, but it is easy to overfit when the input data contains a large number of features.

8.1.6 | A.6 Random Forests

RFs were proposed by Breiman (2001). In essence, RF is an improvement of the decision tree algorithm, which combines multiple decision trees. It uses a bootstrap resampling approach to repeatedly and randomly select samples from the original training sample set N to generate a new training sample set. It generates B decision trees according to the selected sample set to form a random forest. The establishment of each tree depends on an independently selected sample, and each tree in the forest has the same distribution. The result of the new forest depends on the votes and average results of trees:

$$(x) = \frac{1}{B}\sum_{b=1}^{B} h_b(x), \tag{A.7}$$

where $h(x)$ represents the result of a decision tree b.

RF sacrifices the interpretability of decision trees for the shake of improved accuracy, without significantly increasing the computation time. It is insensitive to multicollinearity and the results are robust to missing and unbalanced data. In addition, even if each decision tree is not pruned, the random forest will not overfit.

8.1.7 | A.7 Neural Networks

McCulloch and Pitts (1943) put forward the math model of neuron form and laid the foundation of NN. The NN is a multilayer unsupervised model, which uses the output features of the previous layer as the input to the next layer for feature learning. Better feature representation is achieved by mapping the features of the existing space samples to another feature space after layer-by-layer feature mapping. More specifically, Back-propagation Neural Network (BPNN) is a concept proposed by Rumelhart, Hinton, and McClelland (1986). It is a multilayer feedforward neural network trained using the error back-propagation algorithm. We use a 5-layer neural network model with ReLu activation function to achieve neuron de-linearization. The output of the hidden layer d is given by:

$$H_d = g\left(\sum_{p=1}^{P} w_{dp}x_p + a_d\right), d = 1, \ldots, D, \tag{A.8}$$

where the number of nodes in the input layer is P, the number of nodes in the hidden layer is D. w_{dp} represents the weight from input node p to hidden node d, and a_d is the bias associated with hidden node d. The function $g(x)$ is the activation function for the neural network, defined as the ReLU function:

$$g(x) = \max(0, x), \tag{A.9}$$

The output of the output layer node l is:

$$O_l = \sum_{d=1}^{D} H_d w_{ld} + b_l, l = 1, \ldots, L, \tag{A.10}$$

where w_{ld} is the weight from hidden node d to output node l, and b_l is the bias associated with output node l, and the number of nodes in the output layer is L.

The cost function, used to measure the error of the network, is defined as:

$$E = \frac{1}{2}\sum_{l=1}^{L}(Y_l - O_l)^2, \tag{A.11}$$

where Y_l represents the actual value for output node l

The weight update rule, using gradient descent, is given by:

$$w_{ld} := w_{ld} - \eta\frac{\partial E}{\partial w_{ld}}, \tag{A.12}$$

where η is the learning rate.

BPNNs have multiple nonlinear mapping feature transformations that can be fitted to highly complex functions.

Appendix B

Mean-Variance model (Annualized returns −1 bp)

The annualized returns with 1 bp transaction costs in SPX500 and STOXX600 are presented in Tables B.1 and B.2 separately.

TABLE B.1 | Annualized returns (SPX500 −1 bp).

Model										
IS period	**Num**	**Score type**	**Equal**	**Linear**	**LASSO**	**GA**	**SVR**	**CART**	**RF**	**NN**
3 m	10	ESG	18.81%	23.42%	23.40%	24.13%	23.32%	**28.94%**	26.90%	24.78%
		SDG	20.35%	24.57%	24.74%	23.66%	22.82%	**27.63%**	23.84%	24.38%
		ALL	22.25%	21.90%	21.94%	26.82%	22.80%	22.48%	24.93%	**27.52%**
	30	ESG	16.00%	26.37%	**26.38%**	18.56%	20.30%	13.39%	21.81%	21.22%
		SDG	16.75%	25.46%	**25.49%**	18.99%	13.64%	11.98%	20.71%	22.01%
		ALL	16.95%	26.59%	**26.67%**	19.26%	21.64%	18.16%	20.21%	21.30%
6 m	10	ESG	19.08%	24.84%	24.88%	22.32%	17.18%	23.24%	26.46%	**30.39%**
		SDG	20.77%	5.13%	5.11%	23.28%	23.60%	21.68%	23.11%	**28.23%**
		ALL	19.92%	**34.00%**	33.94%	24.88%	18.67%	19.78%	20.91%	28.46%
	30	ESG	16.92%	16.84%	16.86%	19.97%	16.90%	16.73%	18.50%	**22.65%**
		SDG	18.76%	13.04%	13.03%	20.76%	16.58%	16.24%	**23.25%**	22.30%
		ALL	18.74%	15.56%	15.59%	**20.85%**	15.88%	20.66%	20.33%	20.48%

Note: The table presents the annualized returns for the constructed M-V portfolios out of the score-selected stocks in SPX500 after 1 bp transaction costs. 'IS Period' is the in-sample period used for model construction, which are 3- and 6-month, respectively. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the benchmark equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.

TABLE B.2 | Annualized returns (STOXX600 −1 bp).

Model IS period	Num	Score type	Equal	Linear	LASSO	GA	SVR	CART	RF	NN
3 m	10	ESG	11.78%	13.29%	13.27%	12.61%	**16.86%**	11.89%	14.61%	16.78%
		SDG	11.56%	18.18%	18.17%	13.01%	10.12%	13.68%	16.70%	**18.61%**
		ALL	12.45%	11.22%	11.19%	15.16%	9.70%	10.84%	**16.14%**	13.14%
	30	ESG	11.16%	14.19%	14.12%	13.09%	14.50%	9.34%	12.94%	**19.58%**
		SDG	10.12%	4.95%	4.95%	12.28%	5.97%	6.95%	**14.95%**	12.39%
		ALL	10.37%	4.65%	4.71%	**12.54%**	9.39%	8.61%	11.60%	11.91%
6 m	10	ESG	11.58%	11.00%	11.03%	15.70%	11.24%	9.42%	14.29%	**19.86%**
		SDG	13.16%	9.49%	9.45%	17.80%	17.00%	16.96%	17.56%	**22.03%**
		ALL	14.78%	10.05%	10.04%	17.96%	15.62%	11.90%	18.85%	**22.17%**
	30	ESG	10.89%	9.14%	9.10%	**14.07%**	9.22%	10.65%	12.13%	11.26%
		SDG	13.50%	8.82%	8.82%	16.08%	12.00%	10.91%	14.39%	**18.63%**
		ALL	11.99%	8.68%	8.67%	14.84%	7.59%	10.23%	14.64%	**15.73%**

Note: The table presents the annualized returns for the constructed M-V portfolios out of the score-selected stocks in STOXX600 after 1 bp transaction costs. 'IS Period' is the in-sample period used for model construction, which are 3- and 6-month, respectively. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the benchmark equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.

TABLE C.1 | Annualized returns (SPX500—no transaction costs).

Model IS period	Num	Score type	Equal	Linear	LASSO	GA	SVR	CART	RF	NN
3 m	10	ESG	19.69%	24.28%	24.26%	24.56%	24.42%	**30.16%**	27.91%	25.28%
		SDG	21.21%	25.75%	25.90%	24.06%	23.92%	**28.89%**	24.86%	24.94%
		ALL	23.03%	22.92%	22.92%	27.22%	23.97%	23.72%	26.05%	**28.07%**
	30	ESG	18.20%	30.46%	**30.49%**	19.48%	23.93%	17.87%	25.55%	24.72%
		SDG	18.83%	29.42%	**29.45%**	19.90%	17.19%	16.79%	24.49%	25.60%
		ALL	18.94%	30.61%	**30.68%**	20.23%	25.32%	22.76%	24.03%	24.86%
6 m	10	ESG	19.62%	25.83%	25.87%	22.68%	18.40%	24.58%	27.76%	**31.08%**
		SDG	21.21%	6.34%	6.33%	23.63%	24.98%	23.17%	24.50%	**29.13%**
		ALL	20.39%	35.19%	**35.21%**	25.25%	19.98%	21.32%	22.27%	28.93%
	30	ESG	18.10%	21.23%	21.26%	20.73%	21.19%	22.20%	22.89%	**26.76%**
		SDG	19.92%	17.21%	17.21%	21.50%	20.85%	21.58%	**27.54%**	26.14%
		ALL	19.90%	19.97%	20.01%	21.61%	20.19%	**26.16%**	24.45%	24.48%

Note: The table presents the annualized returns for the constructed M-V portfolios out of the score-selected stocks in SPX500. 'IS Period' is the in-sample period used for model construction, which are 3- and 6-month, respectively. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.

Appendix C

Mean-Variance model (Annualized results—no transaction costs)

Tables C.1 and C.2 present the annualized returns without the consideration of transaction cost in SPX500 and STOXX600 separately, while Tables C.3 and C.4 present the annualized Sharpe ratios.

Appendix D

Mean-Semivariance model (Annualized returns −1 bp)

Tables D.1 and D.2 present the annualized returns of portfolios constructed with the M-SV model.

Appendix E

Mean-Semivariance model (Annualized results—no transaction costs)

Tables E.1 and E.2 present the annualized returns without the consideration of transaction cost in SPX500 and STOXX600 separately, while Tables E.3 and E.4 present the annualized Sharpe ratios.

Appendix F

Alternative in-Sample/Out-of-Sample Split

(See Table F1)

TABLE C.2 | Annualized returns (STOXX600—no transaction costs).

Model IS period	Num	Score type	Equal	Linear	LASSO	GA	SVR	CART	RF	NN
3 m	10	ESG	12.55%	14.45%	14.45%	13.02%	**17.89%**	13.17%	15.62%	17.39%
		SDG	12.29%	19.39%	**19.39%**	13.45%	11.34%	14.91%	17.82%	19.16%
		ALL	13.16%	12.43%	12.42%	15.60%	10.91%	12.18%	**17.31%**	13.64%
	30	ESG	13.10%	18.36%	18.36%	14.05%	18.17%	13.93%	16.83%	**20.26%**
		SDG	11.96%	8.89%	8.86%	13.22%	9.87%	12.01%	**18.63%**	16.36%
		ALL	12.22%	8.66%	8.69%	13.45%	13.18%	13.30%	15.48%	**15.73%**
6 m	10	ESG	12.01%	12.25%	12.28%	16.06%	12.56%	10.81%	17.64%	**20.49%**
		SDG	13.56%	10.73%	10.69%	18.15%	18.40%	18.49%	18.89%	**22.97%**
		ALL	15.15%	11.40%	11.39%	18.33%	17.00%	13.31%	20.17%	**22.85%**
	30	ESG	12.02%	13.66%	13.62%	14.83%	13.25%	15.65%	**16.44%**	15.41%
		SDG	14.50%	13.46%	13.44%	16.87%	16.34%	16.23%	18.78%	**22.72%**
		ALL	13.01%	12.99%	12.98%	15.63%	11.84%	15.50%	19.09%	**19.57%**

Note: The table presents the annualized returns for the constructed M-V portfolios out of the score-selected stocks in STOXX600. 'IS Period' is the in-sample period used for model construction, which are 3- and 6-month, respectively. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.

TABLE C.3 | Annualized Sharpe ratios (SPX500—no transaction costs).

Model IS period	Num	Score type	Equal	Linear	LASSO	GA	SVR	CART	RF	NN
3 m	10	ESG	1.1490	1.0441	1.0426	1.2984	1.1897	1.2553	**1.3554**	1.2951
		SDG	1.2136	1.1769	1.1826	1.1994	1.1938	1.1883	1.2617	**1.2690**
		ALL	1.1972	1.0059	1.0051	1.2585	0.9835	1.0558	1.2355	**1.2793**
	30	ESG	1.0637	1.3137	1.3134	1.0600	1.2035	0.8334	**1.3958**	1.3343
		SDG	1.0689	1.3519	1.3511	1.0560	1.0185	0.8094	**1.4055**	1.3814
		ALL	1.0434	1.2600	1.2622	1.0469	1.2540	1.1257	**1.3420**	1.2928
6 m	10	ESG	0.9877	1.0148	1.0158	0.9603	0.8192	1.0061	1.1962	**1.3232**
		SDG	1.1436	0.2285	0.2278	1.1240	1.1139	1.0403	1.2682	**1.4656**
		ALL	0.9204	0.9195	0.9197	0.9662	0.6693	0.8421	1.0080	**1.2559**
	30	ESG	1.0052	1.0449	1.0458	1.0186	1.1601	0.9660	1.2799	**1.4978**
		SDG	1.0670	0.5438	0.5431	1.0420	0.9615	1.0084	**1.4601**	1.3135
		ALL	1.0770	0.8666	0.8674	1.0406	0.9169	1.0170	1.2429	**1.3791**

Note: The table presents the annualized Sharpe ratios for the constructed M-V portfolios out of the score-selected stocks in SPX500. 'IS Period' is the in-sample period used for model construction, which are 3- and 6-month, respectively. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.

TABLE C.4 | Annualized Sharpe ratios (STOXX600—no transaction costs).

Model IS period	Num	Score type	Equal	Linear	LASSO	GA	SVR	CART	RF	NN
3 m	10	ESG	0.7000	0.6309	0.6309	0.6762	0.8720	0.6452	0.7057	**0.9317**
		SDG	0.6920	0.8922	0.8921	0.7419	0.5329	0.7505	0.9170	**1.0002**
		ALL	0.7102	0.5088	0.5079	**0.8480**	0.5767	0.6058	0.7485	0.8072
	30	ESG	0.7840	0.7076	0.7076	0.8284	0.9894	0.7236	0.9376	**1.0511**
		SDG	0.7293	0.3969	0.3949	0.7958	0.5853	0.5824	**1.0746**	0.9364
		ALL	0.7420	0.3389	0.3397	0.8132	0.7411	0.6720	0.9138	**0.9602**

(Continues)

TABLE C.4 | (Continued)

Model IS period	Num	Score type	Equal	Linear	LASSO	GA	SVR	CART	RF	NN
6 m	10	ESG	0.7010	0.4913	0.4922	0.9115	0.6544	0.5228	0.8529	**1.0226**
		SDG	0.7907	0.4324	0.4300	1.0498	0.9669	0.8187	0.9038	**1.1975**
		ALL	0.9500	0.4553	0.4553	1.1360	0.8855	0.7236	1.0549	**1.1500**
	30	ESG	0.7102	0.4941	0.4917	0.8806	0.7072	0.7227	0.9246	**0.9514**
		SDG	0.9181	0.6098	0.6085	1.0343	1.0197	0.8096	1.1951	**1.1959**
		ALL	0.8048	0.5933	0.5916	0.9611	0.7000	0.7965	**1.0872**	1.0566

Note: The table presents the annualized Sharpe ratios for the constructed M-V portfolios out of the score-selected stocks in STOXX600. 'IS Period' is the in-sample period used for model construction, which are 3- and 6-month, respectively. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.

TABLE D.1 | Annualized returns (SPX500 −1 bp).

Model IS period	Num	Score type	Equal	Linear	LASSO	GA	SVR	CART	RF	NN
3 m	10	ESG	18.81%	9.78%	9.75%	24.17%	**24.68%**	14.30%	23.16%	23.84%
		SDG	20.35%	25.68%	25.76%	22.95%	23.70%	21.36%	**26.15%**	23.86%
		ALL	22.25%	37.35%	**37.39%**	25.74%	13.83%	15.85%	26.62%	26.84%
	30	ESG	16.00%	17.96%	18.19%	**18.85%**	15.37%	14.65%	18.29%	18.36%
		SDG	16.75%	25.68%	25.81%	19.32%	19.01%	19.58%	**26.39%**	20.86%
		ALL	16.95%	**21.12%**	21.06%	19.54%	16.45%	14.21%	21.05%	20.76%
6 m	10	ESG	19.08%	18.45%	18.45%	21.36%	18.29%	**23.11%**	21.14%	21.43%
		SDG	20.77%	16.22%	16.22%	24.24%	17.77%	18.37%	23.97%	**24.25%**
		ALL	19.92%	**31.69%**	31.67%	26.92%	23.79%	23.98%	26.07%	25.80%
	30	ESG	16.92%	26.87%	**26.88%**	20.53%	12.95%	23.13%	20.88%	19.26%
		SDG	18.76%	21.22%	21.11%	20.68%	15.70%	21.41%	21.40%	**23.20%**
		ALL	18.74%	26.72%	**26.74%**	21.02%	15.77%	17.03%	23.48%	24.75%

Note: The table presents the annualized returns for the constructed M-SV portfolios out of the score-selected stocks in SPX500. 'IS Period' is the in-sample period used for model construction, which are 3- and 6-month, respectively. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.

TABLE D.2 | Annualized returns (STOXX600 −1 bp).

Model IS period	Num	Score type	Equal	Linear	LASSO	GA	SVR	CART	RF	NN
3 m	10	ESG	11.78%	10.36%	10.29%	11.89%	12.18%	12.85%	12.76%	**17.05%**
		SDG	11.56%	20.29%	**20.29%**	12.74%	15.77%	18.21%	16.85%	16.23%
		ALL	12.45%	16.13%	**16.13%**	14.49%	9.22%	7.88%	11.85%	13.72%
	30	ESG	11.16%	13.57%	13.58%	12.72%	12.31%	10.68%	12.26%	**15.20%**
		SDG	10.12%	5.07%	5.03%	12.30%	11.01%	13.29%	**13.72%**	10.96%
		ALL	10.37%	6.86%	6.84%	11.98%	**17.96%**	9.50%	10.61%	10.88%
6 m	10	ESG	11.58%	15.92%	15.84%	17.33%	17.96%	13.23%	**19.09%**	17.84%
		SDG	13.16%	17.53%	17.49%	18.81%	9.91%	**19.97%**	19.02%	18.71%
		ALL	14.78%	17.62%	17.61%	18.57%	**21.81%**	21.30%	19.86%	21.76%
	30	ESG	10.89%	7.44%	7.49%	**14.28%**	8.94%	12.23%	11.60%	12.07%
		SDG	13.50%	7.81%	7.89%	16.65%	7.45%	**17.24%**	14.11%	15.88%
		ALL	11.99%	13.48%	13.48%	**14.27%**	8.27%	10.51%	13.97%	14.10%

Note: The table presents the annualized returns for the constructed M-SV portfolios out of the score-selected stocks in STOXX600. 'IS Period' is the in-sample period used for model construction, which are 3- and 6-month, respectively. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.

TABLE E.1 | Annualized returns (SPX500—no transaction costs).

Model IS period	Num	Score type	Equal	Linear	LASSO	GA	SVR	CART	RF	NN
3 m	10	ESG	19.69%	10.69%	10.67%	24.58%	**25.65%**	15.31%	24.14%	24.28%
		SDG	21.21%	26.71%	26.79%	23.37%	24.70%	22.47%	**27.12%**	24.14%
		ALL	23.03%	38.52%	**38.56%**	26.18%	14.84%	17.13%	27.66%	27.45%
	30	ESG	18.20%	22.09%	**22.32%**	19.79%	18.69%	18.64%	21.69%	21.81%
		SDG	18.83%	29.75%	**29.88%**	20.24%	22.50%	24.19%	27.39%	24.48%
		ALL	18.94%	**25.33%**	25.28%	20.54%	19.87%	18.52%	24.46%	24.26%
6 m	10	ESG	19.62%	19.63%	19.63%	21.69%	19.50%	**24.54%**	22.34%	21.95%
		SDG	21.21%	17.52%	17.53%	24.61%	19.02%	19.68%	**25.22%**	24.92%
		ALL	20.39%	32.84%	**32.82%**	27.29%	24.93%	25.31%	27.26%	26.38%
	30	ESG	18.10%	31.26%	**31.27%**	21.29%	17.11%	28.09%	24.89%	23.16%
		SDG	19.92%	25.54%	25.44%	21.37%	19.87%	26.81%	25.43%	**27.16%**
		ALL	19.90%	30.87%	**30.89%**	21.79%	19.81%	22.26%	27.58%	28.61%

Note: The table presents the annualized returns for the constructed M-SV portfolios out of the score-selected stocks in SPX500. 'IS Period' is the in-sample period used for model construction, which are 3- and 6-month, respectively. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.

TABLE E.2 | Annualized returns (STOXX600—no transaction costs).

Model IS period	Num	Score type	Equal	Linear	LASSO	GA	SVR	CART	RF	NN
3 m	10	ESG	12.55%	11.49%	11.42%	12.30%	13.35%	14.02%	13.84%	**17.60%**
		SDG	12.29%	21.37%	**21.37%**	13.15%	16.80%	19.34%	18.01%	19.88%
		ALL	13.16%	17.35%	**17.36%**	14.90%	10.41%	9.07%	12.94%	14.32%
	30	ESG	13.10%	17.73%	17.74%	13.69%	15.80%	15.21%	15.77%	**18.82%**
		SDG	11.96%	9.37%	9.33%	13.28%	14.79%	**18.30%**	17.58%	14.79%
		ALL	12.22%	11.37%	11.35%	12.97%	12.95%	14.04%	14.33%	**14.72%**
6 m	10	ESG	12.01%	17.14%	17.06%	17.70%	19.22%	14.68%	**20.32%**	18.61%
		SDG	13.56%	18.83%	18.80%	19.19%	11.21%	**21.43%**	20.23%	19.33%
		ALL	15.15%	18.95%	18.95%	18.99%	**23.13%**	22.79%	21.23%	22.49%
	30	ESG	12.02%	12.00%	12.06%	15.05%	12.81%	**17.38%**	15.69%	15.67%
		SDG	14.50%	12.17%	12.25%	17.41%	11.62%	**22.76%**	18.33%	20.13%
		ALL	13.01%	18.15%	**18.16%**	15.09%	12.49%	15.98%	18.07%	17.96%

Note: The table presents the annualized returns for the constructed M-SV portfolios out of the score-selected stocks in STOXX600. 'IS Period' is the in-sample period used for model construction, which are 3- and 6-month, respectively. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.

TABLE E.3 | Annualized Sharpe ratios (SPX500—no transaction costs).

Model IS period	Num	Score type	Equal	Linear	LASSO	GA	SVR	CART	RF	NN
3 m	10	ESG	1.1490	0.3151	0.3135	**1.2380**	1.1745	0.6596	1.1881	1.1367
		SDG	1.2136	1.1921	1.1925	1.1675	1.1186	1.0368	**1.3761**	1.3252
		ALL	1.1972	0.9681	0.9689	1.1518	0.6476	0.8261	1.2962	**1.3207**
	30	ESG	1.0637	0.8482	0.8526	1.0403	1.0358	1.0664	**1.1940**	1.1266
		SDG	1.0689	1.3701	1.3772	1.0663	1.3155	0.9643	1.2626	**1.4544**

(Continues)

Model IS period	Num	Score type	Equal	Linear	LASSO	GA	SVR	CART	RF	NN
		ALL	1.0434	1.1003	1.0980	1.0399	0.9778	0.8103	**1.2580**	1.1852
6 m	10	ESG	0.9877	0.7765	0.7759	0.9254	0.9450	**1.1008**	1.0270	1.0366
		SDG	1.1436	0.7299	0.7298	1.1418	0.7451	0.9676	**1.2994**	1.2778
		ALL	0.9204	1.0851	1.0834	0.9712	1.0674	0.9567	**1.1279**	1.0992
	30	ESG	1.0052	1.2436	1.2435	1.0529	0.8740	1.1316	**1.3426**	1.2909
		SDG	1.0670	1.3272	1.3202	1.0219	1.0023	**1.3434**	1.3153	1.3267
		ALL	1.0770	1.5198	1.5200	1.0422	1.0204	1.0579	1.5629	**1.6729**

Note: The table presents the annualized returns for the constructed M-SV portfolios out of the score-selected stocks in SPX500. 'IS Period' is the in-sample period used for model construction, which are 3- and 6-month, respectively. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.

TABLE E.4 | Annualized Sharpe ratios (STOXX600—no transaction costs).

Model IS period	Num	Score type	Equal	Linear	LASSO	GA	SVR	CART	RF	NN
3 m	10	ESG	0.7000	0.5017	0.4987	0.6537	0.6523	0.7046	0.8004	**0.8937**
		SDG	0.6920	0.9738	0.9740	0.7376	0.8343	0.9455	1.0041	**1.1741**
		ALL	0.7102	0.5971	0.5968	**0.7878**	0.5519	0.3554	0.7397	0.7627
	30	ESG	0.7840	0.8183	0.8186	0.8101	0.8425	0.8251	0.8914	**0.9523**
		SDG	0.7293	0.4080	0.4057	0.8057	0.8672	0.9411	**1.0594**	0.8597
		ALL	0.7420	0.5267	0.5251	0.7709	0.7577	0.7130	0.8702	**0.8766**
6 m	10	ESG	0.7010	0.6464	0.6417	**1.0702**	0.8935	0.6861	0.9922	1.0078
		SDG	0.7907	0.7391	0.7371	**1.1534**	0.5650	0.9757	1.0322	1.0890
		ALL	0.9500	0.8378	0.8358	1.2233	1.2184	**1.2382**	1.1315	1.0183
	30	ESG	0.7102	0.4567	0.4583	0.8984	0.7258	0.8456	**0.9412**	0.8622
		SDG	0.9181	0.5460	0.5502	1.0788	0.6513	1.1121	**1.1260**	1.1212
		ALL	0.8048	0.9794	0.9785	0.9168	0.7312	0.7536	**1.0559**	1.0457

Note: The table presents the annualized returns for the constructed M-SV portfolios out of the score-selected stocks in STOXX600. 'IS Period' is the in-sample period used for model construction, which are 3- and 6-month, respectively. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.

TABLE F.1 | Trading performance (Annualized returns–alternative IS/OOS).

Model Index	Num	Score type	Equal	Linear	LASSO	GA	SVR	CART	RF	NN
SPX500	10	ESG	20.257%	16.568%	15.501%	16.906%	16.087%	14.874%	16.502%	**21.107%**
		SDG	19.088%	16.482%	18.520%	17.333%	18.134%	**22.670%**	17.885%	19.927%
		ALL	22.200%	17.018%	18.333%	**23.940%**	19.671%	21.134%	21.795%	22.467%
	30	ESG	20.520%	17.921%	18.272%	21.275%	18.719%	19.433%	21.185%	**21.517%**
		SDG	20.738%	19.143%	20.176%	**26.337%**	24.091%	22.170%	24.047%	25.320%
		ALL	23.138%	29.888%	30.053%	28.813%	26.318%	22.670%	22.985%	**31.278%**
STOXX600	10	ESG	16.261%	**22.330%**	14.445%	17.658%	20.217%	15.576%	15.461%	19.317%
		SDG	15.666%	**23.555%**	14.185%	**17.000%**	15.874%	13.718%	16.415%	17.418%
		ALL	15.866%	15.167%	15.200%	24.287%	18.082%	16.871%	23.937%	**20.386%**
	30	ESG	15.199%	22.330%	22.443%	19.365%	19.412%	22.085%	22.584%	**23.924%**

(Continues)

TABLE F.1 | (Continued)

Model										
Index	**Num**	**Score type**	**Equal**	**Linear**	**LASSO**	**GA**	**SVR**	**CART**	**RF**	**NN**
		SDG	15.783%	23.555%	23.565%	19.914%	18.620%	22.047%	21.932%	**24.285%**
		ALL	17.963%	23.518%	24.305%	19.618%	20.328%	18.946%	19.294%	**25.087%**

Note: The table presents the annualized returns for the constructed M-V portfolios out of the score-selected stocks in SPX500 and STOXX600 based on a 56 and 26 months IS and OOS period respectively. 'Index' refers to the indices to which the selected stocks belong, which are SPX500 and STOXX600. 'Num' represents the number of stocks selected for portfolio construction by different score types. 'Equal' refers to the results of the benchmark equal-weighted portfolios. The values in bold correspond to the best-performing model for each exercise.