ESG Investing: A Sentiment Analysis Approach

Stephane Goutte, Fei Liu, Hoang Viet Le, Hans-Jorg Von Mettenheim

ESG INVESTING: A SENTIMENT ANALYSIS APPROACH

Stéphane Goutte[b], Hoang-Viet Le[a,b,*], Fei Liu[c] and Hans-Jörg von Mettenheim[c,a,d]

[a]Keynum Investments, France
[b]Université Paris Saclay, UMI SOURCE, IRD, UVSQ, France
[c]IPAG Business School, France
[d]Oxford-Man Institute of Quantitative Finance, United Kingdom

Abstract

We analyze the predictability of news sentiment (both general news and ESG-related news) on the return of stocks from European and the potential of applying them as a proper trading strategy over seven years from 2015 to 2022. We find that sentiment indicators extracted from news supplied by GDELT such as Tone, Polarity, and Activity Density show significant relationships to the return of the stock price. Those relationships can be exploited, even in the most naive way, to create trading strategies that can be profitable and outperform the market. Furthermore, those indicators can be used as inputs for more sophisticated machine learning algorithms to create even better-performing trading strategies. Among the indicators, those extracted from ESG-related news tend to show better performance in both cases: when they are used naively or as inputs for machine learning algorithms.

Keywords: ESG, Stock Market Prediction, Sentiment Analysis, Machine Learning, Big Data, GDELT.

*Corresponding author: Hoang-Viet Le, viet.le@keynum.fr

1 INTRODUCTION

Financial time series forecasting, such as financial assets prediction, however, has always been a challenging task due to its volatile and chaotic nature. How to accurately predict the market trend and price movement is still a big question nowadays. Over the last decades, a plethora of research has been trying to refine the methodology for a good prediction. In 1952, the American economist Markowitz first raised the concept of portfolio theory in his paper Portfolio Selection (Markowitz, 1952). In his theory, he used mathematical approaches to formalize investors' preferences to explain investment diversification, and discussed portfolio selection and management systematically. Markowitz's works represent the beginning of using mathematical approaches to discover the logic behind the movement of the economic market. Not only focus on predicting the overall trend of the stock market, but many scholars also attempt to theorize the market price fluctuation by statistical methods. For example, Fung et al. (2002) used a piecewise segmentation algorithm to forecast the trends of the time series. Given the challenges in forecasting financial time series data along with the booming of AI technology, it is no surprise that the method of machine learning has been quickly applied to the field of financial market prediction and become such a popular topic. The work of Vijh et al. (2020) has shown that popular machine learning models such as random forests and neural networks are efficient in predicting US stock prices. Nabi et al. (2020) and Yang et al. (2020) also tried several popular machine-learning approaches to find the superiority of the Gradient Boosting models among others in stock price forecasting.

The input variables used in those mathematical models can come from traditional information such as past price, return, volatility, etc. They can also come from the two traditional schools of stock analysis such as technical analysis and fundamental analysis. Then comes the era of big data where huge volumes of data can be continuously collected from online social networks, multimedia, network sensors, mobile devices, etc. This phenomenon has provided opportunities for researchers to reach a plethora of new type of data that has never been available before. The term 'alternative' was used to describe the data that originate outside of the standard repertoire of market data but are considered useful for the stock price prediction(Hansen and Borch, 2022).

In this paper, we would like to create sentiment indicators from a publicly available source of news database called the GDELT project and use them to find the answer to the following research questions:

- RQ1: Will there be any difference between our newly created sentiment indicators created by general news or ESG-related news? And can they be used to explain the stock price movement?
- RQ2: If it is the case, can it be exploited to create profitable trading strategies?
- RQ3: Can the application of machine learning further improve the performance of trading strategies based on the same inputs?

For the first research question, our results show a significant relationship between both our newly created set of sentiment indicators and the price of around 600 stocks in the STOXX600 index. Based on statistical analysis, both sets of indicators show similar coefficients and levels of significance with the exception of the Activity Density variable where its ESG-related indicators are much more significant compared to the general one. Despite the similar statistical results, when a simple trading experiment was set up using the two sets of indicators, the ESG-related indicators tended to offer better results in risk-adjusted return. Still, trading strategies created using both of the two sets of indicators are profitable and perform much better compared to the STOXX600 index during the same period (February 2015 to September 2022). A more sophisticated trading experiment is then set up for the third research question where a few popular machine-learning algorithms are brought into action using the two sets of indicators as inputs. The result shows the superiority of machine learning models compared to naive strategies in most cases. It also shows that while using both sets of sentiment indicators offers the best result, if we can only choose one set of indicators, the set of ESG-related sentiment indicators is often the better choice.

The remainder of the paper is organized as follows. Section 2 discussed the literature regarding sentiment analysis and other works on similar topics. Section 3 presents data and the methodology for creating both of our sentiment variables. Section 4 provides the statistical results to answer the first research question. Section 5 explains the methodology of how both of our trading experiments are set up, Section 6 provides the results and discussion, and Section 7 concludes.

2 Literature Review

Sentiment analysis is the study of people's sentiments, opinions, and attitudes towards a specific entity such as individuals, organizations, products, etc (Liu, 2015). Since the beginning of the century, as textual data is the fastest-growing form of academic research, sentiment analysis has grown to be one of the most active research areas in natural language processing (NLP).

For the application of sentiment analysis in finance, text data is usually the most common source of data to be used which can either be collected from news feeds or social networks to generate the "sentiment" of that specific piece of information. In the sense of finance, that "sentiment" can either be the positive or negative expectation of the writer of a company, a market, a financial asset, etc. The classification of the "sentiment" can be pre-specified by a dictionary or by summing these weights into document-level sentiment scores (Ke et al., 2019). The sentiment score then can be further analyzed to see if it has any impact on the financial market such as the phenomenon of information transmission (Tetlock, 2014). The general sentiment regarding a specific object (for example, in the case of a company) may even influence the investment decision of a subset of the population toward that company. We cannot forget the infamous short squeeze of GameStop in January 2021 which is primarily triggered by users from a subreddit. There have been quite a few researches discussing the predictiveness of that subreddit's sentiment toward the GameStop stock's return and volatility (Long et al., 2021; Wang and Luo, 2021).

Regarding the sentiment analysis based on news, there are studies from Tuckett et al. (2014) and Shiller (2017) which suggested the positive and negative sentiment from newspapers can be used as an indicator for economic cycles. Baker et al. (2016, 2020) introduced new variables derived from news articles to explain the stock price volatility, investment rates, and employment growth. There are also works of Fraiberger et al. (2021) and Fronzetti Colladon et al. (2020) in which they used the sentiment from Reuters and Italian news articles to predict emerging stock markets and Italian stocks market respectively. And finally, we also have the work of Fronzetti Colladon and Elshendy (2017) and Tilly et al. (2021) where they analyzed news data from the GDELT project to forecast the macroeconomic indices. Several other studies have tried to apply sentimental analysis to the field of trading such as the work of Kazemian et al. (2016) and Mudinas et al. (2019).

Recent works are also finding a way to incorporate sentimental analysis into sustainable investing or ESG investing. Specifically, the work of Schmidt (2019) has shown a link between ESG-related news to stocks' financial performance. Capelle-Blancard and Petit (2019) did their analysis on more than 30,000 ESG news which covers around one hundred listed firms over the period 2002–2010. Their studies have shown that on average, firms facing negative events experience a drop in their market value of 0.1%, whereas companies gain nothing on average from positive announcements. Chen et al. (2020) estimate a stock-level ESG-beta using the stock return history and economy-wide ESG concerns. Most recently, we have the work of Kvam et al. (2022) where they create ESG scores based on ESG-related sentiment extracted from Google Trends, Twitter, and the VIX index. Their result points out the negative relationship between their ESG scores and stock returns but that relationship becomes positive in times of ESG concerns.

The method of extracting sentiment from text is also worth mentioning. Many studies analyzed the text data using the bag-of-words method where a sample of the text data is represented as the collection of its words, disregarding the order of those words in their sentences. However, due to the development of the deep learning methodology, the more sophisticated method of word embedding was introduced to retain the order of information in the text. For example, Wang and Sambasivan utilized sentiment analysis on the dataset collected from StockTwits, a social network for financial investors, by applying supervised learning on the message collection in StockTwits which are pre-classified as “Bullish” or “Bearish” (Wang et al., 2014). Word embedding is also a newer methodology of word representation for text analysis where it transforms the vector of dummy variables of each word into a real-valued vector that encodes the meaning of the word in such a way that the words with similar meanings will be placed closer in the vector space.

3 Data Collection

3.1 Dataset

As our study is about news sentiment as an indicator for stock return prediction, we need two types of datasets for this research. First is the most important choice of our study, the source of sentiment indicators. As we have discussed previously, we can have a sentiment from social networks or multimedia. And in this case, even if we focus on the news sentiment, we still have several different options to choose from. We can scrap news directly from newspaper websites, we can get them from RSS feed or we can get them from news data providers. In this study, we follow the work of Fronzetti Colladon and Elshendy (2017) and Tilly et al. (2021) where we use the GDELT project as the source of our news data.

The Global Database of Events, Languages, and Tone (GDELT) project is considered one of the largest, most comprehensive, and highest resolution open databases of human society created by Kalev H. Leetaru and George Town University. It is a research collaboration of Google Ideas, Google Cloud, Google, and Google News, the Yahoo! Fellowship at Georgetown University, BBC Monitoring, the National Academies Keck Futures Program, Reed Elsevier’s LexisNexis Group, JSTOR, DTIC, and the Internet Archive (Tilly et al., 2021). Currently, the GDELT Project has collected over a quarter-billion world news records in over 300 categories from 1979 to the present. The project’s vision is to leverage this data ”to construct a catalog of human societal-scale behavior and beliefs across all countries of the world, connecting every person, organization, location, count, theme, news source, and event across the planet into a single massive network that captures what’s happening around the world, what its context is and who’s involved, and how the world is feeling about it, every single day” (The GDELT Project, 2022). That is to say, the goal of the project is to record as many world events as possible along with their associated sentiment and context information.

In order to do that, GDELT uses algorithm technology supported by Google Cloud to collect online news from thousands of sources every 15 minutes, translating them to English and adding information regarding every person, organization, location, theme, and emotion connected to them. The project currently offers two main databases: the GDELT Events database and the GDELT Global Knowledge Graph. The former is the collection of events concerning cooperation and conflict relationships between two or more entities and the latter is the collection of news articles (Saz-Carranza et al., 2020). As our study is news-focused, the latter database will be the object of our study.

The GDELT GKG dataset contains more than three-quarters of a trillion news articles, which weights more than 2.5 TB, that are available from February 2015. The dataset offers around 30 different fields of information but due to the enormous size of the dataset, we selected only a few related fields as follows:

TABLE 1. GDELT GKG Dataset Fields Descriptions

Field	**Description**
GKGRECORDID	The unique ID of each article
DATE	The date time in YYYYMMDDHHMMSS format on which the article was published.
ORGANIZATIONS	The list of all company and organization names found in the text of the article
THEMES	The list of all Themes found in the document.
Positive Tone	The percentage of positive words in the article. Ranges from 0 to +100.
Negative Tone	The percentage of negative words in the article. Ranges from 0 to +100.
Tone	The average "tone" of the document as a whole. Ranges from -100 to +100.
Polarity	The indicator of how emotionally polarized the text of the article is. Ranges from 0 to +200
Activity Density	The percentage of active words in the article. Ranges from -100 to +100.
Word Count	The total number of words in the document

As in Table 1, we have the first two fields identifying the article and the time it is published. The ORGANIZATIONS field gives us the list of all companies mentioned in the articles. By using this field, we are able to assign the articles to the selected portfolio of stocks that we will analyze (stocks from the STOXX600 index). There is a lot of news that does not mention any stock company and there is also news that mentions several companies in the same article. After discarding the former, there are around 1.5 million articles remaining. The next five fields are the sentiment indicators that we would use for our analysis. GDELT calculates those indicators using their specific dictionary(lexicon) to identify positive and negative in the documents (bag-of-words method). After that, the overall tone is calculated by subtracting the negative tone from the positive tone. Hence, the tone is a score resulting from the balance between positive and negative words divided (Saz-Carranza et al., 2020). These scores range from -100 (extremely negative) to +100 (extremely positive) with 0 indicating neutral sentiment. There might be two possibilities if the tone is close to zero: it is either the text has a generally low emotional response or the positive and negative sentiment scores are at a similar level and they cancel out each other. These two scenarios can be differentiated using the positive or negative tone variable or the polarity variable (the sum of positive and negative tone which represent how polarized the article is). The final field is the total number of words each document contains.

As this paper not only focuses on general news regarding the stock companies but also aims to study the ESG-related news sentiments, it is also worth mentioning the THEMES field from GDELT. This field specifies the topics that are raised in the article. There can be several themes that can exist in a single and there can also be the case that the article has no theme which means that the topics of the article are not well-recognized by GDELT's algorithms. From this field, we can find different taxonomy glossaries such as the CrisisLex, World Bank Group taxonomies, or GDELT's own labeling system. At the time of writing, there are around 2,600 different themes offered by GDELT.*. Based on the themes of each article, we can then categorized them into different subtopics which is related to the ESG subject. In this paper, we use 17 Sustainable Development Goals (SDGs) which were set up by the United Nations General Assembly (UN-GA) in 2015. Those 17 SDGs are: No Poverty; Zero Hunger; Good Health and Well-being; Quality Education; Gender Equality; Clean water and Sanitation; Affordable and Clean Energy; Decent Work and Economic Growth; Industry, Innovation, and Infrastructure; Reduced Inequality; Sustainable Cities and Communities; Responsible Consumption and Production; Climate Action; Life Below Water; Life On Land; Peace, Justice, and Strong Institutions; and Partnerships for the Goals. All the articles that are founded to contain topics related to those 17 goals are then tagged as ESG-related news and will be used to construct a separate group of sentiment indicators.

As for stock price, we select the components from the STOXX600 index to represent the European stock market. The composition of the two indices is acquired from Blackrock's iShares website.† Regarding the stock price data, they are obtained from Yahoo Finance for the period from February 2015 to September 2022 not only for all the stocks of the indices but also for the price evolution of the indices themselves.

3.2 Data Processing

In this section, we provide details about the transformation process of the data before inputting them into the models. As our sentiment dataset only provides information regarding an individual news article, further processes are needed to aggregate them into our needed time interval (monthly in this case). Here, we calculate the sentiment indicators of each stock company for a specific time interval by taking the average sentiment variables of all the news articles that mentioned that stock company during that time interval. For example, the tone of Nestle in August 2021 is the average tone of all the articles that mentioned Nestle in August. This process is done for all the sentiment indicators except the positive tone and negative tone: Tone, Polarity, Activity Density, and Word Count. After the aggregation, we have 89 months of data for each stock company. For the European STOXX600, we have 570 stocks in total (less than 600 as there is some European companies' news that cannot be detected by our algorithm) with 40482 observations. As for the stock price data, we calculate the log return of every stock as well as the indices based on the open and close prices of each month.

4 Statistical Analysis

A necessary condition for our sentiment indicators to be useful to predict stock prices is that they at least have some correlation with future prices. To test this possibility, we first try to analyze if they have any linear relationship with the stock's subsequent return by very simple regression analysis. Based on the Capital Asset Pricing Theory (Fama and French, 2004), we create our regression formula as follows:

$$R_{i,t} = \beta_0 R_{M,t} + \beta_n Sentiment\_Indicator_{i,t-1} + \epsilon \quad (1)$$

where $R_{i,t}$ is the log return of each individual stock at time t and $R_{M,t}$ is the log return of the given stock index M at time t. $Sentiment\_Indicator_{i,t-1}$ is the value of each sentiment variable at time $t-1$ and ϵ is the error term. We used Pooled OLS method for the STOXX600 dataset as they seem

*The full list of the themes is available at http://data.gdeltproject.org/documentation/GKG-MASTER-THEMELIST.TXT

†https://www.blackrock.com/fr/particuliers/products/251931/ishares-stoxx-europe-600-ucits-etf-de-fund

to be the most appropriate. The results of our panel regression are illustrated in Table 2 and Table 3 for the general sentiment indicators and the ESG-related indicators respectively as follows:

TABLE 2. General Sentiment Indicators Regression Result

Dependent variable	Log Return (t+1)				
	(1)	(2)	(3)	(4)	(5)
Tone	**0.0011*** **			**0.0009*** **	**0.0009*** **
Polarity		**0.0003*** **		**0.0008*** **	**0.0008*** **
WordCount					-1.948e-07
Activity Density			**7.357e-05*** **	**-0.0001* **	**-0.0001* **
Index Return	**1.0802*** **	**1.0804*** **	**1.0804*** **	**1.0802*** **	**1.0802*** **

Table 2 presents the regression results of 6 different combinations of the 5 general sentiment indicators on the STOXX600 dataset. There can be much more combinations but for the sake of visibility, we only keep those that are significant. For example, with the Word Count variable, as it is not significant when introduced as an input, we only show their result on Regression (5) which includes all of the variables. From Regression (1) to (3), we have the regression coefficients of 3 individuals variable (Tone, Polarity, and Activity Density), and all of them are very highly significant ($p < 0.001$) when going by themselves. Regression (4) to (5) shows the 3 variables' coefficients when they are regressed with other variables. From Table 2, we can see that the Tone and Polarity variables are highly significant and show positive coefficients for every regression. Each unit of increase in Tone increases the next month's return of the stock in the STOXX600 index by 0.09% to 0.11% (0.03% to 0.08% in case of Polarity). The Activity Density variable is only highly significant when going alone but becomes less significant when regressed with other variables.

TABLE 3. ESG-related Indicators Regression Result

Dependent variable	Log Return (t+1)				
	(1)	(2)	(3)	(4)	(5)
SDG Tone	**0.0010*** **			**0.001*** **	**0.0009*** **
SDG Polarity		**0.0003*** **		**0.0009*** **	**0.001*** **
SDG WordCount					-6.44e-08
SDG Activity Density			**6.169e-05*** **	**-0.0002*** **	**-0.0002*** **
Index Return	**1.0801*** **	**1.0803*** **	**1.0803*** **	**1.0801*** **	**1.0801*** **

Table 3 illustrates the regression result using the ESG-related sentiment indicators using the same regressions method from Table 2. Here we can see a quite similar result between the two sets of sentiment indicators. Both the ESG-related Tone and the Polarity are highly significant and their coefficients are almost in the same range compared to that of the general set of indicators. The only difference between the two sets is that the Activity Density, in this case, is highly significant in all cases, where it is regressed individually or with other variables.

5 Trading Experiments

After getting the idea about how both sets of the sentiment variable may influence the future return from Section 4, in this section, we would like to set up some trading experiments to answer our remaining research objectives.

5.1 Naïve Trading Strategies

In Section 4, we found that Tone, Polarity, and Activity Density (to a lesser extent) seem to be the most significant sentiment variables from the GDELT dataset. As a result, we set up a very naïve trading strategy based on their coefficients as follows: the previous month's Tone, SDG Tone, Polarity, SDG Polarity, Activity Density, and SDG Activity Density will be used separately as the criteria to select the next month's investment portfolio. There will be 3 different types of trading strategies or ways of selecting a portfolio:

- **Long-Short strategy**: A market-neutral strategy where we long the top n stocks with the highest selected sentiment variable and short the bottom n stocks with the lowest variable.
- **Long-Index strategy**: Another market-neutral strategy where we also long the top n stocks but then we short the ETF of the corresponding index instead of the bottom n stocks.
- **Long Only strategy**: The normal strategy where we only long the top n stocks with the highest selected sentiment variable.

For each type of strategy, we will experiment with 5 different n in the list of $10, 30, ..., 90$. So in total, we will have 45 different trading strategies for two types of sentiment datasets. In order to evaluate the performance of the strategies, we use the annualized Sharpe ratio as the main metric and the annualized return is also given for reference. The performance metrics of the corresponding index will be used as benchmarks.

5.2 Machine learning Approaches

As the general idea of how effective the sentiment variables can be used in portfolio selection, we then use those sentiment variables as inputs for several popular machine learning models to see if they can help improve the performance of the strategies even further. Due to the fact that the regression result between the general sentiment indicators and the ESG-related indicators are quite similar, in this experiment, we decide to create three separate sets of input variables. The first set of input only contains only general sentiment indicators (Tone, Polarity, and Activity Density). The second set contains only ESG-related indicators (SDG Tone, SDG Polarity, and SDG Activity Density), and models created using this set of inputs are given the "_SDG" suffix. Finally, the third set contains all of the variables from both of the two previous sets and is given the "_All" suffix. The goal of using three different sets of input variables is to find out if the ESG-related indicators perform any better than the general one (as suggested by previous analysis) and if using both of them improves the performance significantly.

Regarding the machine learning specification, in this paper, we applied three popular machine learning methods which are linear regression, the support vector machine (SVM), and the Gradient Boosting based LightGBM to predict the next month's return of every stock using the current month's data. The predicted return of each stock will be the criteria to select the investment portfolio for the following months. We will also use the same 3 types of trading strategies (Long-Short, Long-Index, Long-Only) and the same set of n number of portfolio components explained in section 5.1. To prevent the model from overfitting, we split our dataset into two parts: the training set and the testing set. The training period is from February 2015 to December 2018 and the testing period is from January 2019 to September 2022.

All of our machine-learning models are built in Python. For linear regression, there are two popular libraries that are available which are Scikit-learn and Statmodels and we choose the algorithm provided by the Scikit-learn library using the default specification where the l2 regularization is applied to prevent overfitting. The same method is applied for the support vector machine model where we use also use the Scikit-learn library with default specification. For the gradient-boosting decision tree-based approach, there are several popular libraries with their own modified algorithms. The most popular and effective nowadays are the XGBoost library and the LightGBM library. In this study, we chose the LightGBM as it is more recent and offers better training time.

For this experiment, we also use the annualized Sharpe ratio as the main performance metric. The results from our machine learning models are compared directly with the performance of our naive strategies and the corresponding index during the same period to see if they can offer any improvement. As there will be too many results from several combinations of models and portfolio composition that it can get confusing and some models tend to outperform the others repeatedly, we will only present the top 5 best performing models (both machine learning and naive strategies) for each type of trading strategies and provide our conclusion based on that.

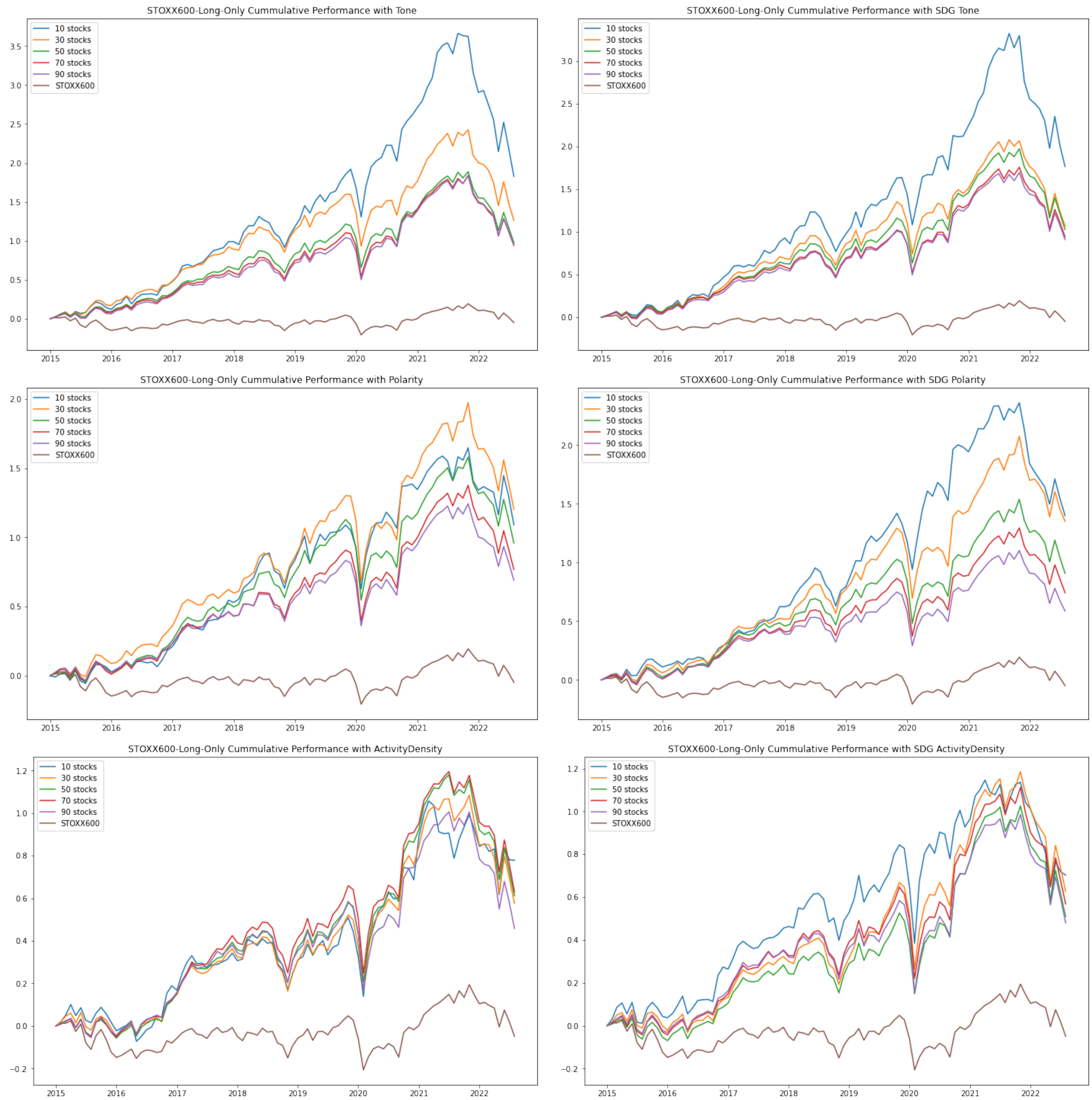
6 Result and Discussion

6.1 Naive Trading Strategies

Table 4 illustrates the performance metrics of the top 10 performing strategies. The strategies are divided by the type of strategies that we discussed in Section 5 and it is sorted in ascending order based on the Sharpe Ratio. The last row provided the performance metrics of the STOXX600 index during the same period as the context for comparison. We can see that the annualized Sharpe ratio of the STOXX600 during this period is only around 0.03 and its annualized return is only around 0.45% which is quite mediocre.

Figure 1 illustrates the cumulative return of our Long Only strategies during the period from February 2015 to September 2022. It is noticeable that the STOXX600 line always goes below other lines. This means all of our Long-Only strategies, no matter the number of stocks or sentiment variables, outperformed the STOXX600 in raw return during the period of study. As of Sharpe ratio, Table 4 gave the same conclusion as we have the top 10 of them ranging from 0.6 to 0.8 which is much higher compared to the 0.03 level of the STOXX600 index. Among the top 10 Long-Only strategies, half of them are from the general indicators sets and the other half are ESG-related. The general Tone with 10 stocks offered the highest yield and Sharpe ratio but three out of the top five strategies used ESG-related indicators. As a result, we can say that both the general indicators and ESG-related indicators are performing similarly for the Long-Only strategy. Another thing is that among the top 10, we can see no appearance of the Activity Density variable (ESG-related or not). This confirms the regression results from our previous analysis that Activity Density seems to be the inferior indicator compared to others. Regarding the effect of adding more stocks into the portfolio, we observe that for the Long-Only strategy, the fewer stocks, the better the return and Sharpe ratio of the portfolio. The best performing strategy, in this case, is based on Tone with only 10 stocks with an annual return of up to 16.57% and a Sharpe ratio of 0.811. The only drawback of these strategies is that they are not market-neutral, which means they can suffer from a drawdown in a bear market. It is obvious that during the first few months of 2020, there was a huge drop in performance for all of the strategies and the STOXX600 index.

Figure 1. Long Only Strategy Cumulative Return



As for the Long-Short type of strategy, the superiority of the strategies compared to the benchmark index is no longer obvious. From Figure 2, we can see that only the top two graphs, which are based on the Tone indicators (general or ESG-related), shows good results as all of the lines go above the line of the STOXX600 index. The next two graphs show the performance of the strategies that use the Polarity indicators and most of their performance is not that good. Unlike the strategies based on Tone that go up gradually, we see a lot of fluctuation and some strategies even perform worse than the STOXX600 which was not even performing well during the period. The last two graphs (Activity Density indicators) show even worse performance as most of the strategies in this case go below the index. This is not necessarily bad news, however, as in this case, if we reverse the signal (longing the stocks with the lowest Activity Density score and shorting those with the highest Activity Density), it is likely that we can have a positive performance. This phenomenon is reflected well in our regression analysis as the Activity Density coefficient is positive when going alone but negative when it is regressed with other variables. Maybe treating them as negative indicators is the right way to go. As this is the Long-Short strategy, from the 6 subfigures, we can see that among them only the Tone indicators show good performance and it is likely that they are the best among the three to differentiate between well-performing and poorly-performing stocks. Table 4 confirms that the top 10 Long-Short strategies are all based on the Tone indicator where the Sharpe ratio ranges from 0.6 to 1.2. Among them, the ESG-related indicator tends to perform better as the three top-performing ones are all ESG-related. The strategies also tend to perform better in raw return the less number of stocks in the portfolio. For the Sharpe ratio, however, it is not that obvious as the best-performing strategy is actually the one with the most stocks (90) with a Sharpe ratio of 1.276 despite not having the best annual return (6.02%). Due to its market-neutral nature, we also see a much better performance of them during the beginning of 2020 despite the big drawdown of the STOXX600 during the same period.

The last type of strategy, Long-Index, is also market-neutral and it seems to perform better than the previous type. From Figure 3, we can once again see that all of the lines are above the STOXX600 line. This means that all of our strategies outperform the STOXX600 index. This is confirmed in Table 4 where we have their Sharpe ratios range from 1.6 to 1.88 which are even higher than the best strategies that we have investigated from the start. However, in this case, Polarity tends to perform better as a portfolio selecting criteria as in the top 10 best performing strategies, 7 of which are based on Polarity. Given the fact that Tone performs better for the Long-Short strategy but Polarity is better for the Long-Index strategy and that the only difference between the two strategies is the shorting portfolio (the former shorts stocks with the lowest score while the latter shorts the index) and the longing portfolio is the same, we can infer that Polarity tends to be better for selecting well-performing stocks but is not as good for identifying poorly-performing stocks. Between the two sets of sentiment indicators, we once again observed better results from the ESG-related indicators. Six out of ten, three out of five, and the best performing strategy are based on ESG-related indicators. The market-neutral effect can also be observed but it is less obvious than that of the Long-Short strategies. We can still see a small drawdown at the beginning of 2020 but the magnitude is way smaller compared to the drawdown of the STOXX600 during the same period.

In conclusion, we observed generally good performance from almost all of the strategies with the exception of Long-Short strategies based on Polarity and Activity Density. Despite the fact that STOXX600 performed badly during the period of study, our portfolios created from its composition outperformed it considerably in both annual return and Sharpe ratio. Among our three types of trading strategies, even though the normal Long Only strategies give the best returns, they actually performed worse in risk-adjusted return in comparison with the two other market-neutral strategies. Among the sentiment variables, Tone generally provides better performance, with the exception in the case of Long-Index strategies where Polarity is proven to be the better choice. Finally, between the two sets of sentiment indicators, despite the fact that their regression results are quite similar, ESG-related indicators consistently provide better performance compared to the general sentiment indicators.

Figure 2. Long-Short Strategy Cumulative Return

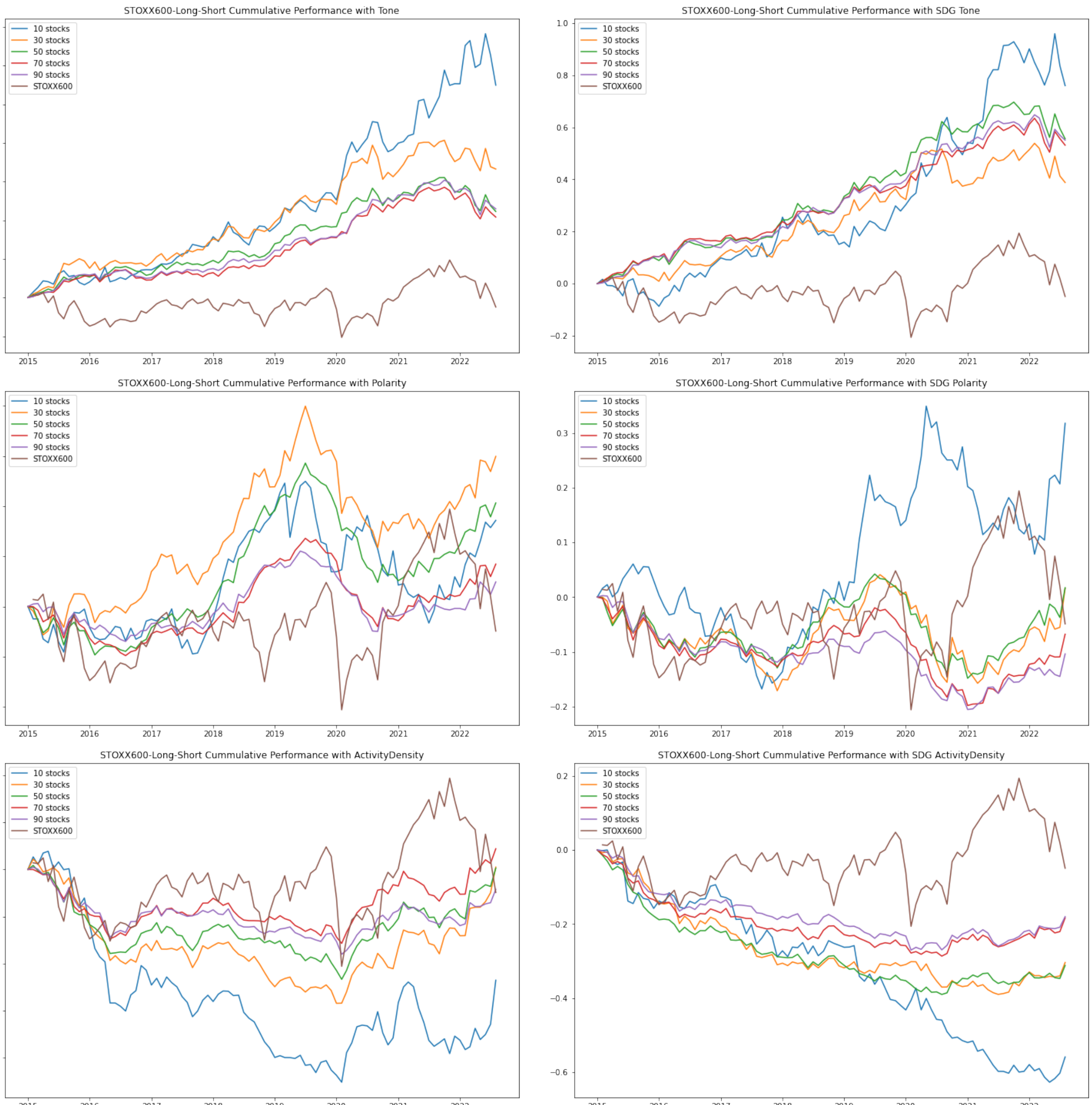


Figure 3. Long-Index Strategy Cumulative Return

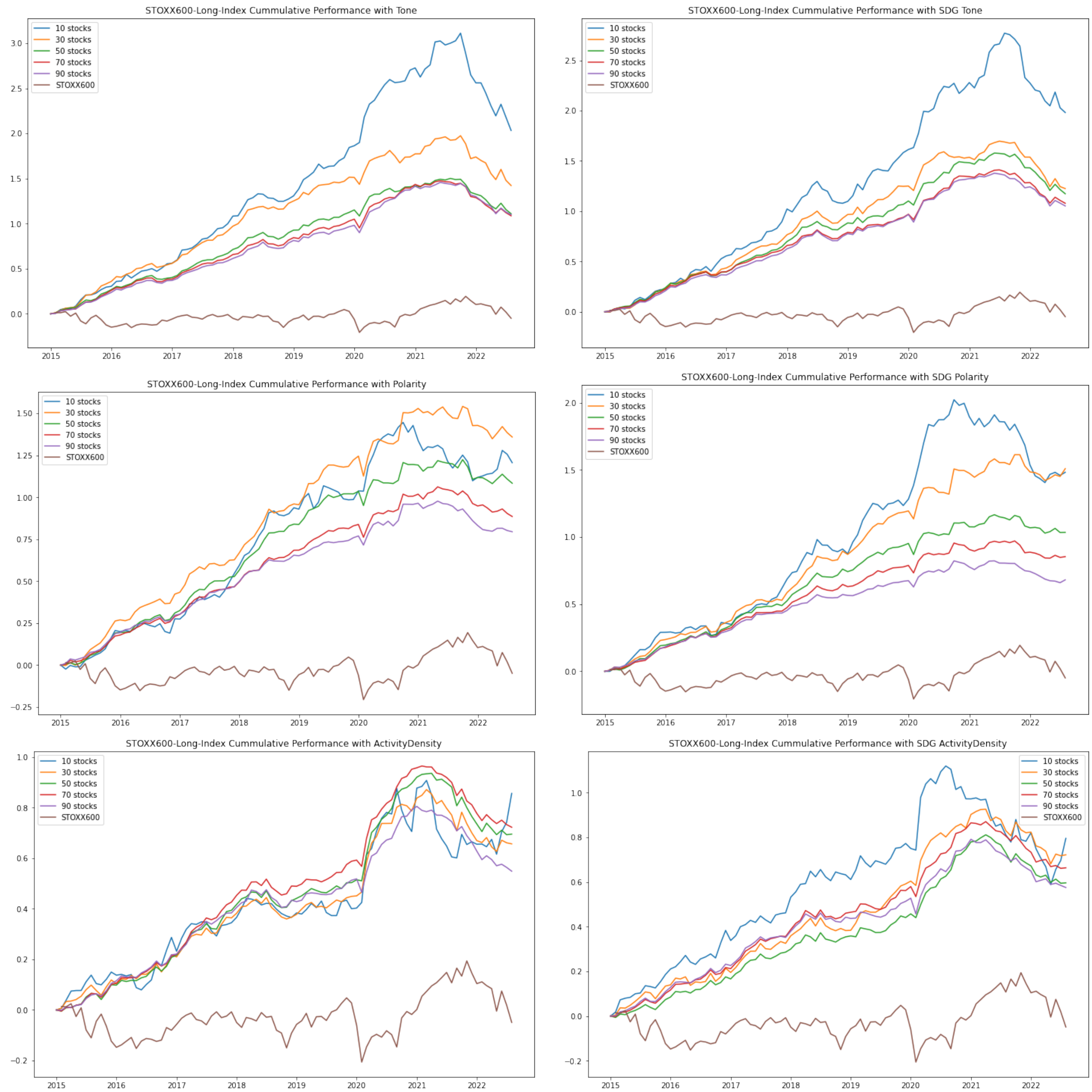


TABLE 4. Top Naïve Strategies Performance

Strategy	No of Stocks	Score Type	Annualized Return	Sharpe Ratio
Long-Short				
	30	SDG Tone	4.69%	0.587
	10	SDG Tone	8.33%	0.706
	50	Tone	5.15%	0.759
	70	Tone	4.82%	0.811
	30	Tone	7.27%	0.814
	10	Tone	10.82%	0.927
	90	Tone	5.20%	0.934
	50	SDG Tone	6.14%	0.951
	70	SDG Tone	5.86%	1.096
	90	SDG Tone	6.02%	1.181
Long-Only				
	50	Polarity	10.69%	0.609
	50	SDG Tone	11.36%	0.611
	30	SDG Tone	11.77%	0.614
	10	Polarity	11.73%	0.643
	30	Polarity	12.58%	0.672
	30	Tone	13.04%	0.687
	30	SDG Polarity	13.47%	0.725
	10	SDG Tone	16.37%	0.773
	10	SDG Polarity	13.51%	0.788
	10	Tone	16.57%	0.811
Long-Index				
	90	SDG Tone	10.02%	1.619
	50	SDG Tone	10.88%	1.622
	70	SDG Tone	10.20%	1.650
	90	Polarity	8.04%	1.671
	30	Polarity	12.09%	1.704
	70	SDG Polarity	8.50%	1.730
	70	Polarity	8.75%	1.759
	50	SDG Polarity	9.86%	1.768
	50	Polarity	10.21%	1.774
	30	SDG Polarity	12.98%	1.877
STOXX600			0.45%	0.030

6.2 Machine Learning Approaches

Table 5 shows the metrics for the top 10 performing strategies regardless if they are machine learning based or not. The strategies are divided by the type of strategies and sorted in ascending order based on the Sharpe Ratio in a similar fashion to the naive strategies section. The last row provided the performance metrics of the STOXX600 index during the same testing period as the context for comparison. Keep in mind that the testing period here is different and shorter than the whole period that we used to calculate the result for the naive approach in the previous section. As a result, the metrics for those naive strategies and the index can be significantly different compared to the last section. We can see that the annualized Sharpe ratio of the STOXX600 during this period is around 0.159 and its annualized return is up to 2.79%. This is quite higher than the whole period which is 0.03 and 0.45% respectively which means during this period from 2019 to 2022, the market is performing much better than in the previous period. Still, a Sharpe ratio of 0.159 is not that good and the index also suffered from a big drawdown at the beginning of 2020.

Figure 4 illustrates the cumulative return of the top 5 strategies for the Long-Only type regarding the STOXX600 dataset. Similar to the result from Section 6.1, almost all of the strategies, in this case, outperformed the performance of the STOXX600 during the given period. From the figure, we can see the line "Linear Regression_All" is above every other line which means the Linear Regression model that used both types of sentiment indicators provides the best performance in return. This is also the case with the Sharpe ratio as it is up to 0.617 higher than other strategies. The second best strategy also comes from the Linear regression model which is created only using ESG-related indicators. The performance of the other top 2 strategies is quite similar with the Sharpe ratio around 0.6. The other linear regression-based strategies that only use general sentiment indicators are nowhere to be found in the top 10 which infers the superiority of the ESG-based indicators in this case. The top two strategies which are both linear regression also show the effectiveness of machine learning-based strategies over the naive approaches in this kind of task. Finally, regarding the effect of adding more stocks into the portfolio, we again observe that for the Long-Only strategy, the fewer stocks, the better the return and Sharpe ratio of the portfolio.

Figure 4. Top 5 Long-Only Strategies Cumulative Return

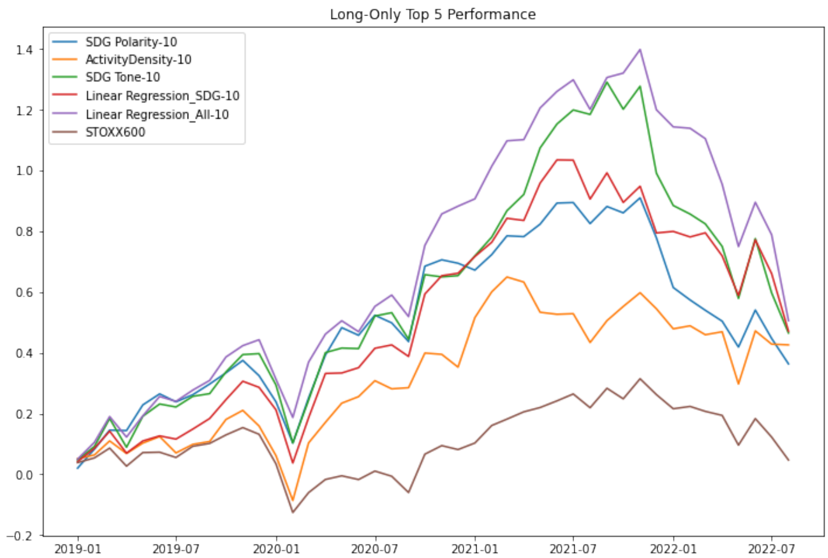
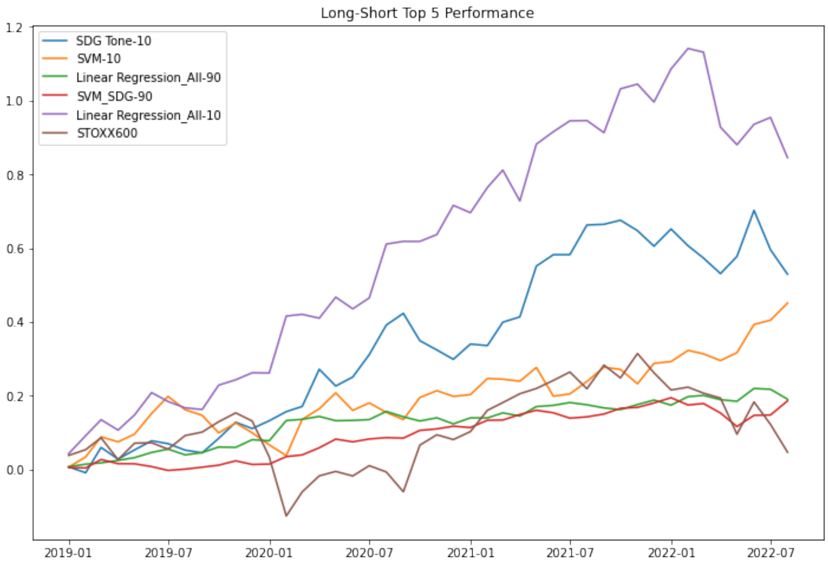


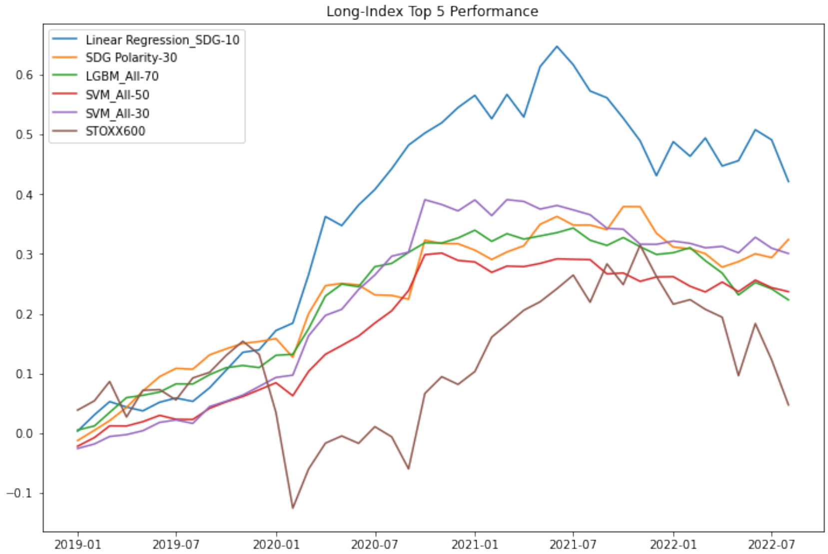
Figure 5. Top 5 Long-Short Strategies Cumulative Return



Next, Figure 5 shows the top 5 performing strategies regarding the Long-Short type. Once again, we have all of the strategies outperformed the index during this period. However, in this case, we can see even better performance from the machine learning algorithms. Among the top 10 performing strategies, 7 of them are created by machine learning models. Four out of five top models are also machine learning-based. As we are discussing the Long-Short strategy, the result shows that machine learning algorithms are effective at separating well-performing stocks and poorly-performing stocks. From Table 5, we can also see that the strategy based on linear regression using all variables with 10 stocks outperformed the others in both raw return and Sharpe ratio. Its Sharpe ratio is 1.276 and the annualized return is up to 19.28%. This strategy surpassed all other strategies (even compared to other types of strategies: Long-Index and Long-Only), especially in return, by a large margin. The second best strategy is based on the Support Vector Machine algorithm that only used the ESG-based indicators. This is better than the fourth strategy that used the same algorithm but was created with the general indicators which infers that ESG-based indicators performed better than the general one.

Finally, for the Long-Index type of strategies, we also notice a similar pattern compared to the Long-Short type. The machine learning models that used all variables continue to be the top-performing ones in both raw return and risk-adjusted return. The best model, however, is not the one using the linear regression algorithm but the SVM algorithms despite the fact that the Sharpe ratios of the two strategies are not that different (1.035 and 1.168 respectively). The linear regression model, in fact, even provides more annualized returns (11.42% compared to 7.64%). The two top strategies both come from the SVM algorithm and the third performing strategy used the LGBM algorithm. The top three strategies' performance is quite similar at the Sharpe ratio of 1.1 and most of the others have a Sharpe ratio of around 1. Among the top 10 performing strategies, 8 of them are created using machine learning algorithms which proves their effectiveness in improving the performance of the strategies. The result again shows that using all variables is better than using either general sentiment indicators or ESG-related indicators. And if you can only choose between the two types of indicator, the ESG-related sentiment indicators tend to provide better results.

Figure 6. Top 5 Long-Index Strategies Cumulative Return



To sum up, despite the fact that the testing period is shortened, we still have similar results from the naive strategies where Tone-based indicators perform the best for Long-Short and Long-Only strategies and Polarity-based indicators perform the best for Long-Index strategies. However, those naive strategies are greatly outperformed by the strategies created by machine learning algorithms. For the strategies created using machine learning, despite its simplicity, the linear regression algorithm is consistently the best model in 2 out of 3 times with the exception of the Long-Index strategy. The SVM algorithm performs well in both the Long-Short strategy and Long-Index strategy but not the Long-Only strategy. Regarding the quality of the indicators, the result indicates that using both types of indicators (general sentiment indicators and ESG-related indicators) offered better performance than using either of them. Between the two sets of sentiment indicators, in most cases, the ESG-related indicators are the better choice.

TABLE 5. Top Strategies Performance

Strategy	No of Stocks	Score Type	Annualized Return	Sharpe Ratio
Long-Short				
	90	LGBM_All	4.72%	0.900
	30	ActivityDensity	8.15%	0.921
	70	Linear Regression_All	5.54%	0.962
	10	Tone	13.97%	0.967
	30	LGBM	7.50%	0.967
	10	SDG Tone	13.16%	0.972
	10	SVM	11.34%	0.980
	90	Linear Regression_All	4.98%	1.061
	90	SVM_SDG	4.87%	1.068
	10	Linear Regression_All	19.28%	1.276
Long-Only				
	10	Tone	11.84%	0.481
	30	SDG Polarity	11.24%	0.492
	30	SVM_All	10.63%	0.498
	30	LGBM_All	10.06%	0.507
	10	SVM	13.09%	0.521
	10	SDG Polarity	10.87%	0.523
	10	ActivityDensity	12.63%	0.554
	10	SDG Tone	14.04%	0.558
	10	Linear Regression_SDG	13.59%	0.599
	10	Linear Regression_All	14.50%	0.617
Long-Index				
	10	SDG Tone	10.97%	0.964
	90	LGBM_All	4.88%	1.014
	50	LGBM_All	5.57%	1.018
	30	LGBM_All	7.08%	1.029
	10	Linear Regression_All	11.42%	1.035
	10	Linear Regression_SDG	10.53%	1.074
	30	SDG Polarity	8.23%	1.087
	70	LGBM_All	5.77%	1.110
	50	SVM_All	6.10%	1.152
	30	SVM_All	7.64%	1.168
STOXX600			2.79%	0.159

7 CONCLUSION

In this work, we analyzed the potential of using two newly created sets of sentiment indicators (one is general and the other is ESG-related) aggregated from news articles gathered from the GDELT database to predict the stock prices in the European stock markets. The results from our regression analysis suggest that there are some significant relationships between most of our sentiment indicators (Tone, Polarity, and Activity Density) and the stock return of the subsequent month, especially in the case of the sets of indicators that are built on ESG-related news articles. We built very simple trading experiments based on these significant sentiment indicators and our results have shown that it is possible to create trading portfolios that can considerably outperform the index using solely these indicators. Even though both types of indicators (general topics or ESG-related) provide similar statistical results, the ESG-related sets of indicators tend to perform better in our trading experiments. Among the sentiment indicators, Tone seems to be better at differentiating the performance between well-performing stocks and poorly-performing stocks as the Long-Short trading strategies based on them outperformed other variables. Polarity, on the other hand, performs better at selecting the stocks with the highest alpha as their Long-Index strategies surpassed other variables. We also use those sentiment indicators as inputs for a few popular machine-learning algorithms to see if we can create even better-performing trading strategies. As it turns out, the indicators are effective inputs for our machine-learning models. Most of the top-performing strategies are created using machine learning algorithms which proves their effectiveness in improving the performance of the strategies. The result also shows that using all variables is better than using either general sentiment indicators or ESG-related indicators. And if you can only choose between the two types of indicators, the ESG-related sentiment indicators tend to provide better results.

References

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring Economic Policy Uncertainty*. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Baker, S. R., Bloom, N., Davis, S. J., and Terry, S. J. (2020). COVID-Induced Economic Uncertainty. NBER Working Papers 26983, National Bureau of Economic Research, Inc.

Capelle-Blancard, G. and Petit, A. (2019). Every little helps? esg news and stock market reaction. *Journal of Business Ethics*, 157:543–565.

Chen, Y., Kumar, A., and Zhang, C. (2020). Dynamic esg preferences and asset prices. *Capital Markets: Market Efficiency eJournal*.

Fama, E. F. and French, K. R. (2004). The Capital Asset Pricing Model: Theory and Evidence. *Journal of Economic Perspectives*, 18(3):25–46.

Fraiberger, S. P., Lee, D., Puy, D., and Ranciere, R. (2021). Media sentiment and international asset prices. *Journal of International Economics*, 133:103526.

Fronzetti Colladon, A. and Elshendy, M. (2017). Big data analysis of economic news: Hints to forecast macroeconomic indicators. *International Journal of Engineering Business Management*, 9:1–12.

Fronzetti Colladon, A., Grassi, S., Ravazzolo, F., and Violante, F. (2020). Forecasting financial markets with semantic network analysis in the covid-19 crisis. *Journal of Forecasting*.

Fung, G., Yu, J., and Lam, W. (2002). News sensitive stock trend prediction. *Advances in Knowledge Discovery and Data Mining*, pages 481–493.

Hansen, K. B. and Borch, C. (2022). Alternative data and sentiment analysis: Prospecting non-standard data in machine learning-driven finance. *Big Data & Society*, 9(1):20539517211070701.

Kazemian, S., Zhao, S., and Penn, G. (2016). Evaluating sentiment analysis in the context of securities trading. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2094–2103.

Ke, Z. T., Kelly, B. T., and Xiu, D. (2019). Predicting Returns With Text Data. NBER Working Papers 26186, National Bureau of Economic Research, Inc.

Kvam, E., Molnár, P., Wankel, I., and Ødegaard, B. A. (2022). Do sustainable company stock prices increase with esg scrutiny? evidence using social media. *SSRN Electronic Journal*.

Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*, pages 1–367.

Long, C., Lucey, B., and Yarovaya, L. (2021). ”i just like the stock” versus ”fear and loathing on main street” : The role of reddit sentiment in the gamestop short squeeze. *SSRN Electronic Journal*.

Markowitz, H. (1952). Portfolio selection*. *The Journal of Finance*, 7(1):77–91.

Mudinas, A., Zhang, D., and Levene, M. (2019). Market trend prediction using sentiment analysis: Lessons learned and paths forward. *CoRR*, abs/1903.05440.

Nabi, R., Saeed, S., and Harron, H. (2020). A novel approach for stock price prediction using gradient boosting machine with feature engineering (gbm-wfe). *Kurdistan Journal of Applied Research*, 5:27–48.

Saz-Carranza, A., Maturana, P., and Quer, X. (2020). The empirical use of gdelt big data in academic research.

Schmidt, A. (2019). Sustainable news – a sentiment analysis of the effect of esg information on stock prices. *Political Economy - Development: Environment eJournal*.

Shiller, R. J. (2017). Narrative economics. *The American Economic Review*, 107(4):967–1004.

Tetlock, P. (2014). Information transmission in finance. *Annual Review of Financial Economics*, 6:365–384.

The GDELT Project (2022). The gdelt story. `https://www.gdeltproject.org/about.html`, Last accessed on 2022-06-02.

Tilly, S., Ebner, M., and Livan, G. (2021). Macroeconomic forecasting through news, emotions and narrative. *Expert Systems with Applications*, 175:114760.

Tuckett, D., Ormerod, P., Smith, R. E., and Nyman, R. (2014). Bringing social-psychological variables into economic modelling: Uncertainty, animal spirits and the recovery from the great recession. *Economic Growth eJournal*.

Vijh, M., Chandola, D., Tikkiwal, V. A., and Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia Computer Science*, 167:599–606. International Conference on Computational Intelligence and Data Science.

Wang, C. and Luo, B. (2021). Predicting $gme stock price movement using sentiment from reddit r/wallstreetbets. *Proceedings of the Third Workshop on Financial Technology and Natural Language Processing*, pages 22–30.

Wang, G., Wang, T., Wang, B., Sambasivan, D., Zhang, Z., Zheng, H., and Zhao, B. Y. (2014). Crowds on wall street: Extracting value from social investing platforms. *ArXiv*, abs/1406.1137.

Yang, J., Zhao, C., Yu, H., and Chen, H. (2020). Use gbdt to predict the stock market. *Procedia Computer Science*, 174:161–171.