Universiteit van Amsterdam
Amsterdam Business School

Hedging the risk against climate change using Twitter and mimicking portfolio approach

Alexandra Kutanina

Master's Thesis to obtain the degree in MSc Finance, track Quantitative Finance
University of Amsterdam
Amsterdam Business School

Author:	Alexandra Kutanina
Student number:	11641274
Email:	kutaninas@bk.ru
Date:	June 2021
Supervisor:	Mario Bersem

Statement of Originality

This document is written by Student *Alexandra Kutanina* who declares to take full responsibility for the contents of this document.

I declare that the text and the work presented in this document are original and that no sources other than those mentioned in the text and its references have been used in creating it.

The Faculty of Economics and Business is responsible solely for the supervision of completion of the work, not for the contents.

Abstract

This research proposes and implements a method for dynamically hedging climate change risk. The Climate News series is created by extracting the Twitter climate sentiment for the last decade on a monthly basis, as well a new factor selection strategy is proposed. Climate change hedging portfolios are then constructed using a mimicking portfolio technique. This is done using the Fama-French factors sorted portfolios, and climate risk exposure sorted portfolios. Later on, the results are compared with the previous studies, and it appears that the traditional way of hedging the climate risk with the industry bets performs better in hedging climate news developments both in and out of sample than the mimicking portfolios created based on the climate characteristics of the assets. In contrast, the new factor selection strategy outperforms both ways of hedging proposed before. The study outlines some potential research possibilities for climate risk management.

1. Introduction

"Clearly, global warming is getting a lot of attention today. And just as clearly, people disagree about whether it is real, whether it is important, and what it means for human societies. What should the non-specialist conclude about this debate? If climate change is real, how much does it matter? Where should our concerns about global warming rank among the other issues we face, such as growing inequality and nuclear proliferation?"
- William D. Nordhaus (Nobel Prize Lecture, 2018)

There have been many concerns in the past few years about climate change, but uncertainty about it is still present in many parts of the world. With all the measures introduced by different governments, it becomes clear that climate change would influence economics, the way companies work, and society in general. Investors are struggling with finding a way to insure themselves against the potential economic consequences of climate change, partly because it is hard to measure the threat of climate change.

Climate change was studied in scientific papers, articled, discussed on social media and in blogs, resulting in a wide variety of opinions and subjects. The common trend is to believe that it poses significant challenges to society and that it is a mother of all negative externalities, which are much bigger than other environmental problems (Anthoff, Tol & Helgeson, 2009). Climate change affects various sectors – agriculture, energy use, health, and others. As Kolk and Pinkse (2004) showed, the industries most affected by regulatory climate risk are oil and gas, mining, metals, utilities, while the physical damages factor is most salient for insurance and food and beverages and, to a lesser extent, finance, and securities. The economic effects are vast, so climate change risk management is crucial for financial institutions.

In recent years, much research has been done that suggests that financial institutions should consider the climate risk when making decisions because it has implications for their portfolios, and some regulatory risks have already started to materialize (Krueger, Sautner & Starks, 2018). For example, research shows that companies will face significant changes because of the Paris Agreement. It was shown by Ginglinger and Moreau (2019) that higher climate risk results in decreased firm leverage, as companies reduce their demand for debt and lenders reduce financing to businesses with the highest risk. Delis, de Greiff, and Ongena (2019) showed that banks have started to include the pricing of carbon risk in their loans, and Seltzer, Starks, and Zhu (2020) have found that polluting firms' credit rating and yield spreads were affected by the Paris Agreement. Sharfman and Fernando (2008) showed that strengthening environmental policies may decrease a company's cost of capital and enhance its value.

Common ways of mitigating climate risk are hedging it using futures or insurance contracts. Engle et al. (2020) have proposed that it is possible to hedge the climate risk using a more accessible to implement approach than using the standard ways of doing that. They use the sentiment derived from the news from the economic journals using textual analysis and publicly traded assets. They propose a method where instead of purchasing a security that directly pays off in the case of a future climate calamity, one would systematically (every month) construct portfolios from the publicly traded assets whose short-term returns hedge the climate change news. They first construct the time series of the extent to which climate news is discussed on social media and then construct a portfolio that overweighs stocks that rise in value when negative news about climate change is published. They believe that if such news concerning climate change materializes again, an investor's portfolio will be well-positioned to profit. If this portfolio is updated regularly based on new knowledge about the relationship between climate news and stock performance, it will eventually result in a portfolio that longs the winners and shorts the losers from climate change. This approach considers the fact that the risk of climate change is non-diversifiable and has a long-run nature, so in the event of a climate disaster, it will be more reliable than futures or insurance contracts since no counterparty gives a guarantee to pay and has a probability to fail its obligation. They urge future researchers to study the alternative sources of news hedging, and the one studied in this paper is social media.

Social media has been snowballing during the past ten years. In 2020, over 3.6 billion people were using social media globally, more than 50% of the Earth's population. *Social media* is web-based and mobile-based Internet applications that make it possible for users to create, access, and exchange user-generated content (Kaplan and Haenlein, 2010). According to the Efficient Market Hypothesis, all the information, including historical and even hidden information, is incorporated in the financial valuation, and investors are rational (Fama, 1970). However, during the past years, more and more economists started to question the Efficient Market Hypothesis because some behavioral and emotional elements are not incorporated in the financial valuation. Prechter and Parker (2005) argue that social mood is one of the factors that play a role in financial decision-making.

As a result, the social mood effect on the financial markets had started being studied by different researchers, mainly using surveys and computational analysis (Nofsinger, 2005; Parker & Prechter, 2005). Large-scale online data was profoundly studied, and many ways of computing indicators of the social mood and sentiment emerged. The computational analysis seems more efficient than surveys since it is more rapid, cost-effective, and accurate (Bollen & Mao, 2011). Moreover, it is possible to predict various phenomena using the computational analysis of the social mood, such as the outcome of presidential elections (Yaqub et al., 2017), commercial sales (Dijkman et al., 2015), epidemics (Ji et al., 2012) and financial markets movements (Tafti et al., 2016; Ranco et al., 2015).

In this research, Twitter is used to measure social mood. Twitter is an American microblogging and social networking service where users can communicate by publishing "tweets", messages that contain a maximum of 280 characters (About Twitter | Our company and priorities, 2021). Besides that, users can publicly share tweets created by other users with their followers, known as a Retweet. It is an effective way to pass the news on Twitter. There is also a reply function, which is a response to another person's tweet. It increases the tweet's visibility since users following the person who replied will see the tweet in their feed. Like function is used to express appreciation for a Tweet. By clicking the Likes button on the user's profile page, it can be seen which Tweets a user liked. Likes increase the tweet's visibility since users see the posts people they follow liked in their feed.

Various researchers used Twitter to describe social and linguistic patterns (Dahal, Kumar & Li, 2019). Dodds et al. (2011) have created the "hedonometer," a tool for quantifying the happiness of each tweet (both positive and negative emotion). Moreover, trending topics include the headlines and news (Kwak et al., 2010), so Twitter is a perfect social media platform to capture the overall sentiment of the news.

Therefore, the central question of this research is: *Can the climate risk measured by Twitter Sentiment and Wall Street Journal Index, proposed by Engle et al. (2020), be hedged using the mimicking portfolio approach?* This study attempts to develop the climate risk hedging portfolios using mimicking portfolio approach and Twitter climate sentiment, whose short-term returns will hedge the climate news on Twitter, which will lead to the long-run exposure to the climate risk. It would make it possible to construct a regression for the investor, for example, in May 2019, where climate extracts the weights for each stock from it, where weight is determined by the company's climate characteristics, book-to-market ratio, market excess return, and size, and construct a well-balanced portfolio with these weights. This portfolio would be used as a hedge for climate news in June 2019.

The hedging ability of two alternative approaches to portfolio construction is also tested in this study. One approach uses ETFs, as well as book-to-market ratio, market excess return, and size. Another approach includes the new factor selection strategy proposed in this research, where factors are stocks that are correlated the most with the climate news on Twitter.

The first step to answer the research question is to convert the Twitter data to a time series to be a reliable hedge target for the climate risk. It is assumed that once information about climate change appears, it is captured on Twitter. Tweets from the top 10 news accounts are taken as a sample. There is a need to construct a measure that captures the extent of the climate change being discussed in the news is calculated as a product from the occurrence of climate news and their sentimental value (positive or negative).

In this research, two different types of climate risk are considered being physical damages from climate change and regulatory risks, associated with the climate change. In this research, news related to both would be analyzed, and no distinguishment between the two is made.

The next step is to implement a mimicking portfolio approach. The goal is to construct portfolios from stocks that are positively correlated with good climate news or negatively correlated with lousy climate news and short sell the stocks that have the opposite correlation with the climate change news. There are different approaches to determine the weights of each stock in the mimicking portfolio, but in this paper, two are used, one being new to the field (coefficient elimination factor selection) and the other one replicated from the study performed by Engle et al. (2020). The primary technique for both is to regress the climate change index on the universe of assets and some other factors with deviations based on the approach. The results of these regressions are coefficients for each stock/factor, and after rebalancing these coefficients, weights for each stock are obtained, and a portfolio is formed. Updating such a portfolio every month would lead to a dynamic hedging strategy with long-run exposure to climate risk. It is done for both indices created by Engle et al. (2020) and the Twitter Climate Change Index developed in this research. Then, the constructed hedge portfolios are compared to alternative hedging portfolios that consist of simple industry bets such as two energy-related ETFs, XLE and PBD, and the Fama-French factors: size, value and market excess return.

The results of this research indicate that the standard approach, where two ETFs, XLE and PBD, are added to the Fama-French Factors, outperform the ESG-characteristic-based mimicking portfolios and hedge around 25.6% of the climate change news on Twitter, whereas the mimicking portfolio approach hedges only 2.2% of the climate news on Twitter. Whereas for the Wall Street Journal Index, mimicking portfolio approach, where the universe of assets is combined with the Fama-French factors, hedges around 8.5% of the climate news, and two ETFs with the same factors hedge 12.6% of the climate news. It is also found that the new factor selection strategy outperforms the market on the out-of-sample period, and the constructed portfolio achieves the out-of-sample correlation of 88.3% with the Twitter climate news. In contrast, the correlation for the strategy involving ETFs is 80.6%, and firm-specific climate characteristics is 54.9%.

This study opens the avenue for future research. First, firm-level climate risk exposures could be studied deeper, and another approach or, perhaps, different data could be used for that. The glossary of climate change could be improved as well since the climate change vocabulary increases with a rapid speed. Future researchers could account for the distinguishment between different types of climate risk and select the securities accordingly. Another suggestion would be to research different hedging strategies as well as factor-selection strategies.

This research contributes to a proliferating literature in the area of the effect of climate change on financial markets and of the effect of financial markets, in turn, on the dynamics of climate change. It also contributes to the existing research on the effect of social media on financial markets. It builds upon the approach of hedging the Wall Street Journal climate news by Engle et al. (2020) and does it by proposing a new climate change news source – Twitter. Both sources for the climate change news are compared in the study. This research also proposes a new factor selection strategy for the mimicking portfolio approach.

2. Related literature

2.1. Climate change risk

There is sufficient empirical and theoretical evidence in the existing literature that suggests that institutional investors should consider the risks caused by climate change when making decisions (Krueger et al., 2020). Bolton and Kacperczyk have provided evidence of the significance of the carbon risks and the pollution of the environment risks in the cross-section of stock returns (2019). Hong, Li, and Xu have proved that climate risks are currently being mispriced in the financial markets (2019). Despite the apparent importance of considering the climate risks while making decisions, institutional investors find it hard to price and hedge the climate risks since most of the instruments available nowadays are not suitable for hedging.

Shive and Forster (2019) have studied the effect of the ownership structure of the U.S. companies on greenhouse gas emissions since the number of public firms in the U.S. has halved during the past 20 years. They have found that public firms and private sponsor-backed firms are more likely to pollute than independent private firms. Also, according to their research, the number of emissions has a negative correlation with mutual fund ownership and board size that suggests that the greenhouse emissions could be decreased with the increased oversight.

Krueger, Sautner, and Starks (2020) study the investors' views and approaches to managing climate risks. They found that the institutional investors believe that climate risks influence their portfolio firms. Besides that, they believe that the regulatory climate risks already started impacting their portfolio firms. A significant part of investors has taken the first steps toward managing climate risks. Most long-term investors and more significant, and ESG-oriented, prefer risk management and engagement to divestment for addressing climate risks.

Addoum, Ng, and Otiz-Bobea studied the effect of climate change on the U.S. corporate sector performance (2019). There is much debate in this area because the U.S. has lately decided to withdraw from the Paris Climate agreement since it would damage the American economy by controlling its carbon emissions. No evidence was found that temperature exposures have a significant effect on establishment-level sales or productivity.

Barnett, Brock, and Hansen (2019) found that belief heterogeneity in the long-run climate change risks significantly affects the real estate prices in the U.S. They have found that houses in the neighborhoods, where most of the population believes in the long-run climate change risk, are on average 7% cheaper than those where the majority does not believe in it, all else being equal. It means that the real estate sector is significantly affected by the climate change risk decades before the actual damages are expected to occur. On the other hand, Murfin and Spiegel (2019) studied whether the threat for the sea level rise is capitalized in the recent transaction prices for residential real estate and whether it is consistent with its price. Surprisingly, they have found no detectable price effects in their research across various specifications and test settings.

Choi, Gao, and Jiang (2019) have discovered that people tend to change their beliefs about climate change to positive when being in an area warmer than the average temperature. Google searches increase as well as stocks of the carbon-intensive companies abnormally underperform the low-carbon ones. Global warming, on the other hand, is a long-term tendency that is not apparent on an individual level. Most people confuse climate with the local weather.

According to Alok, Kumar, and Wermers (2019), the distance between the professional money manager and a place where climate calamity occurred influences the weight of the stocks in the area of the climate disaster professional money managers assign in their portfolios. Investors, which are located closer to the place where the climate event took place, underweight these stocks more than those located further. It is associated with the salience bias. This overreaction can be expensive. Therefore, it is essential to hedge the risk ahead of time.

Andersson, Bolton, and Samama (2016) have presented the investment strategy, which hedges the climate risk for institutional investors that are looking for a long-term passive investment. They have pointed out that most asset managers are unwilling to divest from stocks with high carbon footprints because there is a risk of underperforming one's benchmark since governmental regulations are being postponed and the market does not expect them to be implemented anytime soon. Therefore, they have developed an investment strategy that uses decarbonized indices, but instead of investing in only pure-play indices, it keeps the similar aggregate risk exposure as standard market benchmarks and minimizes the tracking error. The researchers suggest that this strategy helps the investors to hold the free option on carbon since the returns are similar to the benchmark when governmental climate regulations are postponed, and it outperforms the benchmark once these regulations are introduced.

Engle et al. (2020) proposed a revolutionary approach to hedge for climate risk. Instead of hedging it in the traditional way with the help of insurance or futures, which they believe is difficult because climate risk is non-diversifiable and will appear only in the long run, they suggest creating portfolios which short-term returns hedge climate change news throughout the holding period. The investor can hedge the long-run exposure to climate risk by hedging the climate change news every time it occurs. It was accomplished through the development of two complimentary indexes that assess the extent to which climate change is mentioned in the news media. The problem with implementing the mimicking portfolios approach based on these indices is that there is limited data for the climate news; however, there are many assets available. To account for this issue, they use the proxy to account for the firm-level climate characteristics – environmental performance scores from two different data providers. After that, they build the final hedge portfolios by projecting innovations of the climate news measure on the environmental performance scores sorted portfolios and Fama-French factor-sorted portfolios (Excess Market Return, Size, and Value). Later they compare the hedging performance of the universe of assets sorted by climate characteristics with the traditional hedging method with the industry bets (ETFs). They find that the climate characteristic sorted mimicking portfolios perform somewhat better than the traditional approach with industry bets. They find that the return of the hedging portfolio based on the climate characteristics of each firm achieves the out-of-sample correlation of 30% with the WSJ index.

2.2. Twitter sentiment

Griffin and Tversky (1992) suggested that individuals tend to focus a lot on the strength of the available information instead of its statistical weight or credence. In the case of the financial markets, investors overreact to the emotional importance of the news instead of the actual statistical weight of the event described in the news. It leads to misinterpretation of the news, which is emotionally significant but does not affect financial markets. Much literature is available in this field, but the conclusion is that sentiment affects the financial markets in the case of limited arbitrage (Nofer & Hinz, 2015).

Bollen et al. (2011) have shown the connection between Twitter sentiment and the financial markets. Around 10 million tweets were used for the analysis, which was posted in February-December 2008. They found that Twitter sentiment matches the shifts in DIJA values three to four days after the tweets were published. Miital and Goel (2012) have built upon this research and used a more significant sample and advanced machine learning techniques and got similar results.

Social media is now being used as a source of public opinion for various topics. Dahal, Kumar & Li (2019) focus on the climate change topic. They analyze tweets that contain geotags, climate change keywords using volume analysis, topic modeling, and sentiment analysis. They found that the overall discussion on Twitter about climate change is negative, and it is highly negative when individuals react to political or extreme weather events. They also found that climate change tweets in the U.S. contain fewer policy-related topics than others.

Cody et al. (2015) argued that climate change is extensively discussed in science, newspapers, and social media. According to them, scientific papers are too opaque for the public, and newspapers lack accuracy, while on social media, people with diverse backgrounds can share their points of view. They argue that Twitter is a valuable resource for analyzing current events, news, and climate change awareness since more climate change activists are than deniers there. They analyzed the sentiment in tweets that included the word "climate" in reaction to climate catastrophes or any other events related to climate change. The increasing percentage of population is increasingly likely to utilize social media as a source of information, and Twitter conversation is growing more prevalent. They discovered that tweets including the term "climate" are less joyful than all other tweets. They have also found that disasters, climate bills, and oil drilling might lead to a decrease in happiness, climate rallies, a book release, a green ideas contest might lead to an increase in happiness. They conclude that Twitter might be a valuable asset in the ongoing fight against climate change, as well as a valuable research source for social scientists, an uninvited public opinion tool for policymakers, and a conduit for scientists to connect with the public. At the same time, the individuals may use social media to get the news concerning current events and express their views on various problems, including climate change.

Ripberger et al. (2014) tested whether Twitter activity represents the severe weather fluctuations – tornados. They have compared the fluctuations in Twitter activity to the issuance of tornado watches and warnings. They found a high degree of convergent validity, which means that Twitter could be used to understand better the relationship between risk communication, attention, and public reactions to severe weather.

Dewally (2003) have investigated the effect of investment advice on the Internet, namely online discussion groups, on the return characteristics of the recommended stocks. It was found that the recommendations are primarily positive where to buy advice occurs seven times more often than they sell advice. Besides that, most investors follow a momentum strategy, and the stock market does not react to these recommendations.

2.3. Hedging the sentiment

In recent years new platforms occurred, one of which is a microblogging tool StockTwits (StockTwits, 2021). The primary goal of this tool is to allow users to exchange trading advice in brief bursts in real-time. Another similar tool is HedgeChatter (HedgeChatter, 2021), a social media stock analysis platform. It analyzes the conversations of over 7600 U.S. equities and allows investors to determine who is chatting the most about stocks, which are the most targeted, and shows the users who are correctly predicting the price movements (Rao & Srivastava, 2014).

Rao and Srivastava (2014) shave studied the relationship between tweet board literature and financial instruments. They argued that High-Frequency trading companies have recently developed a wide variety of AI bots, capturing the buzzing trends on Twitter. These bots do not go into the meaning of the sentiments and are often not that accurate since they have to help the traders to place the bets faster than other market players. Therefore, their predictions can be pretty misleading: when Anne Hathaway won Oscar, the stock prices of Berkshire Hathaway rose by 2.94%. Therefore, the researchers believe that a high volume of tweets can create short-term effects on stock prices. They have proposed a new tweet hedging strategy and have found a high correlation between Twitter sentiment and stock prices, validated by the Expert Model Mining System with the R-square value of 0.952. After that, they have introduced the market monitoring elements derived from the public moods, which are then used to create a portfolio within a limited risk state during typical market conditions. They use the married put strategy, which depends a lot on whether the market will rise or fall, which is, in turn, predicted using Twitter Sentiment.

Bouri et al. (2017) study whether Bitcoin can be used as a hedge for stock indices, bonds, oil, gold, the general commodity index, and the U.S. dollar index by creating a well-diversified portfolio and employing eight different portfolios optimization scenarios and Monte Carlo simulation. It is found that Bitcoin can be used as a diversifying instrument because the performance measures in almost all portfolio optimization scenarios were better with Bitcoin than without it. The results indicate that Bitcoin is suitable only for diversification purposes, but it can be used as a haven against the extreme down movements in Asian stocks.

Zheng, Osmer, and Zhang (2017) have proposed a sentiment exposure model incorporating the time-series average of past market sentiment. They study how hedge fund managers moved in the market when sentiment changed and the economic value created from hedging the sentiment. They do it using the Baker and Wurgler Sentiment Index and bootstrap analysis. They found that fund managers adjust the market exposure of their portfolios to changes in market sentiment and that the hedge fund managers with the highest negative sentiment exposure outperform the ones with positive by 1.7-2.4% per year.

3. Hypotheses

The main question of this research is the following: *Can the climate risk measured by Twitter Sentiment and Wall Street Journal Index, proposed by Engle et al. (2020), be hedged using the mimicking portfolio approach*? It is divided into three hypotheses:

Hypothesis 1: Mimicking portfolios constructed from the same universe of assets sorted by their ESG-Scores hedge more variation in Twitter Climate Change Index than in the existing Wall Street Journal Climate Change Index.

Hypothesis 2: Mimicking portfolios are a better hedge against climate risk than the existing hedging instruments (XLE and PBD).

Hypothesis 3: When climate change is discussed more on Twitter, firms with higher ESG-scores have lower returns than those with lower ESG-scores.

4. Methodology

4.1. Data

4.1.1. Hedge targets

Engle et al. (2020) made the climate news measure time series publicly available in order to encourage other scholars to investigate alternative approaches to build climate hedges. They have constructed two new sentiment indices, which in this research were combined with the Twitter sentiment index. The first index is constructed as the monthly correlation between 'The Wall Street Journal' text and a fixed climate change vocabulary created from a list of authoritative texts published by various climate organizations. Since this index was used in this research for comparison with the new index proposed, the graph below represents the WSJ Climate News Index from 2010 to 2018.

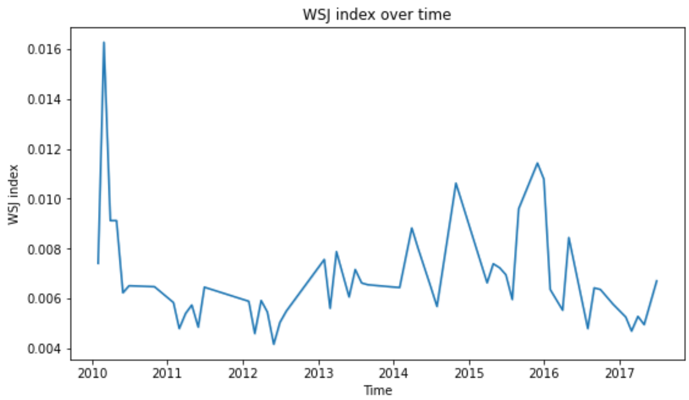


Figure 1: WSJ index over time

It can be observed that climate news coverage started steadily growing in the 2000s. Engle et al. (2020) propose that salient climate events cause the spikes.

The second index is constructed in the same manner, but instead of taking all the news, only the ones with the negative sentiment are selected.

4.1.2. Constructing a third hedge target

In this paper, a new index is created. This index calculates the correlation between tweets in a given month and a set of fixed climate change vocabulary. It uses the historical data collected from Twitter, using a text retrieval algorithm created using mainly pandas, scikit-learn, and NLTK libraries in Python. Twitter is an American microblogging and social networking platform where users can interact with each other using "tweets", limited to 280 characters. Twitter's mission is to give everyone the opportunity to share their opinion without barriers. Many people also use it to read the news and subscribe to the accounts they find interesting. Therefore, an index constructed with Twitter was used as a hedge target for this research. To create it, the steps depicted in Figure 2 are taken and described later in this section (Figure 2).

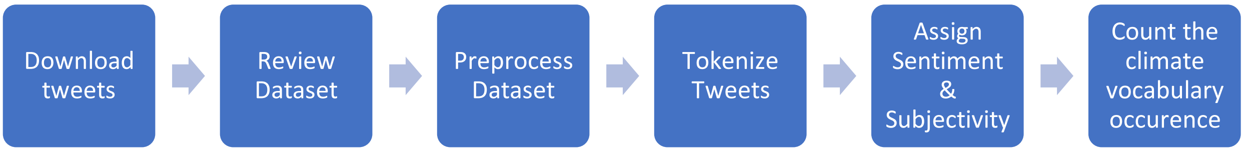


Figure 2 – Twitter index creation steps

4.1.2.1. Downloading the Tweets

In this research, the historical tweets from the top 10 followed Twitter news accounts are collected. It was done using the advanced Twitter scraping & OSINT tool written in Python, Twint (twintproject/twint, 2021), similarly to Xavier and Souza (2018). The advantages of this tool are that it can fetch more tweets than Twitter allows with its API (more than 3200 tweets), it is anonymous, does not have rate limitations, and has a fast initial setup.

4.1.2.2. Reviewing the dataset

In total, 1195201 (1.2M) English language tweets were retrieved. They were collected over the period from January 2007 to April 2021. Each tweet record contains tweet identifier, conversation identifier, date/time of submission (in GMT), user identifier, name, place, language, and text. Mentions, URLs, photos, videos, replies count, retweets, and likes count are also collected and used later in the analysis. An example of how tweets from one of the accounts (@CNN) were downloaded using Twint in Python and a part of a resulting dataset from this query can be seen in the Appendix.

The data from Twitter was collected using the Twint library. Accounts selected for the analysis are the most followed ten news accounts on Twitter: @cnnbrk, @CNN, @nytimes, @BBCbreaking, @SportsCenter, @espn @BBCWorld, @TheEconomist, @NatGeo, @Reuters (SocialBlade, 2021). In total, 1195201 tweets were collected. The graph below depicts the tweet count over time (Figure 3).

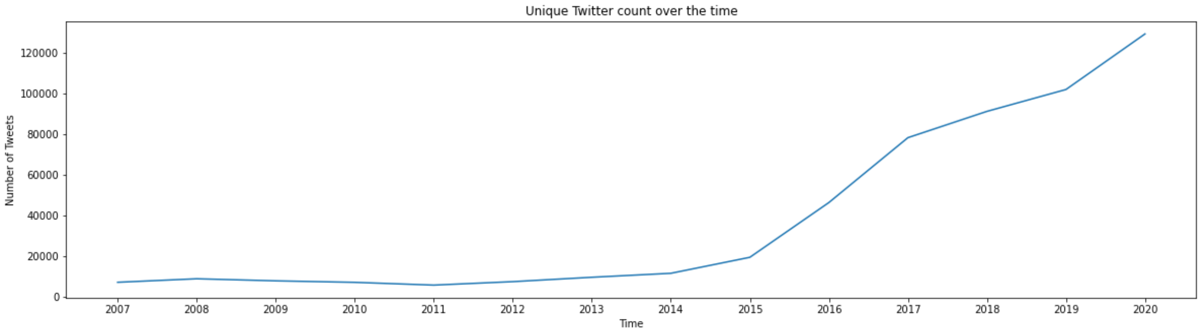


Figure 3: Tweets count over time

Social media usage increased significantly during the past 15 years, so the number of tweets posted on a yearly basis increased accordingly. The number of tweets in 2020 almost reached 160000.

4.1.2.3. Preprocessing the dataset

Twitter sentiment analysis is then done. Its goal is to translate text into sentiment since text with positive sentiment could have a different effect on the financial market than the one with the negative or neutral sentiment. Sentiment analysis lies within Natural Language Processing which extracts, quantifies, and analyzes personal information from the text (Wilson et al., 2005). Machine learning techniques and dictionary-based techniques are the two methodologies used for sentiment classification (Kearney and Liu, 2014). Previous studies indicate no substantial difference in results, particularly when it comes to social media (Hutto and Gilbert, 2014). Tweets in this research combine the data from the climate sector and news analysis, which is a relatively new area, so there is no train data available to train the algorithm. As a result, the dictionary-based method is employed in this study.

The first step after getting the data and reviewing it in the sentiment analysis is data preprocessing. The dataset has many data not needed for the analysis, so data preprocessing is done to clean the text from links, images, hashtags, numbers, and other unnecessary words. It was done in Python using various libraries, namely nltk, bs4, and pandas. The English stopwords from the nltk library were also removed since these are common words that do not add any value to the meaning of the tweet. It can be observed in the figure below how a sample of tweets is changed after the preprocessing step (Figure 4).

	tweet	clean_new
0	"I'm happy the world is finally getting to see [Devin Booker]." @CP3 with high praise as @DevinBook clinches his first postseason 🤝 https://t.co/ELfJLpx2N	happy world finally getting devin booker high praise clinches first postseason
1	After beating the Clippers, the Suns secured a spot in the playoffs for the first time in 11 years 🌞 https://t.co/sH1XtFIBTS	beating clippers suns secured spot playoffs first time years
2	Russ now has 30+ triple-doubles in three different seasons 🤯 Only two players in NBA history have ever had one such season: Oscar Robertson and Wilt Chamberlain. https://t.co/yKy1eYfMXw	russ triple doubles three different seasons players history ever season oscar robertson wilt chamberlain
3	AD interrupted Russ' postgame interview 😩 https://t.co/AB5BhJN3aP	interrupted russ postgame interview
4	The Wizards take down the Lakers 😤 Beal: 27 Pts, 11-18 FG Russ: 18 Pts, 18 Reb, 14 Ast https://t.co/fDe8StiHiy	wizards take lakers beal russ

Figure 4 – Tweet transformation after preprocessing step

Unneeded items, such as links, short words, hashtags, etc., are removed, and only the most essential words for the analysis remain.

<u>4.1.2.4. Tokenizing the dataset</u>

Then tweets were tokenized since some of them still resemble sentences. Tokenizing the text is crucial in sentiment analysis since it breaks out sentences into smaller tokens – words in this case. These words are later used to prepare the vocabulary, which makes it possible to see the occurrence of each word and its added value to the text. The figure below shows how the first five tweets look like after tokenization (Figure 5).

```
0                                [happy, world, finally, getting, devin, booker, high
, praise, clinches, first, postseason]
1                                                       [beating, clippers, suns, secure
d, spot, playoffs, first, time, years]
2    [russ, triple, doubles, three, different, seasons, players, history, ever, season
, oscar, robertson, wilt, chamberlain]
3                                                                                     [i
nterrupted, russ, postgame, interview]
4
[wizards, take, lakers, beal, russ]
Name: clean_new, dtype: object
```

Figure 5- Tweets after Tokenization

After this step, the initial tweets are preprocessed and tokenized and ready for the final analysis.

4.1.2.5. Assigning Polarity and Subjectivity

After that, using the TextBlob Python library based on the pattern library, tweets were assigned with polarity and subjectivity (TextBlob: Simplified Text Processing — TextBlob 0.16.0 documentation, 2021). TextBlob is used in natural language processing for sentiment analysis. Weighted polarity and subjectivity were also calculated. The figure below shows the sentiment assigned to different texts using this library (Figure 6).

tweet	clean	clean_new	Subjectivity	Polarity	Analysis
I am very happy!	i am very happy	happy	1.0	1.0	Positive
this is absolutely terrible	this is absolutely terrible	absolutely terrible	1.0	-1.0	Negative
neutral mood	neutral mood	neutral mood	0.0	0.0	Neutral
climate change is happening	climate change is happening	climate change happening	0.0	0.0	Neutral
climate change can have a negative effect on us	climate change can have a negative effect on us	climate change negative effect	0.4	-0.3	Negative

Figure 6- Sentiment analysis on fake data

Tweets were categorized into three groups: negative (-1), neutral (0), and positive (+1). Then a word cloud with the most common words from the historical news on Twitter was created. The word climate appeared 9013 times. The word cloud with the most common words which appeared in the sample is depicted below (Figure 7).

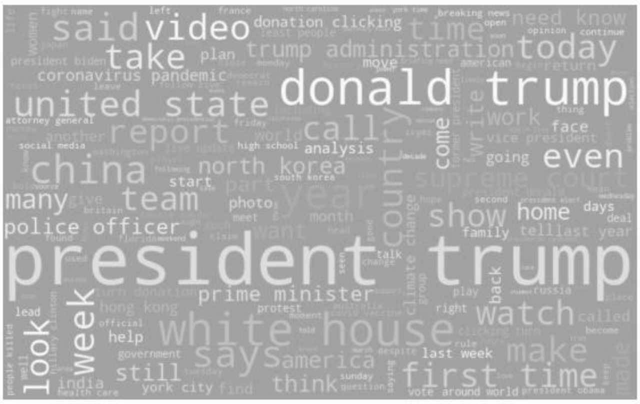
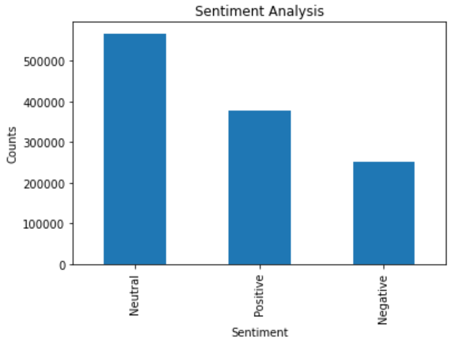


Figure 7: Word cloud

The bar plot with the value counts is depicted below (Figure 8). The histogram below shows the polarity of the tweets with 31 bins (Figure 9).



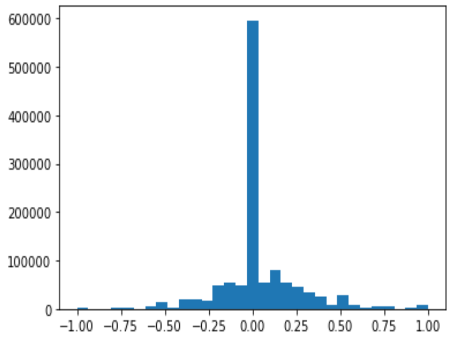


Figure 8: Distribution of sentiment *Figure 9: Histogram of the sentiment*

It can be observed that most tweets are neutral, and there are more positive tweets than negative ones in the given sample. The polarity of the tweets tends to be higher for the positive tweets than for the negative ones.

4.1.2.6. Climate change vocabulary

The glossary with the Climate Change Vocabulary from the Minnesota Department of Health was used to complete the sentiment analysis. Words from the document were parsed into Python using the pdfreader library ("pdfreader", 2021). After that, the number of occurrences of the words from the climate change vocabulary was assigned to each tweet. Each tweet contained between 0 and 7 words from the climate change vocabulary. It appeared that words from the climate change vocabulary had appeared 83838 times in the sample of tweets over time. The graph below shows the occurrence of the words from the climate change vocabulary over time (Figure 10).

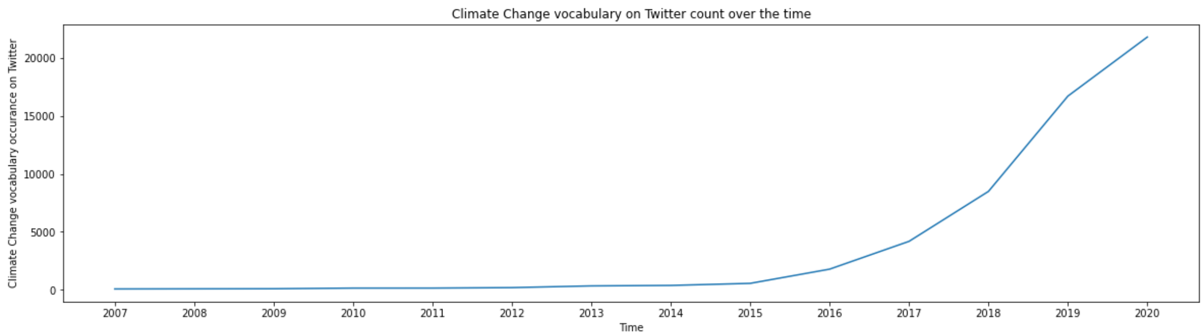


Figure 10: Climate change vocabulary occurrence over time

It can be observed that the occurrence of climate change vocabulary has significantly increased over the sample period of time. The count was relatively stable from 2007 to 2015. However, after 2015 it started growing much faster. It can be explained by the fact that the Paris Agreement was adopted in December 2015, and, as described in the literature review, influenced the financial sector a lot.

4.1.3. Universe of assets and risk-free rates

In order to construct the hedging portfolios, there is a need to define the universe of assets. The monthly individual U.S. stock returns data was downloaded from the CRSP/Compustat Merged Database for this research. The data was collected for the period from January 2010 to April 2021 for 32269 companies. Portfolios were constructed from the U.S. stock returns data from CRSP/Compustat Merged Database. It is crucial to understand that the ticker count over the years was not constant and fluctuated a lot, and it might affect the climate risk exposure measures discussed later.

After collecting the returns, risk-free rates should be estimated. Risk-free rates were downloaded from Yahoo!Finance as a monthly return on the Treasury Yield 10 Years from January 2010 to January 2021.

4.1.4. Climate risk exposures

After choosing the universe of assets for the research, the next step was to measure the climate risk exposure for each firm. Similarly to Engle et al. (2020), the ESG scores created by the prominent ESG data provider, Sustainalytics, were used as such measure. Sustainalytics provides investors with independent ESG and corporate governance research, ratings, and analysis. It aims to encourage investors to develop and implement responsible strategies. The 'Total ESG Score' was downloaded from Sustainalytics for the whole universe of the U.S. securities from January 2010 to September 2018 and was used as a measure for each firm's climate risk exposure. 'Total ESG Score' is calculated using the average of Environment Score, Social Score, and Governance Score. To evaluate each of these scores, Sustinalytics compares each firm to the firms in a similar industry. Weights are assigned to each industry separately based on the exposure of the industry to each ESG risk. Eventually, each firm is assigned a score between 0 and 100, where the higher score means that a firm is more environmentally friendly.

Total ESG Scores, taken from Sustainalytics over the period from January 2010 to September 2018, are plotted below (Figure 11).

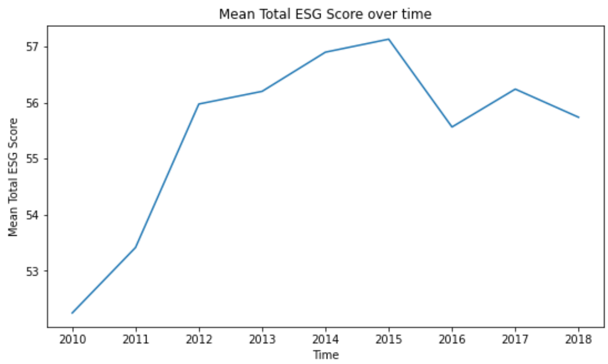


Figure 11: Mean E-Score over time

There is a positive trend in the mean ESG-score each year. However, it might be caused by different firms being added/excluded from the sample. That is why the total number of the firms provided with the ESG score by Sustainalytics each year can be observed below (Figure 12).

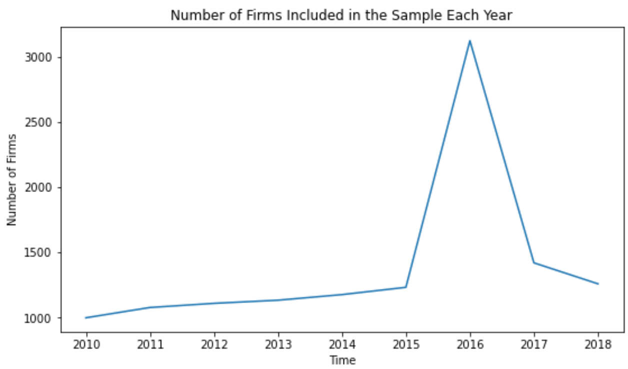


Figure 12: Number of firms included in the sample over time

Based on these two graphs, it can be observed that Sustainalytics was providing the data for roughly above 1000 firms till 2015 with a spike to 3000 in 2016. However, in 2017 the number of firms has declined to 1500 and stayed on the same level in 2018. That might have been the reason for a drop in the mean ESG-score in 2016. Engle suggests another reason for that. Al

(2020) might be that Sustainalytics has been changing its methodology throughout time. However, there is no proof for that.

To escape any complications associated with Sustainalytics' methodology or data not being available for all firms continuously, the 'Total ESG Scores' were ranked for each company monthly, then their ranks were demeaned and rescaled with the range from -1 to 1. This approach prevents cross-sectional dispersion and means changes caused by changes in methodology or constructing the score. However, even the chosen approach could not be the most accurate since a firm can change its ranking because new firms are added/ removed from the sample, but not because its environmental impact has changed (Engle et al., 2020). It can be observed on the graph below that even after demeaning and rescaling, the ESG score showed a drop in 2016 (Figure 13).

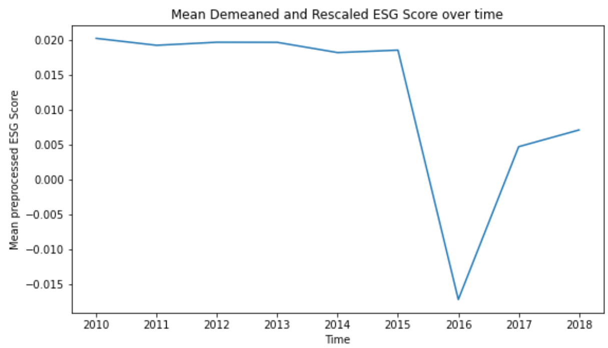


Figure 13: Preprocessed E-Score over time

4.1.5. Indexes

To compare the performance of the Twitter Sentiment as a hedging instrument with the news sentiment used by Engle et al. (2020) and the current climate ETFs (XLE and PBD), the WSJ climate index was downloaded as well as the historical monthly prices from Yahoo!Finance for both ETFs from January 2010 to January 2021.

The data for the ETFs, which was obtained from Yahoo!Finance contained the historical monthly adjusted closing price, open, high low, and volume. Monthly returns were calculated and plotted against time (Figure 14; Figure 15). Two ETFs that were observed are XLE and PBD. XLE is the Energy Select Sector Index representing the energy sector of the S&P 500 Index (ssga.com, 2021). PBD is the Invesco Global Clean Energy ETF based on the WilderHill

New Energy Global Innovation Index. It consists of the companies engaged in the advancement of clean energy and conservation (invesco.com, 2021).

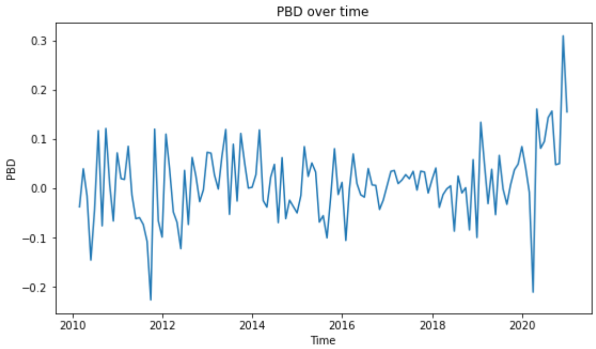


Figure 14: PBD performance over time

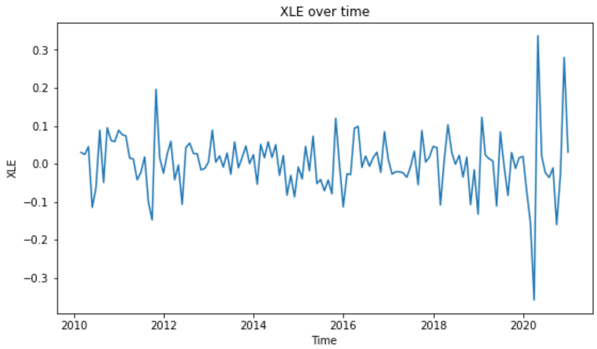


Figure 15: XLE performance over time

It can be observed from these two graphs that the returns have fluctuated a lot during the sample period. However, there is a generally positive trend in both.

4.1.6. Getting the SMB and HML

Additionally, to perform the mimicking portfolio approach, SMB and HML factors have to be estimated. The data for them was downloaded from the Refinitiv database from January 2007 to April 2021.

4.1.7. Combining it all

The last step in the data preparation is to combine all the data that have been gathered and processed. Certain choices were made while doing that, which is explained next. The universe of assets taken from WDRS was collected for the period from 2010 to 2021. However, Sustinalytics ESG-scores are not provided for the period from 2018 to 2020, so this data frame was excluded from the dataset, resulting in the data frame from January 2010 to September 2018. The next step was to combine it with the derived Twitter sentiment, climate vocabulary occurrence monthly. Lastly, the historical monthly prices for the ETFs were added to the dataset and Engle's WSJ climate index and the risk-free monthly rates.

4.2. Theoretical Model of hedging proposed by Engle

The objective of this paper is to construct hedging portfolios for the climate change risk using the news, CN_t. The excess returns of the portfolios were defined by Engle et al. (2020):

$$r_t = (\beta_{cn}\gamma_{cn} + \beta_{cn}(CN_t - E[CN_t])) + (\beta\gamma + \beta v_{t)} + u_t,$$

Where β_{cn} is a risk exposure to the news factor; β is a risk exposure to the other non-tradable risk factors v_t; γ_{cn} is a risk premium for the news factor; γ is a risk premium for the other non-tradable risk factors and u_t is the idiosyncratic error term.

The stated setup assumes that the risk exposures of the assets employed in the estimation remain constant throughout time. As a result, the portfolios have to be built in such a way that their exposures to the underlying risk variables remain consistent. A common method for accomplishing this is to create portfolios by categorizing assets based on their characteristics, and it will be used in this research. Therefore, we define the portfolio returns with firm-level characteristics, which are cross-sectionally normalized, Z_t, as:

$$\tilde{r}_t = Z'_{t-1} r_t$$

Next, the hedge portfolios should be constructed. Hedging portfolios are the portfolios that have unit exposure β_{CN} to climate change risk but no exposure to the other non-tradable risk factors, v_t. This is done using the mimicking portfolio approach (Giglio and Xiu, 2018; Engle et al., 2020).

In this paper, following the mimicking portfolio approach, the risk factor (CN_t) is directly regressed on the excess returns of a sample of portfolios, $\tilde{r}_t$:

$$CC_{CN} = \xi + \omega' Z'_{t-1} r_t + e_t.$$

Each portfolio is obtained using the weights $\widehat{\omega}$ from this regression, and its return is defined as $h_t^{CN} = \omega'\tilde{r}$ and the error $e_t$is a measurement error.

Formed portfolios are called to be a successful hedge for the climate risk if there exists an invertible matrix H such that $\tilde{r}_t = H(CN_t, v_t)$, where (CN_t, v_t) is the vector of all factors.

To apply the mimicking portfolios approach in this research, first, the firm-level characteristics, Z_t, should be defined, which is a climate risk exposure of each firm. For this purpose, the ranked and demeaned E-scores of the companies were used (as described in Section 4.1.2.4.

Another important point to consider is that the climate change risk has to be isolated from the other non-tradable risks. As it was proposed by Fama-French (2008), stocks with a high book-to-market ratio and small caps tend to perform better than the market. Therefore, market, size, and value are included in the equation (3) to account for the non-tradable factors:

$$CC_{CN} = \xi + \omega^{ES} Z_{t-1}^{ES} r_t + \omega^{Size} Z_{t-1}^{Size} r + \omega^{HML} Z_{t-1}^{HML} r + \omega^{MKT} Z_{t-1}^{MKT} r + e_t,$$

where ω^{ES}, ω^{Size}, ω^{HML}, and ω^{MKT} are weights of the e-scores, size, volume, and market portfolios in the mimicking portfolio for CC_{CN}.

The weights are obtained by performing the multivariate regression of the climate change news index on the universe of securities with climate change characteristics, size, volume, and market returns.

The performance of the hedge portfolios constructed using this formula is compared to the performance of two ETFs. It is done using the same method as with the universe of securities (Equation 4).

$$CC_{CN} = \xi + \omega^{XLE} r^{XLE} + \omega^{PBD} r^{PBD} + \omega^{Size} Z_{t-1}^{Size} r + \omega^{Size} Z_{t-1}^{Size} r + \omega^{HML} Z_{t-1}^{HML} r + \omega^{MKT} Z_{t-1}^{MKT} r + e_t,$$

And then, the results compared with the results obtained by hedging the climate news WSJ index proposed by Engle et al. (2020):

$$CC_{WSJ} = \xi + \omega^{ES} Z_{t-1}^{ES} r_t + \omega^{Size} Z_{t-1}^{Size} r + \omega^{HML} Z_{t-1}^{HML} r + \omega^{MKT} Z_{t-1}^{MKT} r + e_t,$$

$$CC_{WSJ} = \xi + \omega^{XLE} r^{XLE} + \omega^{PBD} r^{PBD} + \omega^{Size} Z_{t-1}^{Size} r + \omega^{Size} Z_{t-1}^{Size} r + \omega^{HML} Z_{t-1}^{HML} r + \omega^{MKT} Z_{t-1}^{MKT} r + e_t.$$

Here, the only term different from Equation 3 is the Climate Change News Index, as it takes the news sentiment from the Wall Street Journal.

4.3. Using new factor selection strategy

Previous researchers have used different strategies for factor selection to create mimicking portfolios. In this paper, besides using the Engle et al. (2020) factor selection strategy with the 5 Fama-French factors, the coefficient elimination strategy is used.

With the coefficient elimination strategy, the Climate News index is directly regressed on the existing universe of assets, so Engle et al. (2020) equation hold, however size, volume, and market factors are excluded from there, as well as climate characteristics and companies with the coefficient more than 0.2 were selected for the hedging portfolio, while others are eliminated from the model. A table with the resulting factors can be observed below (Table 1).

Ticker	Coefficient	Ticker	Coefficient
AIM	0,60	NBIX	0,21
AMD	0,26	NTR	0,27
BURL	0,29	ODFL	0,25
CABO	0,30	OLED	0,43
CACC	0,22	OSG	0,59
FDC	0,20	SHOP	0,21
FMCC	0,30	SPB	0,47
FNMA	0,30	SQ	0,34
IQV	0,27	VEEV	0,56
MTCH	0,31	WST	0,23
MTN	0,25	XRX	0,25

Table 1 – Tickers with coefficients > 0.2

It can be observed that 22 stocks were selected after the factor selection, which uses a coefficient elimination strategy. The mimicking portfolio can be constructed by regressing the Climate Change News Index on their returns. The figure below shows the correlations between these stocks (Figure 16).

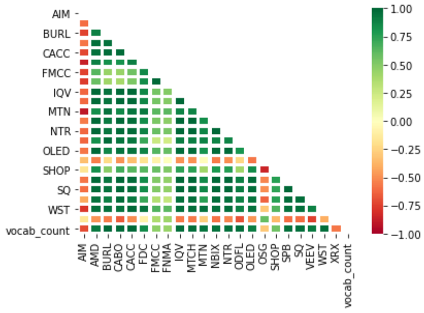


Figure 16 - Correlation between tickers selected with another selecting strategy

It can be observed from this figure that the Climate Change News Index is highly correlated with most of the selected tickers.

5. Results

5.1. Mimicking portfolio formed with the new factor selection strategy

First, the portfolio performance, which was constructed using the new factor selection strategy, was analyzed. As described in Section 4.4, the Climate Change News Index was directly regressed on the assets, and those with higher coefficients were selected to form the hedging portfolio. Weights of such a portfolio could be determined by regressing Climate Change News Index on the returns of the selected stocks and taking the betas from it. The output of the regression can be observed below (Table 2).

Stock	Weight	Stock	Weight
AIM	-0,26	NBIX	0,50
AMD	1,74	NTR	0,27
BURL	0,20	ODFL	0,23
CABO	0,39	OLED	0,35
CACC	0,33	OSG	-0,07
FDC	0,34	SHOP	0,18
FMCC	0,03	SPB	5,48
FNMA	0,02	SQ	1,06
IQV	0,35	VEEV	0,52
MTCH	0,32	WST	0,41
MTN	0,21	XRX	-0,13

Table 2 – Resulting Portfolio weights

Then, the portfolio is constructed using these weights and tests its performance on the out-of-sample period (January 2018 to May 2018).

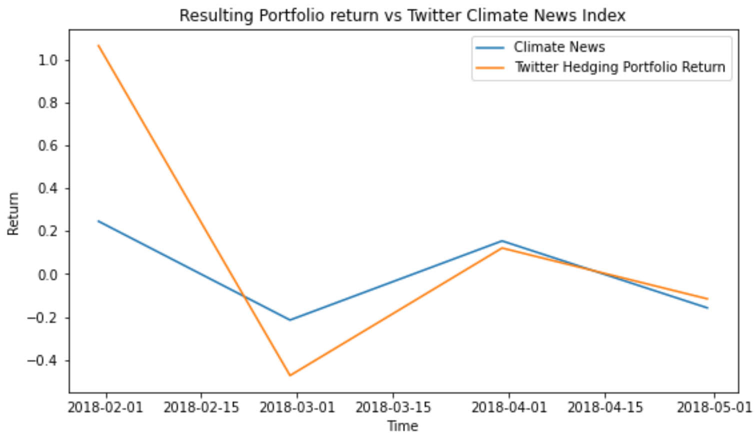


Figure 17 - Resulting Portfolio from the New Factor Selection Strategy

It can be observed from the figure above that the resulting portfolio mimics the Twitter Climate Change News Index very well, which means that the factor selection strategy proposed in this research can be used to hedge the Twitter Climate Change News (Figure 17). The correlation of such a portfolio with the Twitter Climate Change News Index on the out-of-sample period is 88.3%.

5.2. In-sample fit results of the mimicking portfolios proposed by Engle et al. (2020)

After that, the in-sample fit of the four regressions, constructed based on two different hedging strategies, was checked. One of them was constructed using the Fama-French Factors and the universe of assets with ESG-Scores in order to hedge against the Twitter Climate Change Index and the WSJ Climate Change Index, and another one was constructed using two different ETFs, XLE and PBD, as well as Fama-French factors with the same purpose. Table 3 shows the results of all four regressions (Table 3). Columns 1 and 2 show that firm-specific ESG-Scores have a positive and significant relationship with Twitter and WSJ Climate Change indexes. It means that during the times when climate change is actively discussed on Twitter or in the Wall Street Journal, companies with higher ESG-scores would yield higher returns than companies with low ESG-scores. Therefore, Hypothesis 3: " When climate change is discussed more on Twitter, firms with higher ESG-scores have lower returns than those with lower ESG-scores.", is rejected. Besides that, value and size factors have positive and significant relationships with both indices. It means that more prominent companies or companies that have higher value are more exposed to climate change news than companies, which are smaller in size or have a lower value. The R-squared measures suggest that mimicking portfolios formed using the ESG-scores can hedge about 8.5% of the climate news discussed in the Wall Street Journal or 2.2% of those discussed on Twitter. So, Hypothesis 1: "Mimicking portfolios constructed from the same universe of assets sorted by their ESG-Scores hedge more variation in Twitter Climate Change Index than in the existing Wall Street Journal Climate Change Index.", is rejected since the R-squared measure is higher for the WSJ Climate Change Index.

	(1) Mimicking portfolio – WSJ	(2) Mimicking portfolio – Twitter	(3) ETF portfolio – WSJ	(4) ETF portfolio – Twitter
ESG-score	0.0004***	0.0826***		
XLE			-0.0081	-3.2354**
PBD			-0.0048	6.3103***
HML	0.0002***	0.0219***	0.0003*	0.0645**
SIZE	0.00006***	0.0044***	0.0001	-0.0017
MKT	-0.0001***	-0.0155***	-0.00007***	-0.0524*
Constant	0.0066***	0.1326***	0.0067***	0.2138**
Observations	44,671	58,874	84	96
R-squared	0.085	0.022	0.126	0.256

t-statistics in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Columns 3 and 4 show that portfolios based on two different ETFs can hedge around 12.6% of the Wall Street Journal Climate Change News and 25.6% of the Twitter Climate Change News, which is higher than portfolios based on the company-specific ESG-scores. Therefore, Hypothesis 2: “Mimicking portfolios are a better hedge against the climate risk than the existing hedging instruments (XLE and PBD).” is rejected. XLE return has a negative and significant relationship with both indices, which means that XLE return decreases when climate change is discussed more in the Wall Street Journal or on Twitter. PBD return has a positive and significant relationship with both indexes, implying that PBD return increases with more climate news on Twitter or in the Wall Street Journal. The value factor has a positive and significant relationship with both indexes. The size factor has a positive and significant relationship with the Wall Street Journal Climate Change News Index and a negative and significant relationship with Twitter Climate Change News Index, constructed in this research. It implies that companies with higher value have larger exposure to climate change news on Twitter and in the Wall Street Journal, while bigger companies have larger exposure to the climate news in the Wall Street Journal and smaller exposure to the climate news on Twitter.

5.3. Out-of-sample fit

After that, to determine the weight of each stock in the portfolio, the following formula is used: $\omega^{ES} Z_{t-1}^{ES} + \omega^{Size} Z_{t-1}^{Size} + \omega^{HML} Z_{t-1}^{HML} + \omega^{MKT} Z_{t-1}^{MKT}$, where t-1 is the previous time period to the one, where calculation takes place, and weights are determined from the coefficients from the multivariate regression of Climate Change news index on the ESG-Scores, SMB, HML, and excess market return. This portfolio would hedge the investor from the climate risk for one month ahead and be based on the value, size, and the E-Score of the company. The portfolios with the ETFs are constructed in a similar manner using the following formula: $\omega^{XLE} r_{t-1}^{XLE} + \omega^{PBD} r_{t-1}^{PBD} + \omega^{Size} Z_{t-1}^{Size} + \omega^{HML} Z_{t-1}^{HML} + \omega^{MKT} Z_{t-1}^{MKT}$.

These portfolios indicate whether it is possible to hedge the out-of-sample climate news on Twitter and the Wall Street Journal. The periods for the out-of-sample news are January to September 2017 for the Wall Street Journal Climate News and January to June 2018 for Twitter Climate News. The hedging portfolios constructed can be observed below (Figure 18). The top two graphs show the performance of the portfolios formed based on firms' climate characteristics, and the bottom two represent the portfolios constructed using the ETFs.

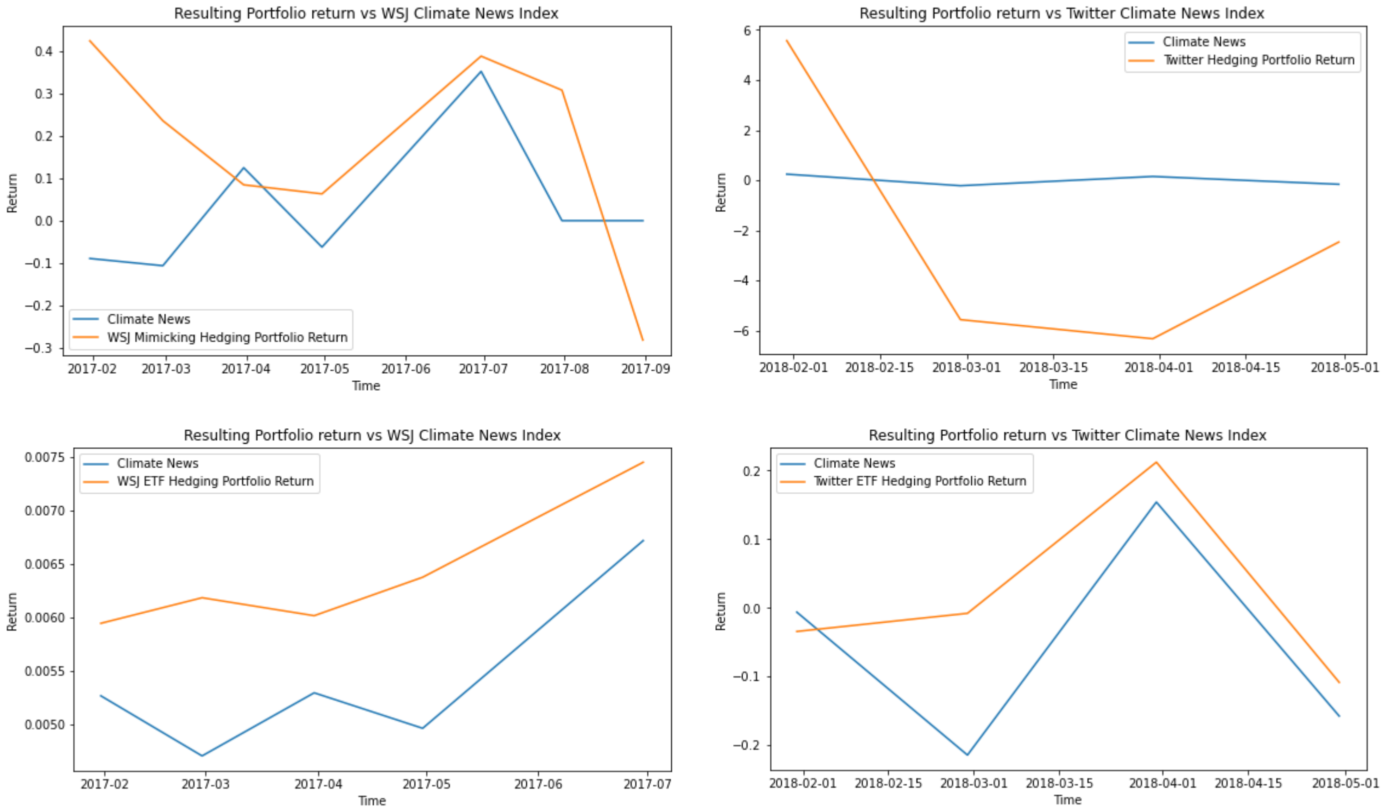


Figure 18 - Constructed portfolios

The right graphs plot the portfolios constructed to hedge the Twitter Climate Change News Index, and the left graphs plot the portfolios constructed to hedge the Wall Street Journal Climate Change News Index. The out-of-sample correlations between the Wall Street Journal Climate Change Index and the ESG-Scores sorted portfolio is 17.8%, and the resulting portfolio constructed using ETFs is 85.9%. The out-of-sample correlations between the Twitter Climate Change Index and the ESG-Scores sorted portfolio is 54.9%, and the resulting portfolio constructed using ETFs is 80.6%. Based on the correlation scores, the best performing portfolio to hedge the Wall Street Journal Climate Change News is the portfolio constructed using two ETFs, XLE and PBD, and Fama-French factors. It is the same for the Twitter Climate Change News Index since the portfolio constructed using the ESG-Scores returns and Fama-French factors outperforms the other one. However, the correlation score for the new selection strategy was even higher (88.3%), so the new selection strategy proposed in this research outperforms the strategies proposed by Engle et al. (2020).

6. Conclusion

In conclusion, this research proposes a new climate change news index constructed using the Twitter Climate Sentiment. It then uses a mimicking portfolio approach to see whether it could efficiently hedge the climate change news using in-sample and out-of-sample performance tests and builds upon Engle et al.'s (2020) work. Later on, the performance of the mimicking portfolio approach based on the climate characteristics of each company is compared to the performance of the industry bets – ETFs. The same analysis is performed for the index provided by Engle et al. (2020).

Based on both in-sample-fit results and out-of-sample fit results, it is found that the portfolios based on ETFs have a higher ability to hedge innovations in climate news than the portfolios formed based on the ESG-scores sorting. The new selection strategy proposed in this research has an even better out-of-sample correlation score, which means that it outperforms the strategies proposed by Engle et al. (2020).

More generally, this article goes deeper into the methodology proposed by Engle et al. (2020) and contributed to the existing literature on the effect of Twitter on the financial markets. Engle et al. (2020) suggest that mimicking portfolios that include climate-specific characteristics outperform the ones with two different ETFs. However, in this study, the opposite is found. The reason for that might be that the selected universe of assets is slightly different and the time frames.

The possible improvement for the result includes adding the more extended period, for which ESG-Scores are collected, perhaps another data provider with the longer and more consistent measure for the ESG-Scores. The data collected from Twitter included the historical tweets from the top 10 news accounts. However, analyzing more news accounts or using another source of climate news would improve the performance of the hedging portfolios. Besides that, reconstructing the portfolios daily might improve the portfolios' performance since there would be much more time periods available for analysis.

This research is building upon the relatively new methodology of constructing the hedging portfolios. It is essential to mention that these portfolios might not be the best performing ones, but this can be viewed as a starting point for future exploration. Things that can be improved include adding new international stocks to the hedging portfolios, including different types of assets (such as bonds or real estate), and the construction of hedge portfolios from both characteristic-sorted portfolios and ETFs.

Integrating another type of data to evaluate firm-level climate risk exposures is another significant avenue for future research. Future researchers can create firm-level climate risk exposures instead of obtaining the data from third parties. Based on previous studies, including information about geographic closeness to potential climate disasters might be a good measure to look at since it causes the salience bias (Alok, Kumar, and Wermers ,2019).

Another area for future research is to come up with alternate definitions of climate change threats. One intriguing topic is whether distinguishing between physical and regulatory climate risks is necessary. A tax on carbon emissions, for example, might lower the need for climate hedging portfolios and, as a result, the cost of hedging the climate risk if implemented widely and at a suitable amount. As a result, effective regulation will reduce the need for climate hedges. The introduction of such a tax or a similar regulatory policy, on the other hand, will produce winners and losers from regulatory risk. Hence, regulatory hedging portfolios may be helpful. The political context may have an impact on the stability of such regulatory hedging portfolios.

The predicted returns of various hedge strategies are a related question. Indeed, when investors increasingly deploy climate hedging portfolios, the price (and hence the predicted returns) of the stocks of the companies with the highest correlation with the climate change news will rise (and hence the expected returns will fall). This reduced predicted return correlates to the climate hedging portfolio's insurance premium.

Quantifying the expense of climate hedging portfolios by observing the corresponding risk premiums will be an intriguing path for future research. It is also fascinating to investigate the general equilibrium impacts of decreased capital costs for enterprises with high E- Scores, directly impacting the climatic trajectory. For example, if green energy companies' cost of capital falls, they may be able to attain efficient scale more quickly, influencing the trajectory of greenhouse gas emissions. Developing structural asset pricing models with such global equilibrium feedback loops appears to be a good research topic.

References

About Twitter | Our company and priorities. (2021). Retrieved 12 June 2021, from https://about.twitter.com/

Addoum, J., Ng, D. and Ortiz Bobea, A., (2019). Temperature Shocks and Establishment Sales. *SSRN Electronic Journal*,.

Alok, S., Kumar, N. and Wermers, R., (2019). Do Fund Managers Misestimate Climatic Disaster Risk?. *SSRN Electronic Journal*,.

Andersson, M., Bolton, P. and Samama, F., (2016). Hedging Climate Risk. *Financial Analysts Journal*, 72(3), pp.13-32.

Anthoff, D., Tol, R., & Helgeson, J. (2009). Discounting for Climate Change. *SSRN Electronic Journal*. doi: 10.2139/ssrn.1726714.

Barnett, M., Brock, W. and Hansen, L., (2019). Pricing Uncertainty Induced by Climate Change. *SSRN Electronic Journal*,.

Bollen, J., & Mao, H. (2011). Twitter Mood as a Stock Market Predictor. *Computer*, *44*(10), 91-94. doi: 10.1109/mc.2011.323

Bolton, P. and Kacperczyk, M., (2019). Do Investors Care about Carbon Risk?. *SSRN Electronic Journal*,.

Bouri, E., Jalkh, N., Molnár, P. and Roubaud, D., (2017). Bitcoin for energy commodities before and after the December 2013 crash: diversifier, hedge or safe haven?. *Applied Economics*, pp.1-11.

Choi, D., Gao, Z. and Jiang, W., (2020). INVITED EDITORIAL COMMENT: Climate Change Implications for the Asset Management Industry. *The Journal of Alternative Investments*, 23(1), pp.8-11.

Cody, E., Reagan, A., Mitchell, L., Dodds, P. and Danforth, C., (2015). Climate Change Sentiment on Twitter: An Unsolicited Public Opinion Poll. *PLOS ONE*, 10(8), p.e0136092.

Dahal, B., Kumar, S., & Li, Z. (2019). Topic modeling and sentiment analysis of global climate change tweets. *Social Network Analysis And Mining*, *9*(1). doi: 10.1007/s13278-019-0568-8

Daniel, K. D., Litterman, R. B., & Wagner, G. (2016). Applying asset pricing theory to calibrate the price of climate risk (No. w22795). National Bureau of Economic Research.

Day Trade Review. (2021). *HedgeChatter Review - Social Media Stock Sentiment Tool*. [online] Available at: <https://daytradereview.com/hedgechatter-review/> [Accessed 14 June 2021].

Delis, M., de Greiff, K. and Ongena, S., (2019). Being Stranded with Fossil Fuel Reserves? Climate Policy Risk and the Pricing of Bank loans. *SSRN Electronic Journal*,.

Dewally, M., (2003). Internet Investment Advice: Investing with a Rock of Salt. *Financial Analysts Journal*, 59(4), pp.65-77.

Dijkman, R., Ipeirotis, P., Aertsen, F., & van Helden, R. (2015). Using twitter to predict sales: A case study. *Arxiv Preprint Arxiv:1503.04599*.

Dodds, P., Harris, K., Kloumann, I., Bliss, C. and Danforth, C., (2011). Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter. *PLoS ONE*, 6(12), p.e26752.

Engle, R., Giglio, S., Lee, H., Kelly, B., & Stroebel, J. (2020). Hedging Climate Change News. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3317570

Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal Of Finance*, *25*(2), 383. doi: 10.2307/2325486

Fama, E. and French, K., (2008). Average Returns, B/M, and Share Issues. *The Journal of Finance*, 63(6), pp.2971-2995.

Finance.yahoo.com. (2021). *Yahoo is now a part of Verizon Media*. [online] Available at: <https://finance.yahoo.com/> [Accessed 14 June 2021].

Giglio, S. and Xiu, D., (2021). Asset Pricing with Omitted Factors. *Journal of Political Economy*, pp.000-000.

Gilbert, C. H. E., & Hutto, E. (2014, June). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf* (Vol. 81, p. 82).

Ginglinger, E. and Moreau, Q., (2019). Climate Risk and Capital Structure. *SSRN Electronic Journal*,.

GitHub. (2021). *twintproject/twint*. [online] Available at: <https://github.com/twintproject/twint> [Accessed 19 June 2021].

Griffin, D. and Tversky, A., (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, 24(3), pp.411-435.

Hong, H., F. Li, and J. Xu. (2019). Climate risks and market efficiency. Journal of Econometrics 208:265–81.

Hong, H., Karolyi, G. A., & Scheinkman, J. A. (2020). Climate finance. The Review of Financial Studies, 33(3), 1011-1023.

Ji, X., Chun, S., Cappellari, P., & Geller, J. (2016). Linking and using social media data for enhancing public health analytics. *Journal Of Information Science*, *43*(2), 221-245. doi: 10.1177/0165551515625029

Kaplan, A., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59-68. doi: 10.1016/j.bushor.2009.09.003

Kearney, C. and Liu, S., (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, 33, pp.171-185.

Kolk, A., & Pinkse, J. (2004). Market strategies for climate change. *European management journal, 22(3),* 304-314.

Krueger, P., Sautner, Z. and Starks, L., (2018). The Importance of Climate Risks for Institutional Investors. *SSRN Electronic Journal*,.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a social network or a news media?. In *Proceedings of the 19th international conference on World wide web* (pp. 591-600).

Mittal, A., & Goel, A. (2012). Stock prediction using twitter sentiment analysis. *Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf)*, *15*.

Murfin, J. and Spiegel, M., (2020). Is the Risk of Sea Level Rise Capitalized in Residential Real Estate?. *The Review of Financial Studies*, 33(3), pp.1217-1255.

Nofer, M. and Hinz, O., (2015). Using Twitter to Predict the Stock Market. *Business & Information Systems Engineering*, 57(4), pp.229-242.

Nofsinger, J. R. (2005). Social mood and financial economics. *The Journal of Behavioral Finance*, *6*(3), 144-160.

Nordhaus, W., (2018). *Climate change: The Ultimate Challenge for Economics*.

Parker, W., & Prechter, R. (2005). Herding: An Interdisciplinary Integrative Review from a Socionomic Perspective. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2009898 pdfreader. (2021). Retrieved 27 June 2021, from https://pypi.org/project/pdfreader/

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The Effects of Twitter Sentiment on Stock Price Returns. *PLOS ONE, 10*(9), e0138441. doi: 10.1371/journal.pone.0138441

Rao, T., & Srivastava, S. (2012). Analyzing stock market movements using twitter sentiment analysis.

Ripberger, J. T., Silva, C. L., Jenkins‐Smith, H. C., Carlson, D. E., James, M., & Herron, K. G. (2015). False alarms and missed events: The impact and origins of perceived inaccuracy in tornado warning systems. *Risk analysis*, *35*(1), 44-56.

Seltzer, L., Starks, L. and Zhu, Q., (2020). Climate Regulatory Risks and Corporate Bonds. *SSRN Electronic Journal*,.

Sharfman, M. and Fernando, C., (2008). Environmental risk management and the cost of capital. *Strategic Management Journal*, 29(6), pp.569-592.

Shive, S. and Forster, M., (2019). Corporate Governance and Pollution Externalities of Public and Private Firms. *SSRN Electronic Journal*,.

Social Blade. (2021). So*cial Blade tracks user statistics for YouTube, Twitch, Instagram, and Twitter! Get a deeper understanding of user growth and trends by utilizing Social Blade*. [online] Available at: <https://socialblade.com> [Accessed 19 June 2021].

Stocktwits. (2021). *Stocktwits - The largest community for investors and traders*. [online] Available at: <https://stocktwits.com/> [Accessed 14 June 2021].

Tafti, A., Zotti, R., & Jank, W. (2016). Real-Time Diffusion of Information on Twitter and the Financial Markets. *PLOS ONE, 11*(8), e0159226. doi: 10.1371/journal.pone.0159226

Textblob.readthedocs.io. (2021). *TextBlob: Simplified Text Processing — TextBlob 0.16.0 documentation*. [online] Available at: <https://textblob.readthedocs.io/en/dev/> [Accessed 19 June 2021].

Wilson, T., Wiebe, J. and Hoffmann, P., (2009). Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis. *Computational Linguistics*, 35(3), pp.399-433.

Xavier, C. C., & Souza, M. (2018, October). Extraction and classification of semantic data from twitter. In *Proceedings of the 24th Brazilian Symposium on Multimedia and the Web* (pp. 15-18).

Yaqub, U., Chun, S., Atluri, V., & Vaidya, J. (2017). Analysis of political discourse on twitter in the context of the 2016 U.S. presidential elections. *Government Information Quarterly*, *34*(4), 613-626. doi: 10.1016/j.giq.2017.11.001

Zheng, Y., Osmer, E. and Zhang, R., (2018). Sentiment hedging: How hedge funds adjust their exposure to market sentiment. *Journal of Banking & Finance*, 88, pp.147-160.

Appendix

```python
c = twint.Config()

c.Username = "CNN"
c.Store_csv = True
c.Output = "CNN.csv"

twint.run.Search(c)
```

Way to download the tweets

	id	conversation_id	created_at	date	time	timezone	user_id	username	name	place
0	1387425149071368195	1387425149071368195	2021-04-28 17:15:07 CEST	2021-04-28	17:15:07	200	759251	cnn	CNN	NaN
1	1387421393726365696	1387421393726365696	2021-04-28 17:00:11 CEST	2021-04-28	17:00:11	200	759251	cnn	CNN	NaN
2	1387417341894569989	1387417341894569989	2021-04-28 16:44:05 CEST	2021-04-28	16:44:05	200	759251	cnn	CNN	NaN
3	1387413856893423616	1387413856893423616	2021-04-28 16:30:15 CEST	2021-04-28	16:30:15	200	759251	cnn	CNN	NaN
4	1387410526226075650	1387410526226075650	2021-04-28 16:17:00 CEST	2021-04-28	16:17:00	200	759251	cnn	CNN	NaN

	tweet	language	mentions	urls	photos	replies_count	retweets_count	likes_count	hashtags	cashtags
0	A juror in the Derek Chauvin trial says every day in the courtroom felt like a funeral https://t.co/rigqDmMLaY	en	[]	['https://cnn.it/3nskYb8']	[]	5	6	42	[]	[]
1	15 of the safest activities for the fully vaccinated â€" with and without face masks https://t.co/vliDeeLUtl	en	[]	['https://cnn.it/3vvZ2ih']	[]	39	30	123	[]	[]
2	The man who put his feet on a desk in House Speaker Nancy Pelosi's office during the Capitol riot will be released from jail to await further court proceedings from home https://t.co/78TenoBUga	en	[]	['https://cnn.it/2Ra3HY0']	[]	151	65	275	[]	[]
3	A bomb alert in southern Germany turned out to be a false alarm after a grenade-shaped object was found to be a sex toy. https://t.co/GajvrG8eez	en	[]	['https://cnn.it/2R4epzB']	[]	121	125	548	[]	[]
4	On Wednesday, the official Covid-19 death toll in India surpassed 200,000. The US President's chief medical adviser, Dr. Anthony Fauci, stressed that helping India was "a responsibility that the rich countries need to assume." https://t.co/i8apDZwYyj	en	[]	['https://cnn.it/2QCp5p7']	[]	59	94	412	[]	[]

	link	retweet	quote_url	video	thumbnail	near	geo	source	user_
0	https://twitter.com/CNN/status/1387425149071368195	False	NaN	0	NaN	NaN	NaN	NaN	NaN
1	https://twitter.com/CNN/status/1387421393726365696	False	NaN	0	NaN	NaN	NaN	NaN	NaN
2	https://twitter.com/CNN/status/1387417341894569989	False	NaN	0	NaN	NaN	NaN	NaN	NaN
3	https://twitter.com/CNN/status/1387413856893423616	False	NaN	0	NaN	NaN	NaN	NaN	NaN
4	https://twitter.com/CNN/status/1387410526226075650	False	NaN	0	NaN	NaN	NaN	NaN	NaN

Sample output of the Twint query