NHH

Norwegian School of Economics
Bergen, Spring 2022

Green Stocks and How to Find Them

Identifying environmentally sustainable IPO firms using textual analysis and assessing their profitability

Martin Rognan Bjørneklett & Asad Mangal

Supervisor: Francisco Santos

Master thesis, Economics and Business Administration

Major: Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgements

This thesis is the finishing piece of our MSc in Economics and Business Administration, with a major in Finance. It marks the end of five enlightening and exciting years at the Norwegian School of Economics (NHH) for us both. We would like to thank our supervisor for valuable insights and ideas throughout the entire semester. Additionally, we would like to thank NHH for providing essential access to databases.

Last but not least, we would like to extend our gratitude to our friends and fellow students for making the last five years unforgettable.

Norwegian School of Economics

Bergen, May 2022

Asad Mangal

Martin Rognan Bjørneklett

Abstract

Sustainable investing has seen exponential growth among investors in recent years. To this end, ESG ratings are the tool used by investors to gauge environmental sustainability in firms. Recently listed firms lack ESG ratings, creating a market imperfection. Additionally, researchers found no correlation between carbon emissions and ESG ratings. This thesis proposes a textual analysis methodology using cosine similarity for computing environmental sustainability scores (E-scores) based on firm activities. Primarily, we provide E-scores for 366 recently listed US firms. We use selection criteria established in IPO literature. Additionally, we compute E-scores for all publicly listed US firms in 2020, approximately 10000 firms.

We indirectly verify the proposed method through an event study of the 2020 US presidential election. Firms with high E-scores had cumulative abnormal returns of 2 percent, while firms with low E-scores achieved only -0.28 percent. Through significance testing, we conclude that this difference in cumulative abnormal returns is statistically significant.

Due to the lack of ESG ratings of recently listed firms, little research has been conducted on the relation between environmental sustainability and underpricing. In this thesis, we use the proposed E-scores to establish this link. During the sample period between 2019 and 2021, we found that firms with high E-scores were underpriced by 1.44 percent, while firms with low E-scores had underpricing equal to -0.73 percent.

Keywords – ESG, Textual Analysis, Cosine Similarity, IPO, Underpricing, Event Study

Contents

List of Figures

List of Tables

1 Introduction

The 2022 report on climate change from the Intergovernmental Panel on Climate Change (IPCC) is the most recent in a long line of scientific research showcasing alarming and impending climate change impacts. As a response, investors are turning towards sustainable investing. Where traditional investing focuses mainly on balancing risk against reward, sustainable investing balances the traditional investment perspective with Environmental, Social, and Governance (ESG) insights in the pursuit of securing better long-term performance. "Sustainable investing is about profits, not taking a stand," argues BlackRock CEO Larry Fink (Norton, 2022).

We find that the existing classifications, even the widely used Bloomberg ESG rating system, do not include recently listed firms. Moreover, the classification system itself is flawed. Dimson et al. (2020) find that E-rating embedded in the ESG has very little correlation across rating providers, mainly due to the lack of standard rating criteria. In addition, Hassan and Romilly (2018) show that a high E-rating is not correlated with activities that reduce carbon emissions. These factors indicate that the existing ESG ratings are not adequate tools for sustainable investors to manage their portfolios confidently, especially when investing in recently listed firms. We believe the lack of environmental sustainability ratings of recently listed firms in the US has severe implications for effective capital allocation by sustainable investors. Consequently, we aim to suggest a method for assessing the environmental sustainability of recently listed firms.

This thesis provides a textual analysis method for computing environmental sustainability scores (E-scores) for publicly listed firms in the US. We use the Section 1: Business from 10K forms filed with the Securities and Exchange Commission (SEC), as these are legally required to describe a firms current activities and business model accurately (SEC, 2022). We find 366 relevant firms that went public in the US between 2019 and 2021, using criteria established by Ritter and Scholar (2021). The majority of these so-called IPO (initial public offering) firms do not have ESG-ratings by established providers. Therefore, the main contribution of this thesis is that it provides absolute and comparable E-scores for these 366 recently listed firms. The E-scores are computed using cosine similarity, a textual analysis method that indicates the likeness between two texts using multidimensional vector spaces (Rahutomo et al., 2012). Each firm receives an E-score between 0 and 1, where a score of 0 indicates an environmentally unsustainable firm and a score of 1 indicates a perfectly sustainable firm. For example, Sunnova Energy, a solar energy firm, receives an E-score of 0.61 and is thus placed among the top one percent most sustainable firms in our IPO sample. Gatos Silver, a silver mining firm, receives an E-score of 0.19 and is placed among the one percent least sustainable firms. The main takeaway is that neither Sunnova Energy nor Gatos Silver has ESG-ratings from an established provider, making their environmental sustainability non-quantifiable and hard to compare for investors (Eccles et al., 2017). The E-scores computed in this thesis quantify their environmental sustainability making it far easier for investors to compare the firms on this characteristic. The relative placement of these two firms makes intuitive, though anecdotal, sense.

To indirectly verify our environmental sustainability measure on a larger scale, we employ an event study of the 2020 US presidential election, replicating the research of Ramelli et al. (2021). They found that "green" firms outperformed "non-green" firms around the announcement of the 2020 election results, using ESG-ratings from Morgan Stanley Capital International (MSCI) as selection criteria for the "green" and "non-green" portfolios. To replicate this event study, we first compute the E-scores of 10024 US firms. The computed E-scores are used as the selection criteria, assigning the firms with the top one percent of E-scores to the "green" portfolio and the bottom one percent to the "non-green" portfolio. Cumulative abnormal returns are computed for each stock using the market model with the S&P 500 as a proxy for the market portfolio (MacKinlay, 1997a). We use cumulative average abnormal returns (CAAR) to compare the two portfolios. Through significance testing, we find that the "green" portfolio significantly outperforms the "non-green" portfolio, in accordance with Ramelli et al. (2021). Specifically, the "green" portfolio had a CAAR equal 2.00 percent during the event window, while the "non-green" portfolio's CAAR was -0.28 percent. The event study findings thus strongly suggest that the computed E-scores are a viable alternative to existing ESG-ratings when assessing environmental sustainability. This thesis adds to the research of Ramelli et al. (2021) by demonstrating that using our E-scores yields similar results as using existing ESG-ratings.

The event study findings legitimate the use of our environmental sustainability measure and the computed E-scores for various practical applications. We demonstrate that investments in IPO firms with high E-scores are correlated with significantly higher first-day returns compared to investments in IPO firms with low E-scores for our sample period between 2019 and 2021. In our sample, the average first-day abnormal return was 1.44 percent and -0.73 percent for firms with high and low E-scores, respectively. In other words, highly environmentally sustainable firms had 2.17 percent greater first-day abnormal returns on average during the sample period between 2019 and 2021. Considering a buy-and-hold investment strategy for the IPO firms in our sample with a one-year horizon, we still find that the firms with high E-scores outperform firms with low E-scores. On average, firms with low E-scores yield first-year abnormal returns of -33.81 percent, while firms with high E-scores yield -8.30 percent. These findings add to the research of Loughran and Ritter (1995), stating that the level of underpricing is contingent on the firms' industry sector. Therefore, one application of the E-scores computed in this thesis for sustainable investors is the management of IPO investments.

The remaining sections of this paper are structured as follows: Section 2 reviews relevant literature on ESG and textual analysis in finance. Section 3 provides an in-depth description of how the datasets employed in this thesis are constructed, and how the raw data is pre-processed before being fit for further analysis. Section 4 presents the analysis, which is structured as an applied methodology section. It presents the methodology, findings, and discussion for each of the three main segments in this thesis, namely (1) textual analysis, (2) event study, and (3) underpricing. Section 5 connects the findings of the three main segments and concludes this thesis. Lastly, section 6 provides a short discussion of limitations and suggests further research.

2 Literature Review

2.1 Environmental, Social, and Governance (ESG)

ESG is the impact firms have on three main areas: the environmental effect, societal consequences, and the governance aspect (Gillan et al., 2021). During the last decade, many providers of ESG ratings have become popular, for example, MSCI. ESG ratings are computed by the providers using a set of criteria for each main area of the ESG. A criticism of ESG-ratings is that the criteria are not universal; different providers use proprietary and distinct criteria (Li and Polychronopoulos, 2020). Although ESG ratings are a promising tool for assessing investments, a 2020 study examined the discrepancy between ESG scores from different rating providers. Emphasising that while bond ratings from a particular issuer are expected to be similar, the same could not be said for ESG ratings, highlighting extensive divergence between scores from different providers (Dimson et al., 2020).

In recent years, ESG ratings have become increasingly impactful in decision-making for finance practitioners. Additionally, these ratings are frequently used by researchers for proxying sustainability (Freiberg et al., 2020). The consequences of discrepancies between providers are severe. Dimson et al. (2020) find that ESG ratings do not provide relevant information regarding sustainability to investors. Moreover, Berg et al. (2019) finds that the divergence in ESG ratings results from measurement differences between rating providers, urging investors to retrieve raw data from many providers and construct their own ESG ratings. A criticism raised towards existing ESG literature and finance practitioners is that they use inconsistent and unreliable ESG ratings (Furlow, 2010). We seek to alleviate this problem by providing a method for evaluating and quantifying the environmental aspect of ESG without using any existing ESG ratings. Moreover, our method is able to assess recently listed firms, something not done by existing ESG rating providers.

The environmental section of ESG measures how firms affect the environment in which they operate, most frequently quantified by the level of carbon emissions. A sentiment has formed in the market during the last decade in which investors expect firms with lower carbon emissions to have greater future profitability (Trinks et al., 2020). Investors expect future regulations and taxes to penalise firms with high carbon emissions (Ardia et al., 2020), thus creating favourable market conditions for firms with lower carbon emissions. To this end, investors use the E-score embedded in existing ESG ratings to select firms with alleged lower carbon emissions. However, Hassan and Romilly (2018) shows that higher E-scores are not correlated with lower carbon emissions. For some ESG rating providers, higher E-scores positively correlate with greater carbon emissions. Li and Polychronopoulos (2020) argue that the inconsistency between higher E-scores and lower carbon emissions is due to the current E-score not accurately reflecting emissions-reducing activities. Existing ESG rating providers assess firm activities and medium and long-term strategies for reducing emissions when computing E-scores. Considering that medium and long-term strategies are not necessarily established, the link between E-scores and reduced carbon emissions becomes tenuous (Kotsantonis and Serafeim, 2019). In this thesis, the methodology for assessing environmental sustainability and the computation of E-scores is based exclusively on firm activities. We use the firms' annual reports filed with the SEC, specifically Section 1: Business, which accurately describe firm activities (SEC, 2022). An important contribution from this thesis is that the provided E-scores are solidly linked with emissions-reducing firm activities.

One of the leading research topics related to ESG within the financial field during recent years is the correlation between ESG and stock performance. A systematic review in the shape of a meta-analysis (Friede et al., 2015) uses empirical findings from 2200 studies and concludes that approximately 90 percent find a non-negative correlation between ESG and financial performance. The findings suggest that a high level of ESG is an important marker of strong financial performance (Ramelli et al., 2021). Building upon this research, we demonstrate that our computed E-scores positively correlate with stronger financial performance, specifically around the 2020 US presidential election.

The topic of underpricing is amply researched in finance; however, the link between ESG ratings and IPO underpricing is sparsely investigated as recently listed firms lack ESG ratings from reputable providers. Existing research attempts to proxy ESG ratings using textual analysis (Fenili and Raimondo, 2021). We add to this literature by providing a more complex and accurate textual analysis methodology. Additionally, we also prove a positive correlation between IPO underpricing and E-ratings.

2.2 Textual Analysis and Cosine Similarity

Textual analysis is best described as the statistician's method of analysing large quantities of textual data using empirical measures (McKee, 2003). Textual analysis was first used in finance when traders started analysing data from social media that was updated in real-time to execute short-term high-frequency algorithmic trading (Colianni et al., 2015). In newer times, textual analysis in finance is also used to measure the broader market sentiment through analysing news articles, buy/sell recommendations from brokers, and firms' annual reports.

In textual analysis literature associated with finance, bag-of-words methods are frequently used. Bag-of-words methods refer to the collection of techniques based on comparing textual data using a dictionary (McKee, 2003). Cosine similarity is one of these bag-of-words techniques. It measures the likeness in word frequencies normalised by the length of the different textual data sources. Lang and Stice-Lawrence (2015) demonstrate this technique's merit by using it to find a statistically significant link between firms' level of disclosure and the firms' level of liquidity. Moreover, Amihud and Mendelson (1986) find that firms' financial performance positively correlate with their level of liquidity. These findings indicate that firms' level of disclosure can have an effect on their financial performance.

Existing literature (Loughran and McDonald, 2011) is focused on finding accurate groups of words indicative of environmentally sustainable rather than assessing specific firms. This thesis is the first publication, to our knowledge, that quantifies environmental sustainability using the cosine similarity technique.

3 Data

This section discloses how we collect the necessary data for our thesis. The analysis performed in the subsequent section requires different types of data collected from different sources. Most of the collected data needs further preparation before being used in the analysis. Therefore, this section also includes the process of preparing the data. The goal is to accurately present data retrieval and preparation so that the datasets used in our analysis can be replicated for further research elsewhere.

3.1 Textual Data

The textual analysis employed in this thesis requires similar formatting of individual data entries. In other words, textual data for each firm needs to be structured and formatted equally. Otherwise, the comparativeness across firms is mostly nonsensical. In the US, all publicly listed firms are required to file annual reports (10Ks) with the SEC. The 10Ks follow a mandatory form and are therefore equally structured. The 10Ks are not to be confused with normal annual reports that firms presents to its shareholders, as these are much less strictly regulated. The 10K forms are legally required to present an accurate description of the firms current activities in the first section, i.e., Section 1: Business (SEC, 2022). With this in mind, we limit ourselves to publicly listed US firms. In the following subsections, we describe the process of identifying firms and collecting textual data for these firms.

3.1.1 Textual Data of Relevant IPO Firms

The SEC offers a database to the public containing all information about firms listed in the US. The database contains various forms about many different aspects of the financial developments of the firms and is called EDGAR. The database has an application programming interface (API) that allows for automated data retrieval. The API has a long list of requirements intended to protect the SEC database from overuse by poorly structured retrieval programs. Attempting to engage the API to retrieve data without following each requirement is met with a ban from the server. Additionally, there are limitations on the maximum amount of raw data one user can retrieve in a given period.

This limitation strongly encourages API users to download only necessary data, i.e., select relevant sections from the wanted forms.

We start by identifying all firms that underwent an IPO in the US during the last ten years. We do this by scraping the SEC database for S-1 forms. Scraping refers to the process of retrieving wanted information from a web resource, which is necessary here as the SEC does not offer S-1 forms for automated download. The S-1 form is a mandatory initial registration form for all firms wanting to issue shares on US stock exchanges. The form includes financial information regarding the security issuance and the firm's central identification key (CIK). We develop the necessary scraper and retrieve all CIKs for firms that filed an S-1 form in the last ten years.

The next step is to engage the API to retrieve 10Ks for all the relevant firms, as identified by their CIKs. We note that many firms do not have 10K available on the SEC database. Firms lack 10Ks due to being too young to have filed an annual report, i.e., firms that went public during the second half of 2021. Additionally, some firms file S-1 forms but do not go public due to bankruptcy or unfavourable market conditions. Removing these firms leaves us with a list of CIKs for firms that went public during the last ten years with a 10K report in the SEC database.

Next, we need to access the 10Ks for the relevant firms. We employ the API supplied by the SEC to retrieve these. A 10K filing typically ranges from 200 to 400 pages, so a considerable amount of time and network capacity is needed to retrieve all reports in their entirety. Additionally, the majority of a 10K contains information not suited for our textual analysis. The section that best describes the business model of a firm is Section 1: Business and is often limited to approximately five pages. We develop an algorithm that separates Section 1 from the rest of the 10K, significantly reducing the contributed network load on the SEC database. However, the algorithm cannot consistently identify and separate Section 1 from 10Ks filed before 2019. This limitation results in our primary dataset of IPO firms ranging from 2019 to 2021.

The primary dataset in this thesis contains textual data for all firms that went public in the US from 2019 to 2021. The reasoning is mainly the technical restrictions imposed by the SEC and the issues with separating Section 1 for older filings. As the main contribution of this thesis is to provide E-scores for newly listed firms that lack assessment from established providers, we argue that the period ranging from 2019 to 2021 is sufficient.

We are interested in operational IPO firms and not simply holding companies. Following the methodology of Ritter and Scholar (2021), we remove all firms that are exclusively (1) holding companies, (2) acquisition companies, (3) investment companies (4) SPACs. We also remove all firms where Section 1 is not identified, i.e., the 10K report is not present in the database. These steps greatly reduce the number of IPO firms in our sample. The number of relevant IPO firms from 2019 to 2021 is 366, and the subsequent analysis is based on these firms.

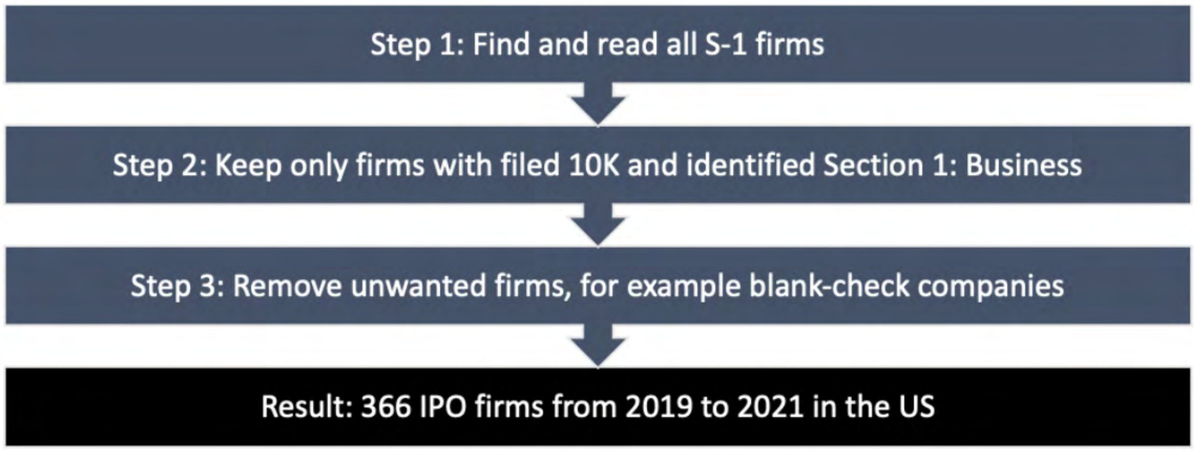

Figure 3.1: Identification of the relevant IPO firms that went public in the US between 2019 and 2021 is explained using four steps. First, engage the API offered by the SEC to read all S-1 forms. Retain only firms with 10Ks and separable Section 1. Remove non-operational firms using criteria established by literature. At last, retain 366 IPO firms to be used in the analysis.

3.1.2 Identifying All Publicly Listed Us Firms

One of the methods we use to indirectly verify the accuracy of our environmental sustainability measure uses a broad event study. We need the Section 1s for all firms listed on US exchanges for this verification process. We use SEC's comprehensive list of all publicly listed firms, identified by firm name and central identification key (CIK). A total of 13737 firms are listed here (SEC, 2022). We apply the same selection criteria (1) - (4) as described earlier, reducing the number of relevant firms to approximately 10000. The event study focuses on the 2020 US presidential election; therefore, we retrieve Section 1s from the 10Ks filed in 2020 for all remaining firms.

Figure 3.2: Identification of all relevant publicly listed US firms is simplified into four steps. First, retrieve all CIKs from the SEC. Retain only firms with 10Ks and separable Section 1. Remove non-operational firms using criteria established by literature. At last, retain 10024 firms to be used in the analysis.

3.1.3 Textual Data for Reference Group

The 10K reports for all reference group firms are retrieved from the SEC, and we isolate the appropriate sections. Textual data we retrieve about the environmentally sustainable firms serve only as a guideline for what we can expect from firms with a higher level of environmental sustainability. Therefore, we aggregate the individual textual data elements from the reference group. We merge all section 1s from the reference group firms into one large reference corpus, which is a form of storing textual data in a manner suited for statistical analysis (McKee, 2003). This reference corpus readily provides us with a dictionary of words and phrases used by reference groups, and paired with the corresponding frequencies; we now have a firm expectation of how the section 1s of environmentally sustainable firms are structured and worded.

3.1.4 Preparing Textual Data for Empirical Analysis

This section discloses the pre-processing steps employed on our textual data. This section can be viewed as a preamble to the next section, in which the textual analysis method for computing firms' environmental sustainability scores is disclosed.

Raw textual data is not suited for use straight away as such data contain many outright harmful or non-valuable elements in the context of textual analysis. Through pre-processing, we limit the occurrence of such elements in our analysis, thus furthering the robustness of our findings. The measures we use in our pre-processing are cross-validated against the methods most frequently used by prominent linguists and are therefore deemed necessary (Roiger, 2017).

Our pre-processing steps are summarised as follows:

I. **Removal of all words consisting of two characters or less**

Abbreviations, transcript errors, and formatting errors result in very short words not suited for measuring likeness (McKee, 2003). Abbreviations occur too infrequently and do not convey the same meaning across the entire textual data sample. The listed errors are, by nature, hindrances to effective textual analysis. Therefore, our algorithm iterates over each word in the textual data and removes these elements.

II. **Removal of all non-alphabetic characters**

In the textual analysis, the critical property we are looking for is the degree of environmental sustainability exhibited through the firms' annual reports. There is no need for numbers, special characters, or symbols in the textual analysis (McKee, 2003). With this in mind, only words consisting of the 26-letter Modern English alphabet are permitted.

III. **Removal of stopwords**

Stopwords are words that are a necessary part of both written and spoken language but that individually convey very little linguistic meaning. Our textual data varies in length for each firm, as Section 1s of the 10K reports are not of equal length across firms. A longer Section 1 will contain more stop words than a short Section 1. The sheer frequency of stop words in the textual data inflates their importance. It would be detrimental to our textual analysis if the length of firms' Section 1 were the deciding factor in our analysis, rather than the content of the Section 1s. The omittance of stop words is essential in the pre-processing procedure to increase accuracy and efficiency.

IV. **Reducing all remaining words to lemma**

All textual analysis is case-sensitive, which means that only exact matches are counted. In practice, "smiling" and "smiled" are counted as two different words.

However, the meaning of a paragraph containing either word frequently is expected to be similar. Therefore, we need to replace each word with its lemma. The lemma of a word is the form in which that word is presented in a dictionary. Comparison between un-lemmatized textual data from two sources is inaccurate (Greco and Polli, 2020). Each remaining word is reduced to its lemma by running all our textual data through extensive dictionaries. Input words are replaced by their respective lemma, i.e., "smile" replaces both "smiling" and "smiled".

V. **Lowercase and excessive spacing removal**

As explained earlier, textual analysis is case-sensitive, meaning that "Electric" and "electric" are two different words. To further negate this issue, all characters are turned to lowercase, effectively sidestepping any problems with case sensitivity.

Pre-processing steps I through III remove words, creating blank spaces in our textual data. In addition, we expect all textual data to exhibit formatting discrepancies and errors to some degree. Both these factors create redundant trailing spaces or newlines. We increase efficiency by removing all trailing spaces and empty new lines, reducing the amount of textual data analysed later.

3.1.5 Example of Textual Data Pre-processing

The pre-processing steps and the pre-processing process may seem arbitrary; therefore, we provide a small demonstration of its importance.

1. *"We design, develop, manufacture, sell and lease high-performance fully electric vehicles and energy generation and storage systems, and offer services related to our products." (Excerpt Tesla 10K report 2021)*

2. *"Industrial Services of America, Inc. is a Louisville, Kentucky-based company that buys, processes and markets ferrous and non-ferrous metals and other recyclable commodities and buys used autos in order to sell used auto parts." (Excerpt Industrial Services of America 10K report 2019)*

3. *"Southwestern Energy Company is an independent energy company engaged in exploration, development and production activities, including the related marketing of natural gas, associated natural gas liquids and oil produced in our operations." (Excerpt Southwestern Energy Company 10K report 2021)*

The processed samples 1-3 below convey the same information about the three firms as the initial samples. The advantage of using the last three samples of textual data is that they are comparable and far shorter, making the following analysis less computationally demanding.

1. *"design develop manufacture sell lease high performance fully electric vehicle energy generation storage system offer service relate product" (Excerpt Tesla 10K report 2021, processed)*
2. *"industrial service america inc louisville kentucky base company buy process market ferrous non ferrous metal recyclable commodity buy use auto order sell use auto part" (Excerpt Industrial Services of America 10K report 2019, processed)*
3. *"southwestern energy company independent energy company engage exploration development production activity include relate market natural gas associate natural gas liquid oil produce operation" (Excerpt Southwestern Energy Company 10K report 2021, processed)*

3.2 Financial Data

We need financial data for the event study of the 2020 US presidential election and the investigation of underpricing in recent IPOs. We extract the CIKs for the firms with the highest and lowest one percent of E-scores for the event study, totaling 230 firms. Similarly, we extract the CIKs for the firms with the highest and lowest five percent of E-scores when computing average underpricing, totaling 36 firms. The API for the SEC database allows for ticker look-up, so we convert all these CIKs into stock tickers. This conversion is necessary as the financial information database requires tickers for firm identification.

We retrieve financial data from the Centre of Research in Security Prices (CRSP) by supplying the database with extracted tickers. All financial analysis in this thesis uses adjusted close price as a proxy for security price. Moreover, movements in the S&P 500 index proxies general market development in the event study, as seen in Ramelli et al. (2021). Conversely, the Nasdaq Composite Index is used as a proxy for market development when computing underpricing for IPO firms, as seen in Loughran and McDonald (2011).

4 Analysis

This section presents the methodology employed in the thesis and the resulting findings. The methodology is applied consecutively, meaning that each of our three main methodology segments is followed by the corresponding results.

We start by defining our environmental sustainability score (E-score) before computing E-scores for our primary sample, which contains IPOs in the US from 2019 to 2021. The main contribution of this thesis is that it provides E-scores for recently listed firms as these firms lack assessments by existing providers. Following this, we provide an indirect verification of our environmental sustainability measure through an event study of the 2020 presidential election, as seen in Ramelli et al. (2021). The event study requires E-scores for the much larger dataset consisting of all publicly listed US firms; these E-scores are also computed and presented here. Lastly, one practical application of the E-scores computed for IPO firms is demonstrated. We look at the level of underpricing exhibited by the IPO firms and investigate its correlation with the respective E-scores.

4.1 Textual Analysis

We employ a score-based approach when quantifying the environmental sustainability of firms. This means that all the firms' annual reports are handed to an algorithm, which computes a score for each firm. The score is comparable across both firms and industries, effectively ranking firms according to their hypothesised environmental sustainability. This section explains the logic behind the algorithm used to calculate E-scores and presents the results thereof.

4.1.1 Textual Analysis Methodology

Our algorithm uses cosine similarity as an estimator for E-score. Cosine similarity is a quantitative linguistic measure of how similar one text is to another (McKee, 2003). In other words, the environmental sustainability score for each firm is dependent on how similar that firm's annual report is to a reference group's combined annual reports (Xia et al., 2015). Therefore, the first step is to select a representative and unbiased reference group of environmentally sustainable firms.

I. Reference group of environmentally sustainable firms

The reference group needs to be representative across sectors, for different market capitalizations, and for the period in which our analysis takes place. In addition, the firms in the reference group must be listed on US stock exchanges as we need access to their 10K reports. We do not want the reference group to be affected by any bias on our part, precluding the manual selection of firms. Instead, we construct our reference group by looking at the firms in the portfolios of exchange-traded funds (ETFs) promoting environmental sustainability. We find several ETFs with the explicit goal of promoting clean energy or environmental sustainability.

The largest fund in this sector (measured by total assets) is the iShares Global Clean Energy ETF (ticker ICLN), composed of many countries' firms in the clean energy sector. The second to largest fund is the First Trust Nasdaq Clean Edge Green Energy Index ETF (ticker QCLN), which is very similar but invests solely in firms listed on the Nasdaq exchange. The Invesco MSCI Sustainable Future ETF (ticker ERTH) is an especially exciting fund. This fund adds firms to its portfolio based on sustainability rather than being limited to the clean energy sector (MSCI, 2021).

The ETF we use as a basis for the reference group in our textual analysis is ERTH. One reason is that the ERTH fund invests in firms that have a measurable impact on global environmental sustainability regardless of their industry. In addition, ERTH only invests in firms that can prove a higher than 75 percent cumulative income stream from environmentally sustainable businesses (MSCI, 2021). We use all the firms listed in the US from the ERTH portfolio, totaling the 23 firms presented in table 4.1. This group of firms will be referred to simply as our reference group hereafter.

We investigate the sensitivity of our textual analysis findings regarding the choice of the reference group and find this to be minimal. We vary the size of the reference group, ranging from all the US stocks in ERTH to only the top 10 measured by portfolio allocation. We also investigate whether significant discrepancies can be found when using different ETFs as the reference group. We find that computed E-scores using the different reference groups mentioned have a correlation larger than 70 percent. In addition, 65 percent of the firms in the top 5 percentile of scores are the same across reference groups. The potential bias of choosing a reference group is minimised in the context of our textual analysis.

Holding	**Percentage of ETF**
Digital Reality	5.49 %
Tesla	5.43 %
Enphase Energy	4.67 %
SolarEdge Technologies	3.10 %
Darling Ingredients	2.36 %
First Solar	1.48 %
Advanced Drainage Systems	1.18 %
NextEra Energy	1.14 %
Ormat Technologies	0.72 %
Switch	0.64 %
Meritage Homes	0.58 %
Badger Meter	0.54 %
KB Home	0.50 %
Renewable Group	0.43 %
Schnitzer Steel Industries	0.36 %
Sunnova Energy	0.35 %
Blink Charging	0.27 %
Energy Recovery	0.15 %
Mercer	0.12 %
Workhourse Group	0.11 %
TPI Composites	0.10 %
Rex American Resources	0.10 %

Table 4.1: This table presents the US stock holdings in the ERTH exchange-traded fund. Note that holdings change continuously with stock prices and investment strategies. Presented holdings and weightins were retrieved on the 14.03.2022 from the MSCI webpage.

II. Creating a document term matrix (DTM)

To measure the likeness of one corpus to another, we convert the corpora into something called a document term matrix. A document term matrix is a mathematical method used to represent and understand corpora (McKee, 2003). The DTM is a matrix where the rows contain documents or textual data references, and the columns contain phrases or terms. Each firm's corpus has its respective row in the DTM, with the last row in the matrix being allotted the reference group corpus. The firms are identified through their central index key (CIK), with the reference group corpus having the only named row.

We exemplify a simple DTM using the samples we pre-processed earlier. The pre-processed samples are shown below. Note that these samples are not translated to the corpora form we employ in our methodology. The reason for this is that the samples are small and contain limited data, meaning that the need for the efficiency gained through the use of corpora is reduced.

1. *"design develop manufacture sell lease high performance fully electric vehicle energy generation storage system offer service relate product" (Excerpt Tesla 10K report 2021, processed)*

2. *"industrial service america inc louisville kentucky base company buy process market ferrous non ferrous metal recyclable commodity buy use auto order sell use auto part" (Excerpt Industrial Services of America 10K report 2019, processed)*

3. *"southwestern energy company independent energy company engage exploration development production activity include relate market natural gas associate natural gas liquid oil produce operation" (Excerpt Southwestern Energy Company 10K report 2021, processed)*

We translate the textual data samples 1-3 into a DTM, as presented below.

Sample/Word	design	develop	sell	electric	energy	...
1: Tesla	1	1	1	1	1	...
2: Industrial Services of America	0	0	1	0	0	...
3: Southwestern Energy Company	0	0	0	0	1	...

Table 4.2: Example of a DTM constructed using the pre-processed textual data samples. The first column specifies which textual data element, called document, that row is in regards to. Thereafter, each column specifies that frequency with which each individual word occurs in that document.

The DTM allows for easy and efficient comparison of textual data. Many of the textual analysis methods used in literature (Lang and Stice-Lawrence, 2015) presumes the computation of a complete DTM. This is also the case when using cosine similarity.

The example DTM presented above is incomplete, as many of the words used in the three samples are not presented in the matrix. The complete matrix for the three short samples presented contains more than forty columns if we catalog only one-word terms. If we included phrases composed of two words each, the matrix would need more than approximately ninety columns. As demonstrated, a drawback of many textual analysis methods that use a DTM is that the methods demand exponential amounts of computational power relative to textual data size.

Our analysis is limited to a DTM mapping exclusively one-word phrases. This is due to a large amount of data and the fact that phrases are not necessarily better at conveying meaning in the textual data compared to single words (Loughran and McDonald, 2011). Even with this limitation, our final DTM contains no less than 30 000 columns derived from unique and accepted words. Table 4.3 shows the top 20 words used the most in the reference group texts, measured by relative frequency. We include them here to showcase (anecdotal) characteristics of the reference group's data.

energy	design	development	material	storage
customer	power	solution	facility	technology
product	solar	geothermal	cost	plant
wind	market	company	business	renewable

Table 4.3: The twenty most relatively often used words in the aggregated reference group texts. These words given an (anecdotal) indication of topics discussed in the combined textual data from the reference group firms.

II. Creating a document term matrix (DTM)

Cosine similarity is a measure of the degree of similarity between two vectors. This measure stems from the mathematical inner product family and serves as a likeness measure in various econometric research (Fenili and Raimondo, 2021). Within the classification field, cosine similarity is frequently used to indicate the degree of similarity between documents. We use cosine similarity to measure the likeness between firms' and the reference group's textual data through the constructed DTM. The general outline of the final DTM is similar to figure 4.2, with approximately four hundred rows and thirty thousand columns.

Imagine a multidimensional space with vectors of different lengths and different directions. Each vector represents the language in one firm's section 1, and each dimension represents a word. The vectors that are most in line with the reference group's vector have the highest cosine similarity. Mathematically, it measures the angle between two vectors on a unit sphere. This translates to the product of two term vectors normalised with the vectors' length (Xia et al., 2015).

$$Cosine\ similarity = \frac{\overrightarrow{firm} * \overrightarrow{reference\ group}}{\overrightarrow{||\ firm\ ||} * \overrightarrow{||\ reference\ group\ ||}} \tag{4.1}$$

In equation 4.1, $\overrightarrow{firm}$ refers to the term vector of one firm, and $\overrightarrow{reference\ group}$ is the term vector of the reference group. Each pair of vectors contains the reference group's term vector, with a variable second vector as we look at each firm. We repeat this calculation for all firms we analyse.

The cosine similarity measure has values between 0 and 1, where a value of 1 indicates identical text and a value of 0 indicates no degree of similarity. In this context, a value of 0 indicates that a firm's Section 1 has no textual similarity to the reference group's combined Section 1s. Thus we expect that the firm is not environmentally sustainable. The inverse is necessarily true for firms with a cosine similarity value equal to 1. We use the cosine similarity value computed for each firm as an E-score. We rank all firms according to their corresponding E-score, resulting in a decreasing list of environmentally sustainable firms.

In order to perform the textual analysis methodology described, we develop an automated algorithm using both R and Python 3. The following section presents and discusses the results from executing the automated algorithm on both the IPO firm sample and the much larger sample of all public US firms.

4.1.2 Textual Analysis Findings

We compute E-scores for the 366 relevant firms that went public on US securities exchanges between 2019 and 2021. We see that the resulting E-scores are centred around approximately 0.31, as seen in Figure 4.1. Furthermore, the highest and lowest calculated scores are 0.61 and 0.12, respectively, indicating that the sample is slightly skewed towards higher E-scores.

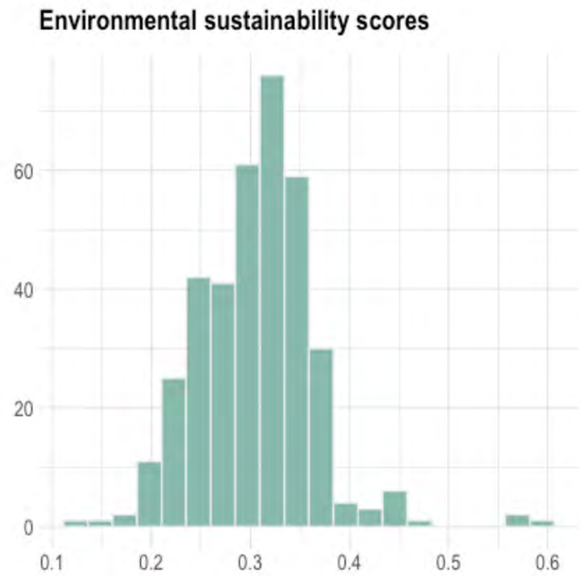

Figure 4.1: This figure presents the distribution of E-scores for the 366 assessed US IPO firms in the sample period between 2019 and 2021. We note the clear outliers in both directions, especially the visible gap between the highest E-scores.

Figure 4.1 presents a histogram of the distributions of firms' E-scores. We note the outliers with E-scores equal to approximately 0.60. Likewise, outliers are found at the lower end of the E-score spectrum with scores around 0.10. However, there is a gap between the outliers on the upper end, which is not present on the lower end of the E-score spectrum. This indicates a significant linguistic difference between the firms with the highest E-scores and the firms following behind. Our automated algorithm indicates that these outlier firms are the most and the least environmentally sustainable in the sample of firms that went public from 2019 to 2021 in the US.

Table 4.4 presents the firms in the top five percentile of environmental sustainability scores. We present these firms with their respective environmental sustainability scores and standard industrial classification (SIC).

Name	Score	SIC(sector)
Array Technologies	0.61	3990 - Miscellaneous Manufacturing Industries
Sunnova Energy	0.61	4931 - Electric & Other Services Combined
Shoals Technologies	0.60	3674 - Semiconductors & Related Devices
Allegro Microsystems	0.59	3674 - Semiconductors & Related Devices
Fluence Energy	0.58	3690 - Miscellaneous Electrical Machinery
Avantor	0.57	2890 - Miscellaneous Chemical Products
Pactiv Evergreen	0.57	2673 - Plastics, Foil & Coated Paper Bags
Pagerduty	0.47	7372 - Prepackaged Software
Mayville Engineering	0.45	3460 - Metal Forgings & Stampings
Sitime Corp	0.45	3674 - Semiconductors & Related Devices
Parsons Corp	0.44	7373 - Computer Integrated Systems Design
Montauk Renewables	0.43	4932 - Gas & Other Services Combined
Compass Diversified	0.41	2510 - Household Furniture
Agrify Corp	0.40	0700 - Agricultural Services
Onewater Marine	0.39	5531 - Retail-Auto & Home Supply Stores

Table 4.4: This table presents the firms with highest computed E-scores using our cosine similarity measure. We use the dataset containing the 366 relevant IPO firms identified between 2019 and 2021 for the US stock market. Additionally, the standard industrial classification codes are retrieved from the SEC database.

At the top, we find Array Technologies, a firm that produces state-of-the-art solar tracker technology . Following close behind, we find Sunnova Energy, another solar technology firm (SEC, 2022). We see a clear trend of renewable energy firms performing very well in our ranking of environmental sustainability scores. However, examining the entire sample, there is no significant grouping of industries in the ranking.

Table 4.5 presents the firms in the bottom five percentile of environmental sustainability scores. We present these firms with their respective environmental sustainability scores and standard industrial classification (SIC).

Name	Score	SIC(sector)
Aspirational Consumer Lifestyle	0.22	4522 - Air Transportation
Appharvest	0.22	0100 - Agricultural Production-Crops
Snowflake Computing	0.21	7372 - Prepackaged Software
EVmo	0.21	7374 - Services - Computer Processing
DMY Technology Group	0.20	7990 - Services
CM Life Sciences	0.20	8000 - Health Services
Super League Gaming	0.20	7900 - Amusement & Recreation
Samsara Luggage	0.20	5990 - Retail Stores, NEC
Gohealth MD	0.20	8741 - Management Services
Forum Merger III Corp	0.19	3711 - Motor Vehicles
Fellazo Corp	0.19	2750 - Commercial Printing
Gatos Silver	0.19	1040 - Gold and Silver Ores
One Clean Planet	0.17	8200 - Educational Services
FS Development Corp	0.16	2834 - Pharmaceutical Preparations
890 5th Avenue Partners	0.12	4899 - Communications Services

Table 4.5: This table presents the firms with lowest computed E-scores using our cosine similarity measure. We use the dataset containing the 366 relevant IPO firms identified between 2019 and 2021 for the US stock market. Additionally, the standard industrial classification codes are retrieved from the SEC database.

At the bottom, we find 890 5th Avenue Partners, an investment firm specialising in telecommunications investments. Following close behind is FS Development Corp, a pharmaceutical firm (SEC, 2022). There is no trend among these firms with the lowest environmental sustainability scores, as opposed to those with the highest scores where we noted a tendency of renewable energy firms.

We note the clear trend of renewable energy firms placing high in our E-score ranking of all publicly listed US firms, figure 4.6. This trend is more prominent when assessing the larger dataset than seen in the IPO sample.

Name	**Score**	**SIC(sector)**
Daystar Technologies	0.76	3674 - Semiconductors & Related Devices
Ses Solar	0.74	3433 - Heating Equipment
Perfectenergy International	0.74	3674 - Semiconductors & Related Devices
Worldwide Energy	0.74	3990 - MISC. Manufacturing Industries
Solar Enertech	0.72	3674 - Semiconductors & Related Devices
Spire Corp	0.69	3674 - Semiconductors & Related Devices
Harry's Trucking	0.66	3990 - MISC. Manufacturing Industries
Regent Technologies	0.64	1311 - Crude Petroleum & Natural Gas
Ascent Solar Technologies	0.63	3674 - Semiconductors & Related Devices
Clear Skies Solar	0.62	3433 - Heating Equipment
Open Energy Corp	0.62	3674 - Semiconductors & Related Devices
Windgen Energy	0.61	6794 - Patent Owners & Lessors
Sunedison Corp	0.61	4911 - Electric Services
Sunnova Energy	0.61	4931 - Electric & Other Services
Real Goods Solar	0.60	1700 - Special Trade Contractors

Table 4.6: This table presents a sample of the firms with the highest computed E-scores using our cosine similarity measure. We use the dataset containing all relevant firms from the US stock market present in 2020. Additionally, the standard industrial classification codes are retrieved from the SEC database.

The textual analysis in itself is accurate. We perform test runs where we compute environmental sustainability scores of two hypothetical annual reports, where the only difference is one single sentence containing the word "energy". A difference in computed scores, though very small, is observed. Thus, we are confident that the textual analysis methodology is finely tuned and accurate. This is not synonymous with the actual assessment of the firms' environmental sustainability scores being accurate. Instead, the accuracy of the assessment is contingent on the representativeness of the reference group's textual data and the resulting dictionary thereof. As we are using the largest ETF in the environmental sustainability sector (MSCI, 2021), we are confident that the reference group's textual data is representative of environmentally sustainable firms.

The objectivity of the firms' assessment is contingent only on the choice of the reference group. There are no other exogenous parameters in the textual analysis or the computation of E-scores. We reduce any bias by leaving the choice of reference group firms to an external source, i.e., the managers of the ERTH fund.

One weakness of our textual analysis, and thus the assessment of firms' environmental sustainability, is due to the effect of negative screening. We rank firms by their computed environmental sustainability scores. Keep in mind that the firms with the most accurate environmental sustainability scores are the firms that receive the highest scores (Wei et al., 2016). Our environmental sustainability score measures the likeness of a firm's textual data to the reference group's textual data. If two firms' textual data is very unlike the reference group's, then the ranking of these two firms will be contingent on a small degree of difference in likeness. Therefore, the firms with the highest scores are more environmentally sustainable than those with the lowest scores. This indicates that the ranking of firms in the lower percentiles is less robust. This effect is discussed in the literature, as Trinks et al. (2020) finds that negative screening often leads to inaccuracy in ESG ratings due to similar reasons.

In order to indirectly verify or backtest the environmental sustainability score, we perform an event study of the 2020 US presidential election. The motivation for this is the boost in the financial performance of stocks in the environmental sustainability sector observed by Ramelli et al. (2021). We argue that if our assessment is accurate and objective, firms with high E-scores will have better financial performance in the period around the 2020 election compared to firms with low E-scores. We now use financial and textual data from all publicly listed US firms to perform this event study.

4.2 Event Study of the 2020 Presidential Election

Ramelli et al. (2021) suggests that "the green industry", or rather the environmentally sustainable sector, experienced a significant increase in stock prices when Joe Biden won the 2020 presidential election. According to Randall (2020), this increase resulted from the economic reforms embedded in Biden's executive order with the express purpose of catalysing the environmentally sustainable firms in the US. The challenge investors face is identifying firms belonging in the "green industry" successfully. We solve this challenge here by using our environmental sustainability measure.

We use event study methodology to investigate the effect of Joe Biden winning the 2020 presidential election on two groups of stocks, as seen in Ramelli et al. (2021). The two groups are (1) the firms with the top percent of E-scores and (2) the bottom one percent of E-scores. These indicate that group (1) should outperform group (2) in the relevant period, given that the firms in group (1) are environmentally sustainable and the firms in group (2) are not.

Indirectly, the event study thus serves as a verification method for our environmental sustainability measure. We underline that this verification method is indirect and that the conclusions gained are indications and not proof.

4.2.1 Event Study Methodology

The event study methodology is popular in finance and accounting research when investigating the relation between security prices and relevant economic events (MacKinlay, 1997b). Assuming a market subject to the semi-strong form of the Efficient Market Hypothesis (EMF), the security prices of affected firms should immediately reflect the change in risk, future cash flows, or underlying growth (Fama, 1970) induced by the event. A challenge is isolating the effect of the relevant event from arbitrary and unrelated developments in the market. The event study methodology has improved through various research, so that general market price movements and other confounding events are removed, neutralising this problem (MacKinlay, 1997b).

MacKinlay (1997b) claims that there is no consensual structure to event studies. However, a general chronology of event studies is discernible by looking into the literature. The studies start by defining the event that is thought to have affected stock prices in the period of interest. In our case, that event is the election of Joe Biden as president in 2020.

The period of interest is called the event window, of which the length is a parameter. Issues with confounding events and unrelated market movements is lessened with shorter event windows (Armitage, 1995). Dann et al. (1977) found relevant information to be reflected in stock prices within fifteen minutes, suggesting smaller event windows. However, MacKinlay (1997b) states that choosing an event window slightly larger than the specific period of interest is favourable. I.e., an event window that starts one day before and ends one day after the event is more accurate as it picks up on price movement after market close. We also reflect on the possibility that information may be leaked or made available to the different segments of the market at different times, thus increasing the need for an enlarged event window.

There was a large amount of disinformation and ambiguity regarding the 2020 election results (Benaissa Pedriza, 2021). Therefore, our period of interest is not when Biden was sworn into office on January 20th. Nor can we limit it to the day Biden won a majority of the electoral votes, November 7th. Instead, we set our event window to be centred around November 7th, defining this date as $t = 0$ in event time. In addition, we add a one day to the event window in both directions, thus expanding our event window to [-1, 0, 1] in event time. In real-time, which is trading days in our case, this translates to the dates 5th, 6th, and 9th of November 2020.

The event is identified, and now the next step is to define the selection criteria for observations included in the study. We define only two portfolios to be compared. We have no selection criteria for volume, price, or price development for the stocks in the portfolios. However, we use the data sample consisting of all publicly listed US firms in 2020 for this event study. The two portfolios consist of (1) firms with top one percent of environmental sustainability scores and (2) firms with the bottom one percent of environmental sustainability scores. Hereafter called **portfolio 1** and **portfolio 2**, respectively.

Stock returns cannot be compared directly, as firms are more or less risky. Investors expect to be compensated for risk through higher returns in a rational market. Stocks also have different levels of correlation to the overall market, i.e., different betas. To accurately measure the effect of an event on a firm, event studies use abnormal returns. Abnormal returns are defined as the difference in observed return and the expected return of a given stock at the time of the event. The expected return is not observed and is thus a hypothetical value that needs to be calculated. The abnormal return for each firm at a given time can be defined as follows:

$$AR_{i,t} = R_{i,t} \; - \; E(R_{i,t}|X_t) \tag{4.2}$$

Where $AR_{i,t}$, $R_{i,t}$, and $E(R_{i,t}|X_t)$ are the abnormal, observed, and expected returns, respectively, for a given time. The expected returns are hypothetical returns that the stock would have if the event had not occurred. These expected returns are computed through a normal return model, which uses historical data to generate expected returns. In event study methodology (MacKinlay, 1997b) there are two main models used to compute expected returns, namely **(1) the market model** and **(2) the constant mean return model**.

The idea behind the market model is that stock returns are contingent on market returns. Individual stocks have different normal returns since they exhibit different levels of correlation with the market. According to the market model, the normal return for a stock can be determined by adjusting market returns by the level of correlation between that individual stock and the market (MacKinlay, 1997b). The constant mean return model assumes that a stock's price development will align with its long-term growth. According to the constant mean return model, normal returns are simply the average of the observed returns over a period of time. Both these models use the estimation period to compute the needed parameters, which are (1) the level of correlation and (2) the average observed returns.

Deciding upon a length for the estimation period presents a trade-off. We want the estimation period to be short enough to represent the recent stock price development. However, the estimation period must also be long enough to counteract random spikes in variance (Strong, 1992). We chose to have an estimation period of 250 trading days. The estimation period needs to end a few days before the event window, as we do not want to bias our parameter estimates. In event time, our estimation window can be expressed as [-260, -10], as presented below.

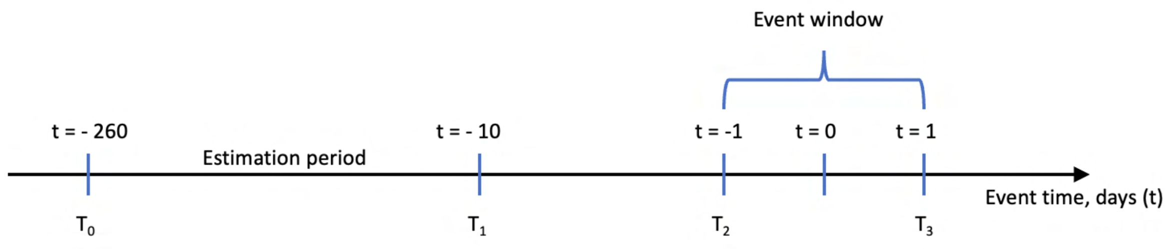

Figure 4.2: This figure presents the estimation period and event window in our study of the 2020 US presidential election as measured in event time. The estimation period starts at T_0 ($t = -260$), and ends at T_1 ($t = -10$). The event window starts at T_2 ($t = -1$) and ends at T_3 ($t = 1$). Ten trading days are left out as to avoid bias in estimation window.

We use the market model to compute expected normal returns in our analysis. There is no census among the researchers in the event study literature regarding the number of factors to use in the market model. Sharpe (1998) argues that more factors can increase the accuracy of coefficients. At the same time MacKinlay (1997b) states that adding more factors does not necessarily decrease expected normal returns variance - thus not adding value to the study. We weigh the arguments and decide to use a single-factor market model. Using this model, we express the expected normal returns for any stock as described in equation 4.3.

$$E(R_{i,t}|X_t) = \alpha_{i,t} \ + \ \beta_i \ * \ R_{m,t} \ + \ \epsilon_{i,t} \tag{4.3}$$

Equation 4.3 expresses the normal return for a stock as generated by the single-factor market model. Alfa and beta, the market parameters, are estimated using ordinary least square (OLS) regression. The stock returns are the dependent variable in the regression, and the market portfolio returns are the independent variable. The alpha is determined as the constant term and the beta as the coefficient of the market portfolio. Moreover, epsilon is the assumed zero mean disturbance term. S&P 500 proxies the market portfolio.

In equation 4.2, we expressed the abnormal returns as the difference between the actual observed returns and the hypothetical normal returns. Using the market model equation and the estimated model parameters, we now estimate the abnormal returns as follows:

$$\widehat{AR_{i,t}} = R_{i,t} \ - \ (\widehat{\alpha_{i,t}} \ + \ \widehat{\beta_i} \ * \ R_{m,t}) \tag{4.4}$$

We are not interested in differences in individual stocks' abnormal returns but rather the difference between the two portfolios' aggregated abnormal returns. We compute the average abnormal returns during the event window for both portfolios, as shown in the equation 4.5. This step supplies us with the average abnormal return for each portfolio on each day of the event window.

$$\overline{AR_t} = \frac{1}{n} * \sum_{i=1}^{n} \widehat{AR_{i,t}} \tag{4.5}$$

The event window in this study is not one day, meaning that we have to aggregate the average abnormal returns for each portfolio throughout the three-day window. We aggregate the abnormal returns from an individual stock level to cumulative average abnormal returns (CAAR) for each portfolio, as follows:

$$\overline{CAR} = \sum_{t\ =\ -1}^{1} \overline{AR_t} \tag{4.6}$$

Following equation 4.6, we have an appropriate measure for comparing the two portfolios of interest, and thus we conclude the event study methodology here. We now move on to the findings and conclusions from the event study. In the following discussion of our findings, we use several statistical tests, described in detail in the appendix A1.

4.2.2 Event Study Findings

PORTFOLIO	NUMBER OF FIRMS	CAAR	CAAR SIGNIFICANTLY LARGER THAN ZERO?
1	115	2.00 %	Yes
2	115	-0.28 %	No

Figure 4.3: A summary of how Portfolios 1 and 2 performed, measured by CAAR, during the three-day event window around the 2020 US presidential election. Portfolios 1 and 2 are made up of firms from the sample of all publicly listed US firms. Portfolio 1 consists of the 115 highest ranked firms, while Portfolio 2 is made up of the 115 lowest ranked firms, measured by E-score. We find that Portfolio 1's CAAR is significantly larger than zero, while Portfolio 2's is not.

We find that portfolio 1, consisting of firms with high E-scores, achieved a CAAR equal to 2.00 % through our three-day event window. Portfolio 2, consisting of firms with low E-scores, had a CAAR of -0.28 %. We perform the cross-sectional t-test for each portfolio's CAAR. Based on results from these tests, we conclude that portfolio 1's CAAR is significantly larger than zero, while portfolio 2 does not have a CAAR significantly larger than zero.

The next step is to compare the CAARs of the two portfolios against each other rather than against zero. First, we perform an f-test with a null hypothesis of equal variance. The f-test estimator we compute is equal to approximately 1.79, with the critical value of a two-tailed f-test converging towards 1.26 when the sample size becomes as large as ours. We conclude that the two portfolios' returns do not exhibit the same variance. Knowing now that the variance of the two portfolio returns is unequal, we use the two-sample one-tailed t-test with the adjustment allowing for unequal variance (appendix A1.3). The computed t-test estimator for the observed CAARs has a value of 7.27, while the critical value is 1.65 at the appropriate degrees of freedom. We conclude that the CAAR of portfolio 1 is significantly larger than portfolio 2's CAAR.

TEST	PURPOSE	NULL AND ALTERNATIVE HYPOTHESIS	CONCLUSION
Two-tailed f-test	Compare portfolio return variances	H0: equal variances HA: unequal variances	Discard the H0 Observations indicate unequal portfolio variances
One-tailed t-test with unequal variance	Compare portfolios' CAARs	H0: equal CAARs HA: portfolio 1 CAAR significantly larger than portfolio 2 CAAR	Discard the H0 Observations indicate that portfolio 1 CAAR is significantly larger than portfolio 2 CAAR

Figure 4.4: A summary of the significance testing conducted on the event study findings. A two-tailed f-test concludes that the difference between Portfolio 1 and 2's variances is significantly non-zero. Following this, a one-tailed t-test concludes that Portfolio 1's CAAR is significantly larger than Portfolio 2's. Note: the p-value in both tests is less than 0.1%.

The observed security prices in our sample indicate beyond any reasonable doubt that portfolio 1 outperformed portfolio 2 in the event window around the 2020 US presidential election. If our textual analysis assessment of firms' environmental sustainability was faulty, one could argue that portfolio 1 and portfolio 2 would be two arbitrary combinations of 115 US stocks. Arbitrary portfolios with 115 stocks are likely diversified enough that their CAAR should be approximately equal and very likely zero during any period. The probability of observing a gap in CAARs as large as the one in our sample, given that our assessment was arbitrary, is the p-value of our t-test. The computed p-value is less than 0.1 % in our t-test, indicating that random stock selection is highly improbable.

In this section, we replicate the study in Ramelli et al. (2021) that found “green” firms outperformed “non-green” firms around the 2020 US presidential election. Our findings align with theirs, as portfolio 1 outperformed portfolio 2 in the event window. This indicates that portfolio 1 consists of firms with a high level of environmental sustainability, and portfolio 2 does not. The practical implication is that investors can use our textual analysis measure of environmental sustainability to identify sustainable firms effectively. In the following section, we demonstrate a practical application of our computed E-scores by looking at the underpricing of newly listed US firms.

4.3 IPO Underpricing in the US

Underpricing, measured as the first-day stock return, is an established phenomenon in the US stock market. Several factors affect the extent of underpricing for a firm, such as the industry sector (Ritter and Scholar, 2021). At the time of writing this thesis, limited research has been conducted on the correlation between firms’ level of environmental sustainability and level of underpricing; however, the findings (Fenili and Raimondo, 2021) indicate that environmental sustainability positively correlates with more significant IPO underpricing. The challenge for investors is to identify the level of environmental sustainability for newly listed firms, as existing providers of ESG-ratings often lack assessments of recent listings. Anecdotally, only a small minority of IPO firms in our sample have an ESG-rating from Bloomberg.

In this section, we aim to demonstrate that investors can use our E-scores to identify and obtain stocks in firms that are environmentally sustainable and that thus exhibit underpricing to a larger extent. We do this by comparing the underpricing of IPO firms in the top five percent of E-scores to the underpricing of the bottom five percent. The first-year returns of these groups are also mapped out and viewed in juxtaposition, yielding an indication of longer-term relative performance.

4.3.1 Underpricing Methodology

First-day abnormal returns

First-day returns are the standard measure of underpricing in IPO research (Ritter and Scholar, 2021). First-day returns can be understood as the percentage difference between the closing price on the first day of a stock's trading day and the IPO issuance price. Ignoring dividend payments, first-day returns can be expressed as shown below:

$$R_{i,1} = \frac{P_{i,1}}{P_{i,IPO}} - 1 \tag{4.7}$$

Raw first-day returns do not factor in systematic market movements and can therefore not be compared across firms that underwent their IPO at different points in time (Fama, 1970). We use the Nasdaq Composite Index to proxy general market movements, computing the market return for any given day t as presented in equation 4.8.

$$R_{m,t} = \frac{P_{m,t+1}}{P_{m,t}} - 1 \tag{4.8}$$

We can now compute the daily abnormal returns for each stock. The abnormal first-day returns are defined as the observed first-day returns for each stock subtracted by the corresponding daily return of the market, here proxied by the Nasdaq Composite Index.

$$AR_{i,1} = R_{i,1} \ - \ R_{m,1} \tag{4.9}$$

We need an aggregated measure of the abnormal first-day returns as we are interested in comparing two portfolios of stocks rather than individual stocks. Portfolio 1 contains the firms in the top five percent highest environmental sustainability scores among the IPO sample, while portfolio 2 contains the firms in the bottom five percentile. We aggregate the abnormal returns for the group with high environmental sustainability scores and those with the group with low scores. We are using equal-weighted portfolios when comparing these groups. Thus, we compute the first-day return for each of these portfolios as shown in equation 4.10:

$$AR_{P,1} = \frac{1}{n_P} \ * \ \sum_{i=1}^{n_P} \ AR_{i,1} \tag{4.10}$$

First-year buy-and-hold abnormal returns ($BHAR_1$)

We use BHAR to measure the two portfolios' performance over a longer investment horizon. In literature, many researchers assume the continuous equal weighting of stocks (Choueifaty et al., 2013). This is not a realistic assumption, as the extraordinarily frequent rebalancing of a portfolio is expensive. Therefore, we compare the portfolios under the assumption of a simple buy-and-hold investment strategy with equal initial weights.

The first step is to compute BHAR for each stock. These are computed as the observed first-year returns of each stock subtracted by the market's return in the same period. We compute BHAR for each stock using the formula presented in equation 4.11.

$$BHAR_{i,1.year} = \frac{P_{i,250}}{P_{i,IPO}} - \frac{P_{m,250}}{P_{m,IPO}} \tag{4.11}$$

The next step is to calculate the BHAR of the two portfolios. We have assumed a buy-and-hold investment strategy where all stocks' initial weights in each portfolio are equal. A portfolio's BHAR is the arithmetic average of each inherent stock's BHAR. We compute the BHAR for each of the two portfolios with the formula presented in equation 4.12.

$$BHAR_{P,1.year} = \frac{1}{n_P} * \sum_{i=1}^{n_P} BHAR_{i,1.year} \tag{4.12}$$

We are also interested in looking at how the portfolios perform during the first year. In other words, we need a measure of the cumulative abnormal returns associated with investing in either of the two portfolios. The idea is to compute how the value of one dollar invested in the portfolios would develop through the year. We do this by computing the portfolio's daily compounded buy-and-hold abnormal return. This return is defined as shown in equation 4.13.

$$CBHAR_{P,t} = -1 + \prod_{i=IPO}^{t} \left(1 + \frac{1}{n_P} * \sum_{i=1}^{n_P} BHAR_{i,t}\right) \tag{4.13}$$

4.3.2 Underpricing Findings

We look at the degree of underpricing observed in IPOs during the last three years. For this study, we use the dataset of IPOs described in section 3.1.1. Moreover, we link underpricing to environmental sustainability through the computed E-scores. Portfolio 1, presented in table 4.4, contains the firms placed among the top five percent in terms of E-scores. Likewise, portfolio 2, presented in table 4.5, contains the firms in the bottom five percentile of scores. By comparing portfolios 1 and 2, we highlight the difference in the level of underpricing contingent on environmental sustainability, as measured by our computed E-scores.

PORTFOLIO	NUMBER OF FIRMS	FIRST-DAY AR	FIRST-YEAR BHAR
1	18	1.44 %	- 8.30 %
2	18	- 0.73 %	- 33.81 %

Figure 4.5: This figure presents a summary of the underpricing exhibited by portfolios 1 and 2. Portfolio 1 consists of the 18 firms with the highest E-scores from the IPO sample, and Portfolio 2 consists of the 18 firms with the lowest E-scores. Portfolio 1 has a first-day abnormal return of 1.44 %, compared to the -0.73% of Portfolio 2. Environmentally sustainable firms are more underpriced compared to unsustainable firms in the US between 2019 and 2021.

The first-day abnormal return of portfolio 1 is 1.44 % versus - 0.73 % for portfolio 2. The first-day abnormal return for each portfolio is calculated as the mean of all first-day abnormal returns of individual stocks in each portfolio. This means that, on average, the first-day returns are larger by 2.2 % for the IPO firms in portfolio 1 compared to the firms in portfolio 2. The first-year abnormal return of portfolio 1 is - 8.30 % versus - 33.81 % for portfolio 2. We note that both portfolios perform poorly with a buy-and-hold strategy with a longer horizon. However, portfolio 1 still fares better than portfolio 2, as there is a staggering 25.5 % difference in first-year BHAR.

The financial data and the findings discussed show that the observed average underwriting in a group of firms is correlated with the level of environmental sustainability scores achieved by the firms in question. The average first-day increase in stock prices for the firms in portfolio 1 is larger than those in portfolio 2. This translates to firms in portfolio 1, i.e., firms with higher environmental sustainability scores, exhibiting harsher underwriting. The first-year abnormal buy-and-hold return is not necessarily an effect of underwriting. The much larger investment horizon allows many unrelated factors to influence the portfolio returns. However, it is interesting that portfolio 1 outperforms portfolio 2 with a large margin even with this longer investment horizon. We argue that firms with higher levels of environmental sustainability will perform better on the stock market due to a shifting sentiment in the global economy towards sustainability (Randall, 2020).

Figure 4.6: The development of $ 1 invested in portfolios 1 and 2 for one year. Portfolio 1 outperforms portfolio 2 significantly, indicating that firms with high E-scores on average yield larger first-year returns.

One dollar invested into each of the portfolios 1 and 2 develops throughout the first year as shown in the figure 4.6. The difference in the two portfolios' cumulative buy-and-hold abnormal returns is maximised around 116 trading days post-IPO. There is also a very clear trend of portfolio 1 outperforming portfolio 2.

The data sample strongly indicates a significant correlation between the level of underwriting and the level of environmental sustainability. We cannot be confident that our computed E-scores have a causal effect on observed underpricing, but the correlation is clear-cut. The findings presented in this section substantiate our claim that the computed E-scores can be used as a profitable tool for sustainable investors.

5 Conclusion

In this thesis, we propose a textual analysis tool for assessing the environmental sustainability of firms listed on US stock exchanges. The tool is a more versatile alternative to established sustainability measures, such as Bloomberg's ESG rating. The most important contribution of the proposed assessment method is that it facilitates the assessment of recently listed firms. We demonstrate this fact by providing E-scores for IPOs in the US during the sample period between 2019 and 2021. In addition, we prove that our textual analysis assessment is viable for a much larger sample as we also provide E-scores for all publicly listed US stocks.

Using an event study of the 2020 US presidential election, we are able to verify our environmental sustainability assessment indirectly. We successfully identified a group of environmentally sustainable firms that outperformed both the general stock market and other unsustainable firms during the three-day event window. The event study replicates Ramelli et al. (2021), in which “green” firms are compared to “non-green” firms using ESG ratings and abnormal returns. We construct our “green” and “non-green” portfolios using E-scores, and we find similar abnormal returns for these two portfolios as seen in Ramelli et al. (2021). Moreover, we conclude through significance testing that our “green” portfolio outperforms the “non-green” portfolio. This indicates that our E-scores can be used to identify environmentally sustainable firms.

There is little established literature on the relation between environmental sustainability and IPO underpricing due to very few IPO firms having ESG ratings. However, Loughran and Ritter (1995) finds that different industry sectors exhibit different levels of underpricing. Our findings indicate that the effect of being environmentally sustainable is positively correlated with the level of underpricing across sectors. We find that firms with a high level of environmental sustainability had first-day abnormal returns of 1.44 percent, compared to firms with low environmental sustainability having first-day abnormal returns of -0.73 percent from 2019 to 2021 in the US. Our findings add to the existing literature by providing insight into how environmental sustainability can affect IPO underpricing across sectors and longer-term financial performance.

6 Limitations and Further Research

An important limitation of the methodology in this thesis is that it builds on the assumption that firms accurately present their activities in their Section 1: Business filed with the SEC. However, this concern is addressed by SEC's regulations: "The company writes the 10-K and files it with the SEC. Laws and regulations prohibit companies from making materially false or misleading statements in their 10-Ks" (SEC, 2022). Nevertheless, the proposed method will only be able to provide environmental sustainability scores for firms that have filed 10Ks with the SEC.

Another limitation of our environmental sustainability assessment is that it is only indirectly verified through the event study. Firstly, we assume that the research by Ramelli et al. (2021) is accurate and that "green" firms indeed surged during the 2020 US presidential election. However, several researchers agree with this finding (Dubois et al., 2020). Secondly, there is the possibility of a type-1 error, meaning that our event study findings are purely random.

The textual analysis results indicate that an assessment of firms' environmental sustainability is possible through textual analysis. Ideally, we would attempt to refine our textual analysis methodology further, but that is outside the scope of this thesis. Suggested further research on the textual analysis front would be to eliminate bias and further the objectivity of the assessment tool regarding the choice of the reference group. Other textual analysis methods could be combined with the cosine similarity measure we use in this thesis, such as the smooth inverse frequency method.

References

Amihud, Y. and Mendelson, H. (1986). Liquidity and stock returns. *Financial Analysts Journal*, 42(3):43–48.

Ardia, D., Bluteau, K., Boudt, K., and Inghelbrecht, K. (2020). Climate change concerns and the performance of green versus brown stocks. *National Bank of Belgium, Working Paper Research*, 1(395).

Armitage, S. (1995). Event study methods and evidence on their performance. *Journal of economic surveys*, 9(1):25–52.

Benaissa Pedriza, S. (2021). Sources, channels and strategies of disinformation in the 2020 us election: Social networks, traditional media and political candidates. *Journalism and Media*, 2(4):605–624.

Berg, F., Koelbel, J. F., and Rigobon, R. (2019). *Aggregate confusion: The divergence of ESG ratings*. MIT Sloan School of Management Cambridge, MA, USA.

Choueifaty, Y., Froidure, T., and Reynier, J. (2013). Properties of the most diversified portfolio. *Journal of investment strategies*, 2(2):49–70.

Colianni, S., Rosales, S., and Signorotti, M. (2015). Algorithmic trading of cryptocurrency based on twitter sentiment analysis. *CS229 Project*, 1(5).

Dann, L. Y., Mayers, D., and Raab Jr, R. J. (1977). Trading rules, large blocks and the speed of price adjustment. *Journal of Financial Economics*, 4(1):3–22.

Dimson, E., Marsh, P., and Staunton, M. (2020). Divergent esg ratings. *Journal of Portfolio Management*, 47(1):75–87.

Dubois, H., Heath, A., and Edelman, J. F. (2020). Esg & 2020 us. presidential election. *Harward Law School Forum on Corporate Governance*.

Eccles, R. G., Kastrapeli, M. D., and Potter, S. J. (2017). How to integrate esg into investment decision-making: Results of a global survey of institutional investors. *Journal of Applied Corporate Finance*, 29(4):125–133.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417.

Fenili, A. and Raimondo, C. (2021). Esg and the pricing of ipos: Does sustainability matter. *Available at SSRN 3860138*.

Freiberg, D., Rogers, J., and Serafeim, G. (2020). How esg issues become financially material to corporations and their investors. *Harvard Business School Accounting & Management Unit Working Paper*, 1(20-056).

Friede, G., Busch, T., and Bassen, A. (2015). Esg and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 5(4):210–233.

Furlow, N. E. (2010). Greenwashing in the new millennium. *The Journal of Applied Business and Economics*, 10(6):22.

Gillan, S. L., Koch, A., and Starks, L. T. (2021). Firms and social responsibility: A review of esg and csr research in corporate finance. *Journal of Corporate Finance*, 66:101889.

Greco, F. and Polli, A. (2020). Emotional text mining: Customer profiling in brand management. *International Journal of Information Management*, 51:101934.

Hassan, O. A. and Romilly, P. (2018). Relations between corporate economic performance, environmental disclosure and greenhouse gas emissions: New insights. *Business Strategy and the Environment*, 27(7):893–909.

Hoeffding, W. (1948). A non-parametric test of independence. *The annals of mathematical statistics*, pages 546–557.

Kotsantonis, S. and Serafeim, G. (2019). Four things no one will tell you about esg data. *Journal of Applied Corporate Finance*, 31(2):50–58.

Lang, M. and Stice-Lawrence, L. (2015). Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics*, 60(2-3):110–135.

Li, F. and Polychronopoulos, A. (2020). What a difference an esg ratings provider makes. *Research Affiliates. https://www. researchaffiliates. com/documents/770-what-a-difference-an-esg-ratings-provider-makes. pdf.*

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of finance*, 66(1):35–65.

Loughran, T. and Ritter, J. R. (1995). The new issues puzzle. *The Journal of finance*, 50(1):23–51.

MacKinlay, A. C. (1997a). Event studies in economics and finance. *Journal of economic literature*, 35(1):13–39.

MacKinlay, A. C. (1997b). Event studies in economics and finance. *Journal of economic literature*, 35(1):13–39.

McKee, A. (2003). *Textual analysis: A beginner s guide*. Sage.

MSCI (2021). Prospectus: Invesco MSCI Sustainable Future ETF. https://connect.rightprospectus.com/Invesco/TADF/46137V407/P, accessed = 2022-04-20.

Norton, L. (2022). Larry Fink: Sustainable Investing Is About Profits, Not Taking a Stand. https://www.morningstar.com/articles/1075068/larry-fink-sustainable-investing-is-about-profits-not-taking-a-stand, accessed = 2022-02-12.

Pagan, A. R. and Schwert, G. W. (1990). Testing for covariance stationarity in stock market data. *Economics letters*, 33(2):165–170.

Rahutomo, F., Kitasuka, T., and Aritsugi, M. (2012). Semantic cosine similarity. In *The 7th International Student Conference on Advanced Science and Technology ICAST*, volume 4, page 1.

Ramelli, S., Wagner, A. F., Zeckhauser, R. J., and Ziegler, A. (2021). Investor rewards to climate responsibility: Stock-price responses to the opposite shocks of the 2016 and 2020 us elections. *The Review of Corporate Finance Studies*, 10(4):748–787.

Randall, D. (2020). Analysis: Green is the color of money for funds betting on a Biden win. https://www.reuters.com/article/us-usa-election-funds-analysis-idUKKBN276190, accessed = 2022-04-28.

Ritter, J. R. and Scholar, C. E. (2021). Initial public offerings: Updated statistics on long-run returns.

Roiger, R. J. (2017). *Data mining: a tutorial-based primer.* Chapman and Hall/CRC.

SEC (2022). Investor bulletin: How to read a 10-K - SEC. https://www.sec.gov/files/reada10k.pdf, accessed = 2022-02-12.

Sharpe, W. F. (1998). The sharpe ratio. *Streetwise–the Best of the Journal of Portfolio Management*, pages 169–185.

Strong, N. (1992). Modelling abnormal returns: A review article. *Journal of Business Finance & Accounting*, 19(4):533–553.

Trinks, A., Mulder, M., and Scholtens, B. (2020). An efficiency perspective on carbon emissions and financial performance. *Ecological Economics*, 175:106632.

Wei, C.-H., Leaman, R., and Lu, Z. (2016). Beyond accuracy: creating interoperable and scalable text-mining web services. *Bioinformatics*, 32(12):1907–1910.

Xia, P., Zhang, L., and Li, F. (2015). Learning similarity with cosine similarity ensemble. *Information Sciences*, 307:39–52.

Appendix

A1 Significance Testing

To determine whether the firms with high (low) environmental sustainability scores have different returns, we first investigate whether the CAARs for these two portfolios are significantly larger than zero. Thereafter, we test whether the two portfolios have equal CAARs. These tests will allow us to conclude whether **(I) the two portfolios are significantly positively affected by the 2020 election result** and whether **(II) the degree to which the portfolios are affected differs**.

A1.1 (I) Cross sectional T-test

A cross-sectional t-test is a parametric test (Hoeffding, 1948). The underlying assumption of parametric tests is that the sample data, the CAARs for each portfolio, can be accurately described using distribution parameters. Distribution parameters are mean and variation, for example. Some researchers argue that non-parametric tests may be better than parametric tests when working with returns data. However, data shows that excess returns are normally distributed to a high degree, meaning that they have clear parametric tendencies. CAAR is a form of excess return, and the research (Pagan and Schwert, 1990) encourages the use of parametric tests on this returns measure. We use the cross-sectional t-test to determine whether the CAARs for the two portfolios are significantly larger than zero.

Since we decided to use a parametric t-test, our CAARs are assumed to be distributed with an expected value of 0 and a non-zero variance. In addition, the parametric t-test assumes that there is no correlation between the CAARs for the two groups. The research of Strong (1992) indicates that the assumption of independent samples is valid when using abnormal returns, in this case, measured by CAARs.

The residual variance for each stock in the two portfolios is not known. We need to estimate the variance of each stock before aggregating individual stock variance to the portfolio level. Per MacKinlay (1997b), we use the sample variance of each stock supplied by the market model during the event window to estimate the residual variance. The variance of the CAARs can then be expressed mathematically as the sum of the sample adjusted error term of each stock's cumulative abnormal return, as shown in equation A1.1.

$$S^2_{\overline{CAR}} = \frac{1}{n-1} * \sum_{i=1}^{n} \left(CAR_i \; - \; \overline{CAR} \right)^2 \tag{A1.1}$$

We now have a cumulative returns measure, CAAR, and a suitable estimate of this measure's variance. A t-statistic is a suitable test estimator for evaluating our null hypothesis of non-significant CAARs (Strong, 1992). The t-statistic we use is defined as follows:

$$t = \sqrt{n} * \frac{\overline{CAR_i}}{\sqrt{S^2_{\overline{CAR}}}} \tag{A1.2}$$

The t-statistic, equation A1.2 is calculated for each portfolio. The absolute critical value of a one-tailed t-test with a five percent significance level is retrieved from the t-distribution with degrees of freedom equal to the number of observations subtracted one. If we compute a t-statistic with a greater absolute value than the critical value, we must dismiss the null hypothesis of non-significant CAARs. In that case, our financial data will indicate that the true CAAR is larger than zero for the portfolio in question.

A1.2 (II) F-test for variance

Comparing the means of two groups is perhaps the most common statistical test in any field of research. In our thesis, we are interested in comparing the CAARs for the two defined portfolios. Comparing means is empirically done using two-sample t-tests. A two-sample t-test assumes that the samples' means are normally distributed and their variances are independent. The test estimator used in a two-sample t-test is dependent on the relative variance of the two portfolios. Therefore, we first test whether the two portfolios have equal variances through a two-tailed f-test.

A two-tailed f-test for equality of two variances is used to test whether two portfolios' variances are equal, which is the null hypothesis. The two-tailed f-test tests against the alternative hypothesis that the two variances are unequal in either direction. The test estimator is called F and is one portfolio's sample variance divided by the other's. The further this estimator is from 1, the larger the difference in observed variance. Once the discussed F-test is performed, we will know whether the two portfolios have statistically equal variances or not, allowing us to continue with the comparison of CAARs.

$$F = \frac{S_1^2}{S_2^2} \tag{A1.3}$$

A1.3 (II) T-test for comparing CAARs

The point of the t-test is to check whether the difference between the two groups' means is significantly different. In our case, we want to test whether the CAAR of portfolio 1 is significantly larger than the CAAR of portfolio 2. The difference in CAARs for the two portfolios is adjusted for the two portfolios' variances. Based on the performed f-test, we assume either equal variance and pool the variances or assume unequal variance. The rest of this section describes both variations of the two-sample t-test.

$$T = \frac{\overline{CAR_1} - \overline{CAR_2}}{\sqrt{\frac{S_1^2}{n_1} - \frac{S_2^2}{n_2}}} \tag{A1.4}$$

$$df = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}} \tag{A1.5}$$

The unequal variance two-sample t-test estimator does not adjust the difference in CAARs on a pooled variance but instead adjusts the difference for the two portfolios' variances (Hoeffding, 1948). The estimator used in this unequal variance two-sample t-test is t-distributed with degrees of freedom computed by equation A1.5. We are performing a one-tailed test with a five percent significance level; thus, the critical value can be retrieved from the t-distribution at the calculated degrees of freedom at the appropriate significance level. If the absolute value of the test estimator shown above exceeds this critical value, we dismiss the null hypothesis of equal CAARs. This would indicate that the CAAR of portfolio 1 is significantly larger than the CAAR of portfolio 2.