NHH

Transparency and its Relation to Sustainability Performance

A textual mining approach to sustainability performance of mutual funds and disclosure in underlying companies

Dag Håvard G. Bergsland

Thomas Hansen

Supervisor: Nataliya Gerasimova

Master thesis in Finance

NORWEGIAN SCHOOL OF ECONOMICS

Abstract

Over the last decade, sustainability has truly moved from niche to mainstream when it comes to attracting attention among investors and policymakers. The number of funds investing under environmental, social and governance considerations has surged, and with that, the call for transparency is stronger than ever. Corporate statements and reports are therefore a valuable resource as they represent a wealth of information regarding companies' operations. There are two main purposes of this thesis: The first is to create a tool that captures ESG-related disclosures in annual 10-K reports of underlying companies in mutual funds. Secondly, to see if disclosure relates to sustainability performance, represented by the score in the Morningstar sustainability rating (MSR). The sample consists of 118 US mutual funds, observed over a three-year timeframe, from 2016 to 2018. The first research question examines if the level of disclosure in underlying companies can predict sustainability performance of funds. Our results indicate that there is a relationship between the level of disclosure in underlying firms and sustainability performance for the following investment categories: US large cap blend, US large cap growth, US large cap value, US mid cap, and finance. For sector-specific categories such as healthcare, consumer goods and services and technology, no significant relationship is found. The explanatory power of textual disclosure score on sustainability performance of funds is limited but the model shows potential for more precise predictions for certain investment categories. Estimates appear to be less accurate for more volatile funds for which the difference between MSR and ESG disclosure score is larger. We also find that "green labelled" funds in our sample have better sustainability performance than conventional funds, while we find no difference in the disclosure score. Lastly, despite the increasing amount of sustainable investing, our data does not suggest an increasing trend of ESG-disclosures in 10-K filings over the sample period.

Preface

This Master thesis is written as part of the Finance master's program at the Norwegian School of Economics (NHH).

The paper uses textual data analysis to examine whether the ESG-related statements companies make in their annual reports align with their performance in the Morningstar Sustainability Rating. The choice of topic emerged as a result of our mutual interests for finance, sustainability and programming.

The process has been challenging and time-consuming, especially creating a code that successfully retrieves, structures and analyses the data. With a wish to create a tool complex enough to be considered a real contribution to the field of research, we needed to balance this with the timeframe and limitations to our skills in efficient programming. Retrieving and processing our data in the right way became just as important as analysing it afterwards.

We would like to acknowledge our advisor, Nataliya Gerasimova, who has provided us with helpful feedback and consultation during this academic work. We also want to thank Mattias Ekstrand from Morningstar for giving us important insight regarding the Morningstar Sustainability Rating and Morningstar Direct.

Dag Håvard G. Bergsland

Thomas Hansen

Norwegian School of Economics

Bergen, June 2019

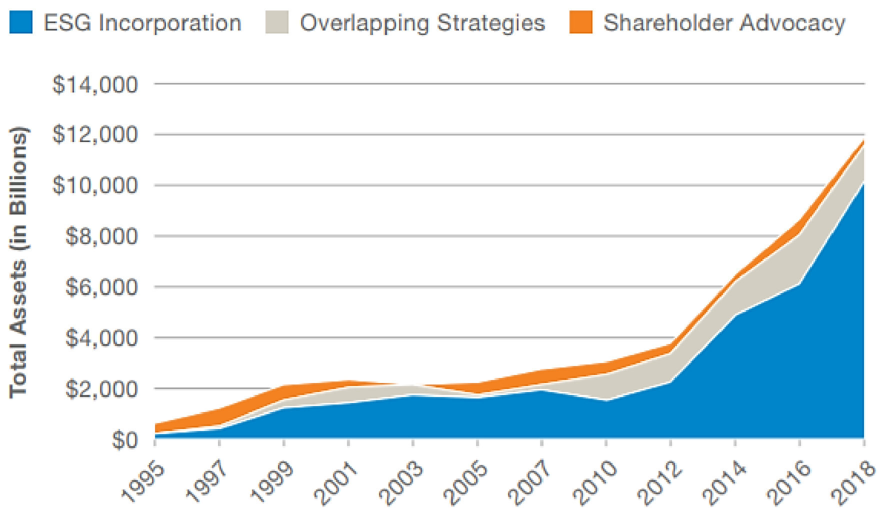
Contents

1. Introduction

In recent years, sustainability topics have received increased attention among investors and policymakers, which in turn pressure companies to integrate environmental, social and governance (ESG) criteria in their operations (Mooney, 2018). The trend is underpinned by a US study from 2017, indicating that three-quarters of all investors and 86% of Millennials are interested in sustainable investing. Furthermore, 38% reported that sustainability had already been a factor in their investment decisions (Morgan Stanley, 2017). While socially responsible funds and related performance studies have been around for decades, it is not until the recent years SRI and implementation of ESG criteria have truly moved from niche to mainstream (Hamilton, Jo & Statman, 1993; J.P. Morgan, 2018).

In terms of assets under management, Europe has traditionally been the prevailing actor in this field, followed by the fast-growing US market (Global Sustainable Investment Alliance, 2016). In the US, assets under management using SRI strategies grew 38%, to a total of $12.0 trillion, from 2016-2018. This represents 26% of the total US assets under professional management (US SIF Foundation, 2018). As indicated in Figure 1, many of these assets applied various ESG criteria in investment considerations. Numerous ESG ratings have emerged in response to this massive rise in sustainability awareness among investors – with perhaps the most prominent one being Morningstar and their widely known “globes”.

Figure 1 - Sustainable and Responsible Investing in the US



(US SIF Foundation, 2018)

Existing SRI-related research has often been designed to reveal potential costs or gains of sustainable versus conventional investing. The theoretical framework of modern portfolio theory implies that the isolated effect of imposing constraints, such as screening processes, to an investment portfolio would increase the associated idiosyncratic risk (Humphrey & Tan, 2014). However, the results from empirical studies of SRI funds' performance are ambiguous and inconclusive – as highlighted in the following literature review section.

This thesis makes use of textual data analysis to retrieve and analyse 10-K reports of publicly traded US-based companies. We use this data to create our own "textual-disclosure" measure for a sample of US mutual funds. Next, we use the textual disclosure score to examine if the level of disclosure can predict the level of sustainability, represented by the Morningstar Sustainability Rating. The intention is to examine if funds with more transparent underlying companies exhibit superior ESG-performance versus less transparent counterparts. Additionally, we test whether typical "green labelled" funds perform better than conventional funds in the mentioned aspects. Another objective of this thesis is to reveal whether the stakeholders' increased pressure has resulted in more transparency regarding companies' sustainability activities. More specifically, this is to be examined by reviewing the disclosure trend in our sample funds over a three-year time frame.

This analysis contributes to the literature by merging textual mining methods with sustainability research to create a tool for analysing funds. This tool builds on several concepts in earlier literature and brings them together to compute a quantifiable measure meant to reflect the level of disclosure in annual reports. Our results indicate that there is a relationship between the level of disclosure in underlying firms and sustainability performance in the following investment categories; US large cap blend, US large cap growth, US large cap value, US mid cap and finance. In sector-specific categories such as healthcare, consumer goods and services and technology, no significant relationship is found. The textual disclosure score cannot be regarded as a precise predictor of sustainability performance of funds but shows potential in certain investment categories. We also find that "green labelled" funds in our sample have better sustainability performance than conventional funds, while no difference is found in the disclosure score. For further research, the tool can be customized, e.g. by using a more sector-specific dictionary.

2. Hypothesis Development

The following section introduces the concepts that form the background for the research questions that are to be examined in this thesis. The code in the software R is altered according to each research question. The general methodology of retrieving and processing data and the alterations are described in detail in the Methodology-section.

2.1 Socially Responsible Investing

Socially responsible investing (SRI) is a generic term that describes an investment approach that aims to merge societal and financial gain. SRI adds a set of social, environmental, ethical and governance factors, as well as financial return, to the investment decision-making process (Louche & Hebb, 2014). The modern sense of the term SRI has its origin from the US during the 1960s when investors ceased to invest in stocks that opposed their standards. This equity boycott is an example of a negative screening process, which was often the standard in the early days of SRI (Matloff & Chaillou, 2013). Since then, the field of SRI has evolved into a complex universe, scattered with acronyms that describe the multitude of diverse interpretations on the subject. In fact, EuroSIF (n. d.) defines seven different strategies to sustainable investing, which are all collectively referred to as SRI-strategies**.** Among them are positive screening and shareholder action. *In this study, sustainable companies are defined as companies that aims to create long-term value for shareholders while managing ESG-risks and opportunities in an effective matter. Terms as SRI-investing, sustainable investing, responsible investing and ESG-integration are used interchangeably to describe investors efforts to implement sustainability into their investment considerations.*

2.2 Sustainability Metrics and Transparency

Different priorities among investors, combined with the complexity of measuring a company's overall sustainability, means there is no absolute truth in how to measure performance in this regard. There exists a complexity in measuring sustainability in an objective and meaningful way, partly due to the lack of universal agreement on relevant sustainability criteria, and how to measure these criteria. Consequently, ESG fund-ratings from competing research providers are inconsistent, with a relatively low correlation – as opposed to the case of financial ratings (Wigglesworth, 2018). In short, quantification of sustainability is a highly sophisticated process and it is unlikely that existing ratings can tell the whole truth about a fund's underlying impact.

Nevertheless, leading investment research firms such as Morningstar and MSCI (formerly Morgan Stanley Capital International) have launched fund sustainability ratings to facilitate the evaluation of investment objects exposure to sustainability risks. These ratings use a set of environmental, social, and governance (ESG) criteria as performance indicators (Morningstar, 2018a; MSCI, 2016). It is worth mentioning that although the ratings are featured as "sustainability ratings", they only consider a limited scope of ESG related risks. While such ratings have been subject to criticism, they have undoubtedly gained traction among an increasing pool of investors concerned about risks linked to sustainability (Thompson, 2019). These investors often search beyond financial statements for a more complete picture of a company's ability to create value over the long term (SASB, 2016).

A pre-condition for assessing sustainability practices is transparency. Transparency and accountability advocate ethical behaviour among economic actors (Zsolnai, 2012). One definition of transparency is "*timely and reliable economic, social and political information, which is accessible to all relevant stakeholders*" (Kolstad & Wiig, 2009). Due to the lack of regulatory frameworks, there is also great divergence when it comes to companies' transparency and their approach to ESG reporting. While regulatory forces have driven European companies to greater transparency through the EU Non-Financial Reporting Directive (implemented from 2018), the regulatory demands for transparency have been more modest in the US. Despite this, 92% of companies listed on the S&P 500 offered some accessible sustainability information while 78% issued sustainability reports (IRRCI, 2018).

Pressure from consumers and employees can materialize in more transparent reporting. Existing literature suggests that companies exposed to pressure from stakeholders are more likely to present reports with a higher degree of transparency (Fernandez-Feijoo, Romero & Ruiz, 2014). This characteristic relates to the legitimacy theory (LT), which CSR literature refers to as a potential driver of voluntary disclosure. (Cuganesan, Ward & Guthrie, 2007).

The first research question in this thesis will address the link between sustainability metrics and transparency on a fund level. The purpose is to identify a possible relationship between these two ways of assessing sustainability in an investment context. More specifically, our research question is:

Are there more ESG-disclosures in the underlying companies of sustainable funds, and is a measure of ESG-disclosures in underlying companies a possible predictor for sustainability performance of a fund? Are there any differences across investment categories?

Our intuition regarding this link is that sustainable funds are likely to hold more transparent companies, relatively to their less sustainable peers. However, we find no research supporting this view. One study that compares individual companies' sustainability rating (Newsweek Green Ranking) with the frequency of sustainability words in annual reports finds no significant correlation (Wen, 2014). Another study performed on Australian mining and production companies finds a positive correlation between the level of emissions and environmental disclosures (Clarkson, Overell & Chapple, 2011). Note that the latter study only evaluates emissions. Consequently, no other ESG-related aspects are considered.

2.3 Difference in Scores

As an extension to the previous research question, we want to explore if some of the difference in the textual disclosure score and the Morningstar ESG score can be explained by fund characteristics such as fund size (million USD), age (days since inception), and the number of holdings. Additionally, we include some risk factors based on the Fama & French (1993) three-factor model: Beta as a measure of volatility for each fund, growth versus value style investing, and small size versus big size investing. We ask the question: *How does the difference (represented by delta) between textual disclosure score and sustainability score vary according to funds' characteristics?*

The purpose of this research question is to gain a better understanding of what might cause the difference observed between the two scores. By computing the absolute difference between the scores, we can examine this variation while ignoring the direction of it. We do, for instance, expect that funds with more holdings will reduce some of the variation in both the textual ESG score and MSR by diversifying and thus have a lower score difference.

2.4 Green Labelled Funds

One key characteristic of the mentioned sustainability ratings is that they do not only assess funds with a stated SRI mandate, so-called "green labelled" or "green billed" funds. The methodology of these ratings facilitates identification of conventional funds with similar, or even higher, ESG performance than funds with the "SRI-label". This means investors can tilt their portfolio towards more sustainable funds in any given fund category, instead of choosing from the relatively limited pool of SRI-funds. Despite the flaws of these ratings, one could argue that these tools can be useful to complement traditional financial analysis, while also help "mainstreaming" SRI considerations among conventional investors.

The concept of "green labelled" funds is, however, still an interesting characteristic of a fund. A fund's name is the most rudimentary information about a fund, and most likely the first thing a potential investor evaluates. Therefore, the name of a fund often projects some intrinsic information about the applied investment strategy. From this assumption we define the next research question: *Do "green labelled" funds exhibit superior performance in the MSR and/or our disclosure metric?*

The intention behind this research question is to examine if the funds that appear sustainable by name are in fact more sustainable than the rest of the sample. The non-green-labelled funds are from now referred to as *conventional funds* in this context. We expect to find a statistically significant difference in sustainability score between "green labelled" and conventional funds. This outcome would be in accordance with previous research that compared the environmental impacts of conventional and sustainable investment funds (Koellner, Suh, Weber, Moser & W. Scholz, 2008). Moreover, if sustainable funds tend to score better on the textual disclosure score, we also expect a difference between "green labelled" and conventional funds in the disclosure score.

2.5 Time Trend

The final research question derives from a corporate responsibility survey stating there has been a growth in sustainability reporting in the US which is partly driven by investor and shareholder interest in sustainability, forcing companies who have not previously reported to start practising this kind of disclosure (Blasco & King, 2017). We aim to answer whether this trend can be identified in the textual disclosure metric created in this thesis on a fund level in the past three years. In other words: *Has there been an uptake of ESG-disclosures in 10-Ks in recent years?*

Considering the increased interest in responsible investing, we expect to see a rise in the level of disclosure during our sample period as companies adapt to investors and shareholders desire for transparency. A survey of CR reporting from KPMG finds that the 81% of the 100 largest US companies integrated CR information into their financial reports in 2017 – a significant increase from 30% in 2015 (Blasco & King, 2017).

3. Literature Review

Previous research on the topic of SRI, socially responsible funds, ESG investing, and other alternatives to integration of ESG criteria in the investment process is mainly focused on the financial performance of such investments compared to conventional investments. The results of existing literature are mixed. Some find significant underperformance (Ciciretti, Dalò & Dam, 2017), and others overperformance (Friede, Busch & Bassen, 2015; Henke, 2016; Eccles 2014).

The methodology of each study varies in the geographic market selection, time-period, and asset classes (Morningstar, 2016a). They also vary in how sustainable investments are selected and classified as sustainable. One method being used in studies for classifying sustainable investments is the use of sustainability ratings such as the Morningstar Sustainability Rating (Dolvin, Fulkerson & Krukover, 2017) or MSCI ESG fund metrics (Breedt, Ciliberti, Gauldi & Seager, 2018). Though these ratings are becoming more widely applied in research and in the market, questions are being raised concerning the quality and consistency of such metrics (van Steenis, 2019; Allen, 2018). As these ratings are adopted among investors, they also influence capital allocation. However, when large inconsistency between ratings are found, questions about the usefulness of such ratings to investors are being raised (Thompson, 2019). The cause of this inconsistency is that a single ESG score is meant to represent a wide variety of different aspects. When rating companies use their own unique methodologies, metrics, weightings, and definitions of what constitutes ESG (Doyle, 2018), inconsistency is bound to occur.

Though measures such as greenhouse gasses emissions and water usage are easily quantifiable, the effect of certain environmental programs or human rights and anti-corruption policies are harder to measure and compare between companies. As Chvatalová, Kocmanová & Dočekalová (2011) mention in their paper on corporate sustainability reporting:

> *"To be comparable across all companies, and thus useful for mainstream investment analyses, it is important that economic, environmental, social and governance data is transformed into consistent units and is presented in a balanced and coherent manner in ESG indicators" (2011, p. 246).*

Organizations such as the Global Reporting Initiative (GRI) works towards implementing and guiding businesses and governments towards standardization in ESG reporting. However, these standards are often voluntarily implemented by companies and to what degree the standards are followed and interpreted differs (Chvatalová, et al., 2011). Since there are no universal standards to ESG reporting, and the measures that are easily quantifiable does not capture the whole extent of ESG performance, analysts are drawn towards largely unstructured data on companies' approach to sustainability in their operations.

Currently, the most substantial source of data related to sustainability measures and performance is through disclosures in reports published by the company itself. The number of corporations reporting sustainability information has been growing rapidly in the last two decades (Eccles, Krzus, Rogers & Serafeim, 2012). This growth in sustainability disclosure can be partly attributed to the legitimacy theory (LT).

LT is derived from the concept of organisational legitimacy and can be described as an organisation's continuous seek to ensure that they operate within the bounds and norms of their respective societies. It is based on the idea that there exists a social contract between a company and the society in which it operates. Thus, if the company violates this social contract by not operating in a legitimate manner, society will withdraw the contract. This can materialize in less demand for the product, higher financing costs, or higher taxes (Cuganesan et al., 2007). As a result of this, managers in general search for an alignment between corporations' activities and prevailing public values and views – or at least to create an appearance of such. Research has also suggested that some managers view voluntary disclosure of environmental information as a measure to "head-off" public pressure (O'Donovan, 2000). Another study finds evidence implying that companies with high consumer visibility and political risk are more likely to disclose CSR information. Moreover, the same companies are likely to employ more disclosure strategies aimed at shifting public perceptions and expectations, and deflecting attention (Cuganesan et al., 2007.)

Existing literature state that greater transparency in sustainability issues can add value to companies by improving their reputation, motivating employees, enabling differentiation and reduce the risk of negative publicity. A study by Morgan Stanley (2017) finds that millennials are two times more likely to purchase items from, or even invest in, companies that are perceived as sustainable. The same study indicates that millennials were three times more likely to have sought employment with a sustainability-minded company. In other words, being perceived as sustainable and transparent should be an important concern for any firm. On the contrary, obtaining and communicating sustainability data to stakeholders can be a costly and time-consuming process for corporations - especially due to the stated complications of measuring ESG-performance (Dubbink, Graafland, & van Liedekerke, 2008). Although modest pressure from policymakers can stimulate transparent reporting methods, this is not given. Unnecessary rigid and demanding regulations could also feed an attitude of minimal compliance and distrust (Dubbink et al., 2008). From the stakeholders' perspective, the lack of external assurance in voluntary sustainability reports is considered as a key concern (IRRCI, 2018). Another inadequacy is that governance risk is often better reported than environmental and social risk, which in turn creates an imbalance. As investors are increasing their reliance on ESG factors, studies emphasis investors' demand for more uniformed standards in the reporting of such metrics (Nelson, 2019).

The surge of disclosures in sustainability reports and annual reports have made the task of manually reading and analysing this information across industries and markets an overwhelming task (Horuckova & Baudasse, 2017; Ching, Gerab & Toste, 2014). In addition to being cost-intensive and time-consuming, human-based methods are affected by the intuition of each individual researcher and can consequently be prone to biases (Van den Bogaerd & Aerts, 2011). To address these issues textual analysis methods are becoming widely utilized and acknowledge research techniques (Aureli, 2017). With easier access to more computational power and the right tools, these techniques can process data faster and to a better degree mimic the advantages of traditional techniques. Most importantly, they can be applied to convert the increasing amount of unstructured textual ESG information and sources into quantitative measures (Loughran & McDonald, 2011; Horuckova & Baudasse, 2017).

The most commonly utilized textual analysis methods include text mining and content analysis. Though content analysis and text mining are inherently different methods, they share some fundamental similarities, and both are applied in numerous studies that analyse trends, patterns, content and sentiment in sustainability reports, annual reports and media (Aureli, 2017). Aureli (2017) explores the differences in more detail and find that content analysis is the more common method in research but is employed more often in qualitative studies. Content analysis focuses more on sentences and the content around a single term, rather than solely the term itself. The context dimension makes it more suitable for identification of complex concepts in documents with large amounts of unstructured text. Hooks & van Staden find in their paper from 2011 a high correlation between content analysis measures and quality of the information in environmental reports, further proving the effectiveness and reliability of this technique.

Text mining, on the other hand, relies even more on computer programs and algorithms to process larger amounts of textual data. It distances itself further from human involvement, and single keywords are transformed into quantitative data. This thesis deal with a relatively large fund sample over a three-year period, the cumulative amount of textual data consists of more than 5 500 10-K forms, each often with a length of around 200 pages. For this reason, the computational power of a textual mining approach is deemed favourable and necessary. In the following section, we will further explore some of the research done using text mining approaches on sustainability disclosures.

Aureli, Medei, Supino & Travaglini (2016) use text mining to analyse sustainability disclosure after an industrial crisis. Similar to this thesis, a glossary of terms related to sustainability and their frequency in reports is used as a basis for the textual analysis. Te Liew, Adhitya & Srinivasan (2014) use a textual-based analysis method to unveil sustainability trends and practices in four main sectors of the process industry. Once again term frequency is the basis of the analysis. More closely related to this thesis Wen (2014) compares ESG ratings of individual companies and term frequencies in annual reports. These studies demonstrate a wide variety of applications in which textual analysis produces relevant data from a large amount of text. The structure of the data and the tests performed vary depending on what the researchers are looking for. However, in each study, a term frequency count lay the grounds for converting textual data into quantitative data for further analysis.

This thesis uses annual 10-K reports filed by U.S. companies as a basis for its textual data analysis. As discussed in the data section of the thesis, 10-K reports are comprehensive annual reports and a primary source of information about the company’s activities accessible to all stakeholders. Additionally, a set of reporting rules makes 10-Ks more structured and thus better suited for a text mining approach as all firms are required to report a minimum level of information.

With this thesis, we contribute to the existing literature by examining the relationship between transparency and sustainability on fund level. This contrasts with previous transparency-studies performed on individual companies (Clarkson et al., 2011; Wen, 2014). As a consequence of having 118 funds in our sample, we are analysing reports from 2 027 companies – making the source data more extensive than in the reviewed literature. Another aspect we bring to the literature is the assessment of several investment categories. Wen (2014) uses the same regression across different industries, while Clarkson et al., (2011) are only assessing the transparency/emission-relation within the mining and manufacturing industry. A key contribution of our thesis is the construction of a tool that can create a quantifiable measure of sustainability disclosure of a fund's underlying companies.

4. Data

4.1 Overview of Data Sources

The two main data sources used in this thesis involve fund data retrieved from Morningstar and 10-Ks filed to the US Securities and Exchange Commission (SEC). Morningstar is a US-based provider of investment research and is regarded as an influential force in asset management (Marriage, 2017). Using Morningstar's investment platform, Morningstar Direct, we are able to retrieve present and historical sustainability ratings of US open-end mutual funds. Our textual data analysis is based on annual 10-K reports from the companies represented in the sample funds, which are downloaded via SEC's EDGAR database.

4.2 Sample Selection

With data available for thousands of global funds, a sample selection is required before further analysis is done. The process of choosing a sample is a result of both the required nature of the data, as well as an active selection of the remaining funds. The steps of this screening process are displayed in table 1 below.

Table 1: The screening process

	Screening action	Sample size
1	Initial sample size	293 161
2	Open-end funds	276 465
3	Equity funds	105 651
4	Investment Area – the US	11 749
5	Complete historical scores	6 332
6	Qualified funds, duplicates removed	1 201
7	Final selection	118

Morningstar provides data for 293 161 global funds through its platform Morningstar Direct. An initial screening process directly on the platform reduces the sample significantly before it is downloaded, and further screening is performed locally. The first criterion involves limiting the sample to open-end mutual funds. In comparison to Exchange traded funds (ETFs), mutual funds are more actively managed. This is an important criterion as we are interested in funds that are actively picking stocks that are performing well regarding ESG criteria and transparency, and then comparing them to funds that might be actively picking stocks based on other principles.

Next, we screen the sample by limiting it to funds that are categorized as equity funds. Thus, we remove any funds that invest in fixed income, commodities, property, the money market, etc. This is because our analysis requires 10-Ks, which are only available for companies, and not any of the other investment groups. The last initial screening limits our sample to funds that are primarily invested in U.S companies. Again, 10-Ks are only available for U.S companies, and thus we have limited our research to the U.S market. At the end of this initial cleaning, we are left with 11 749 funds.

The following part of the sample selection process is done locally in the software R. The code used in this process can be found in Appendix 1. Since we base our selection on the Morningstar Sustainability Rating and use this data later in the analysis, we require that the remaining funds have a quarterly sustainability rating that goes back to Q1 2016. After contacting Morningstar, we were able to retrieve the available historical sustainability scores for funds. By doing this, the sample size is reduced to 6 332 funds. Of the remaining funds, a majority is what we in our case will define as duplicates. Though they differ slightly in name, such as "*AB Equity Income A"* and "*AB Equity Income B"*, they have the exact same weighted holdings, and thus also the same sustainability rating. We remove all duplicates by assuming that if two funds have the exact same portfolio sustainability score and the same number of holdings, they are indeed duplicates.

The remaining 1 201 funds in our sample are all eligible to be used in our analysis. However, though computational textual analysis is faster than any human-based method, downloading and processing 10-Ks is still a time-consuming procedure. To limit the time spent on processing funds, a subsample of funds is selected.

Our final selection of funds is chosen based on the investment category in which they are assigned by Morningstar, and by their Morningstar Sustainability Rating. The top four categories are all U.S. Equity funds; large cap blend, large cap growth, large cap value and mid cap. In each of these categories, the top five rated, the bottom five rated and the five funds around the median is chosen for further analysis. Similarly, a selection of funds in the four top sector categories are chosen. These sectors are technology, financials, consumer goods and services, and healthcare. The five top, bottom and median rated funds are selected in each of these categories as well. Ultimately, the selection process is designed to leave us with 120 open-end funds that are to be used in the analysis.

4.3 Summary Statistics

Table 2 contains summary statistics from Q4 2018 for both the 1 201 qualified funds, as well as the 118 funds in our final sample. The statistics are helpful to determine if there are selection biases present in our final sample. Table 2 shows that in terms of sustainability score and the three ESG pillars, the final selection is an acceptable representation of the whole range of scores. In terms of size, both in value and number of holdings, our final selection is on average smaller than those in the full population. However, after further examination of the data, and as partly seen from the large difference between the third quarter statistics and maximum values, much of this difference between the final selection and full population can be attributed to extreme outliers in the full sample. Considering that we do not want to include these outliers in the final selection, we deem the final selection as an adequate representation of the population of funds we want to assess.

Table 2: Summary Statistics Q4, 2018

		Min	1st Quarter	Median	Mean	3rd Quarter	Max
Sustainability Score	Full sample	35.70	44.13	45.71	45.47	46.72	57.64
	Selection	38.81	42.32	44.78	45.26	47.21	57.64
Environment Score	Full sample	41.27	50.71	53.93	52.81	55.52	65.28
	Selection	41.63	50.31	53.63	53.33	56.66	65.28
Social Score	Full sample	41.89	49.82	52.44	51.57	53.61	63.76
	Selection	41.89	47.48	50.27	51.06	54.28	63.76
Governance Score	Full sample	37.61	48.86	50.35	50.35	51.81	59.92
	Selection	37.61	48.15	50.40	50.52	52.84	59.92
Fund Size (Million USD)	Full sample	0.28	128.1	492.0	3 709.2	1 808.9	671 889.9
	Selection	0.28	85.2	312.4	1 425.5	1 320.9	19 075.1
Number of Holdings	Full sample	13	48	74	149.9	123	3515
	Selection	18	39	63	86.6	101	478
Age (days)	Full sample	1 034	5 676	7 509	8 253	9 748	34 502
	Selection	1 431	6 424	7 749	8 490	10 412	30 359

4.4 Holdings

As the textual data analysis in this thesis is based on 10-Ks filed by U.S. companies, the holdings of each fund are downloaded. The holding data is acquired through the Morningstar Direct platform. Quarterly holdings from Q1 2016 to Q4 2018 are downloaded for each fund. Holdings are quantified as weightings of total fund distribution. Weightings are later used to proportionally weight the ESG scores derived from the textual data analysis. Of the 120 funds in the last part of the selection, holdings were available and downloaded for 118 of them. The two remaining funds were dropped before further processing due to missing holdings.

4.5 Morningstar Sustainability Rating

Since we want to examine whether businesses' transparency in the 10-Ks align with their measured ESG-performance, we need a rating we can relate the results from the textual analysis to. After evaluating the alternatives, we decided to use the Morningstar Sustainability Rating (MSR) – partly due to Morningstar's standing as a leading provider of investment data. Since the launch in 2016, the rating has become a well-known tool among investors, offering quantification of ESG performance to over 20 000 funds ETFs and mutual funds globally (Morningstar, 2016b). Morningstar's fund ESG-rating is constructed on company-level scores provided by a third-party company, Sustainalytics.

Sustainalytics is a prominent actor within the field of ESG and corporate governance research. By examining various disclosure forms, and in some cases, direct outreach to the companies, Sustainalytics construct ESG-reports which cover over 9 000 companies across 42 industries. (Ezeokoli, Layne, Statman & Urdapilleta, 2017). These ESG-reports provide qualitative analysis and quantitative ratings that assess the extent to which companies address relevant environmental, social and governance issues (Systainalytics, n.d.; Ezeokoli et al., 2017). Within each of the three E, S and G pillars, companies are evaluated in three dimensions; preparedness, disclosure, and performance. "Preparedness" measures commitment to handling ESG risks through stated policies and programs, while "disclosure" reflects to which extent a company is transparent in its ESG activities and reporting. "Performance" is estimated using numerous sector-adjusted quantitative and qualitative indicators. Each sub-score follows an industry-specific weight matrix, meaning that the aggregated company-score reflects a company's ESG performance within that industry – enabling for peer-to-peer comparison.

Sustainalytics also consider ESG-related controversies by deducting companies' scores according to the involvement in significant controversies (Hale, 2016; Ezeokoli et al., 2017).

Morningstar utilizes this company-level data to construct portfolios corresponding to the funds in the Morningstar Sustainability Rating. The "Portfolio Sustainability Score" is given by:

$$\text{Portfolio Sustainability Score} = \text{Portfolio ESG Score} - \text{Portfolio Controversy Score} \quad \textbf{(1)}$$

Portfolio ESG Score and Controversy Score is an asset-weighted average of the company-level ESG scores in the respective fund. Due to Sustainalytics' unique combinations of indicators for each peer group to reflect the relative ESG-performance of companies in the same industry, they are not directly comparable across industries. To make the ESG scores alike across peer groups, Morningstar normalizes the scores of each group using a z-score transformation (Morningstar, 2018a). These z-scores are used to generate normalized ESG scores on a 0-100 scale, with a mean of 50, as follows:

$$Z_{Peer} = \frac{ESG_A - \mu_{Peer}}{\sigma_{Peer}} \quad \textbf{(2)}$$

$$ESGNormalized_A = 50 + (Z_{Peer} \times 10) \quad \textbf{(3)}$$

Sustainalytics does not obtain data from every company present in one of the 20 000 funds the MSR cover. Morningstar deals with this issue by requiring that at least 67% of a portfolio's assets under management must have a company ESG score to receive a Portfolio ESG score (Morningstar, 2018a). The funds who are rated in the MSR are given a score between 1 and 5, where 3 is the peer group average, while 5 and 1 represent the top and bottom 10%, respectively. A score of 5 means the fund score at least two standard deviations above average in its peer group, and vice versa. Figure 2 defines the distribution of funds:

Distribution	Score	Descriptive Rank	Rating Icon
Highest 10%	5	High	
Next 22.5%	4	Above Average	
Next 35%	3	Average	
Next 22.5%	2	Below Average	
Lowest 10%	1	Low	

Source: Morningstar

4.6 10-K reports

All US companies that are listed on a national securities exchange[1] are obliged to file a 10-K report to the SEC annually (EY, 2017). The form of 10-K is required to follow a set of SEC rules. While this makes the report more generic and less visually appealing, the required structure makes it appropriate for textual analysis. The 10-K report is more comprehensive and detailed than annual financial reports, which are primarily meant for shareholders. 10-Ks, on the other hand, address a wider range of stakeholders. 10-Ks are filed through the Electronic Data Gathering, Analysis, and Retrieval system (EDGAR). From here anyone can access and download the data (SEC, 2019). Filings are accessed using the EDGAR index files. Yearly index files are published indexing all public filings. Index files contain information such as company name, form type, CIK (Central Index Key), and file name. These traits facilitate automated crawling of the EDGAR database.

The disclosures of the 10-K are often related to description of the business, risk factors, properties, legal proceedings, financial data and management's discussion and analysis of the financial conditions (SEC, 2011). The U.S., unlike the EU (the Non-Financial Reporting Directive), have not implemented cross-state regulations that require disclosure of environmental, social and ethical aspects (European Commission, n.d). However, under existing federal securities laws and regulations, companies are obligated to disclosure ESG-issues that are likely to have a material effect on the businesses and their operations (SEC, 2011). Investor Responsibility Research Center Institute (IRRCI, 2018) also found that 23% of the companies on the S&P 500 voluntarily address sustainability in 10-Ks. This number only includes those firms who disclose issues beyond what is already regarded as obligated, i.e. material issues.

4.7 ESG Dictionary

From each of the 10-K reports used in this thesis a term frequency count is produced. This process is explained in detail in the methodology section of the thesis, but the end result is structured as a Document-term Matrix (DTM). A DTM is an especially useful structure for

[1] Companies that hold assets equivalent to at least $10 million and have more than 2 000 equity security (or 500 non-accredited) holders are also obliged to file a 10-K report to the SEC annually (EY, 2017)

information retrieval, term weighting and document clustering in textual analysis (Xu, Liu & Gong, 2003; Shahnaz, Berry, Pauca & Plemmons, 2006). Along one axis are the documents, which in this case are the 10-K reports. Along the other axis are terms that appear in the documents. This axis includes our specific selection of terms that are meant to represent ESG topics. This list of terms is referred to as a glossary or dictionary.

The ESG dictionary used in this thesis is primarily based on a dictionary created by Baier, Berninger & Kiesel (2018). The dictionary is created by procedures developed through existing textual analysis literature and is also based on 10-K reports. Terms pass through several screening steps and relevance tests before being included in the final selection. The ESG dictionary created by Baier et al. (2018) includes 482 terms. Additionally, another 19 terms are added by us to bring the final dictionary up to a total of 501 ESG related terms. Terms are categorized according to the E, S and G pillars and distributed as follows: 69 Environmental terms, 156 Social terms, and 276 Governance terms. The dictionary is found in Appendix 3.

5. Methodology

This section of the thesis will explain the procedures we are performing in R to answer our research questions. In short, the process can be divided into three components. The first step is data retrieval and cleaning of the data. In this part, we retrieve the data from the annual 10-K filings, remove stop words and count the frequency of relevant terms in each individual document. In the second step, term weighting is used to adjust the value of keywords according to their relative frequency across the sample. Lastly, the ESG-scores are standardized, before the scores from each company filing are weighed to match the holdings of sampled funds. All procedures are done in R, and the complete code can be found in Appendix 1.

5.1 Data Retrieval in R

All funds in the final selection are broken down to their holdings from 2016, 2017 and 2018. 10-K reports are then downloaded for each company corresponding to the year they are held by a fund. If a company is held over several years, the 10-K report for each year is downloaded. This is done using an automated crawler of the EDGAR database using R code.

One central feature of R is the number of packages available for specialised techniques and capabilities in the code. One of these packages is the “edgar” package (Lonare & Patil, 2017). Key functions in this package are used to assist in downloading annual reports, and to construct term frequencies from 10-K reports. EDGAR master index files are downloaded for each year for easier access to document locations in the database. The master index files use companies’ Central Index Key (CIK) as an identifier, while holdings downloaded from Morningstar identify firms by tickers. A conversion is therefore applied to match CIK and tickers before 10-Ks are downloaded. Holdings that do not have a 10-K filing in EDGAR are dropped.

The SEC operates with a filing deadline on 10-K reports of 60-90 days after the end of the company’s fiscal year. Thus, a company is required to file its 10-K report within the first quarter of the year if their fiscal year aligns with the calendar year. Holdings are reported on the last day of each quarter. A 10-K report filed within the first quarter of a year is therefore used as the base for the textual analysis ESG score for all quarters that year. However, if the fiscal year of a company diverts from the calendar year or a report is filed late, the 10-K report might not be filed the first quarter. In this case, the annual report from the previous year must be used for the textual analysis, until the new report is filed. Holdings are adjusted for this and matched with their appropriate 10-K filings in EDGAR.

Next, 10-K reports are downloaded, and term frequencies are constructed using the “edgar” package in R. The function “getFilings” used for downloading 10-Ks is slightly edited to bypass the need for a user input each time a report is downloaded. Next step is to use the function “getWordFrquency” from the “edgar” package on the downloaded report. This function converts the report into a corpus and cleans the text. The cleaning removes punctuations, numbers, excess whitespaces and English stop words. Stop words are the most common words in English such as “the”, “to”, “of”, “and”, etc. Removing these helps reduce noise later in the analysis. All remaining terms are returned along with their respective frequency in the report. Additional cleaning is performed to remove irrelevant terms that describe style and design of the report such as font style and size.

Using the ESG dictionary, all ESG terms are extracted from the term frequency count. Terms that occur in the dictionary but not in the 10-K report are given an NA value. The results of the ESG term frequency count is saved as a new column in a Document-term Matrix. The result is also tagged with the CIK of the company and the year in which the 10-K was filed. A 10-K filed by Apple Inc in 2017 would, for example, be tagged as “320193_2017”. This is repeated for all holdings in every fund. In the end, 5 556 annual reports are converted to ESG term frequency lists, and results are stored in a Document-term Matrix. A subset of the final Document-term matrix is shown below, in table 3.

Table 3: Extract of the Document-term matrix

	62709_2016	316709_2017	80424_2015	866787_2016	899689_2017	217346_2018	821026_2016	1413329_2016
agricultural	NA	NA	NA	NA	NA	1	21	4
charitable	NA	NA	NA	NA	NA	2	NA	NA
community	NA	16	NA	NA	3	NA	NA	3
conduct	22	28	41	15	55	23	2	17
education	4	5	NA	NA	1	NA	NA	1
environmental	7	NA	59	8	73	95	15	14
ethical	NA	NA	1	5	NA	NA	NA	1
healthcare	8	2	3	1	NA	28	5	NA
integrity	NA	6	5	2	1	NA	NA	NA
misconduct	1	10	2	NA	14	NA	NA	17
renewable	NA	NA	NA	NA	2	NA	2	NA
safety	6	5	4	13	12	14	6	8
social	1	4	4	4	2	NA	NA	7
transparent	NA	NA	NA	NA	NA	NA	NA	NA

5.2 Term Frequency – Inverse Document Frequency

Term weighting schemes are often used in textual analysis to evaluate the importance of certain keywords in a sample of documents. Our ESG dictionary contains words with great differences in frequencies in our sample. To mitigate overestimation of the information value in high-frequency words, we therefore apply term weighting (Loughran & McDonald, 2011). One common method of term weighting is called *tf-idf*. In its simplest form, the formula for this weighting process can be expressed as (Jurafsky & Martin, 2009):

$$w_{i,j} = tf_{i,j} * idf_i \qquad \textbf{(4)}$$

$$idf_i = \log(\frac{N}{n_i}) \qquad \textbf{(5)}$$

$$Where:$$

$\mathrm{tf}_{\mathrm{i,j}} = \text{Term frequency of word } i \text{ in document } j$

$\mathrm{N} = \text{the total number of documents in the sample}$

$\mathrm{n}_{\mathrm{i}} = \text{the number of documents in which the term } i \text{ occurs}$

The first term of the formula, tf (term frequency), simply counts the number of times a keyword occurs in each document. The more frequent a term appears in the document, the higher the weight. The last term, idf (inverse document frequency), alters the importance of a word based on how often the words are used across all documents in the sample. Terms that occur in fewer documents receive a higher weight. This lowers the value of common words. In our case ESG-words such as “health” and “audit” is used in close to all documents in our sample. These words receive a lower weight than for example “biodiversity”, which only appears in 21 documents, or “minorities”, which only appear in 40 documents. Instead of using the raw term frequency count, a log normalization is used as suggested by Loughran & McDonald (2011). Additionally, log average term frequency is used to make sure important keywords that might not be used very often in each document are still weighted proportionally to its importance, as suggested by Umemura & Church (2000). The final weighting scheme used in this thesis is shown below as both an equation and as programmed in the R code.

$$w_{i,j} = \frac{1+\log(tf_{i,j})}{1+\log(ave_{i\in j}(tf_{i,j}))} * \log(\frac{N}{n_i}) \qquad \textbf{(6)}$$

$$where:$$

$$ave_{i\in j}(tf_{i,j}) = \frac{\sum_j tf_{i,j}}{n_i}$$

Figure 3: tf-idf Function in R

```
TF.IDF <- function(corpus) {
  tf.t <- apply(corpus, 1, function(x) sum(x, na.rm = T))
  df.t <- apply(corpus, 1, function(x) length(which(!is.na(x))))
  avtf <- tf.t/df.t
  tf <- apply(corpus, 2, function(x) ((1+log10(x))/(1+log10(avtf))))
  tf <- as.data.frame(tf, stringsAsFactors = F)
  D <- length(corpus)
  idf <- apply(corpus, 1, function(x) log10(D/(length(which(!is.na(x))))))
  tf.idf <- apply(tf, 2, function(x) x*idf)
  tf.idf <- as.data.frame(tf.idf, stringsAsFactors = F)
  return(tf.idf)
}
```

Table 4 shows the same subset of the document-term matrix as in table 3 when the tf-idf weighting scheme is applied. Words such as “conduct” and “safety” which appear in a large proportion of the document are given a lower weight as a result of the inverse document frequency part of the function.

Table 4: Effect of the tf-idf weighting scheme.

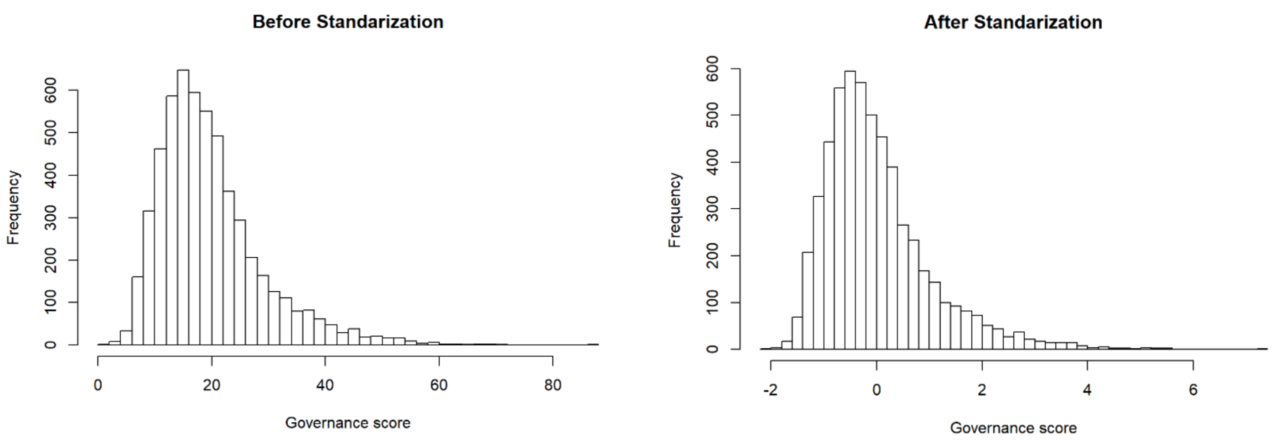
	62709_2016	316709_2017	80424_2015	866787_2016	899689_2017	217346_2018	821026_2016	1413329_2016
agricultural	NA	NA	NA	NA	NA	0.3235055	0.7512507	0.5182752
charitable	NA	NA	NA	NA	NA	0.4806105	NA	NA
community	NA	0.2238896	NA	NA	0.1500427	NA	NA	0.1500427
conduct	0.0036833	0.0038479	0.0041084	0.0034217	0.0043090	0.0037136	0.0020458	0.0035072
education	0.2718364	0.2882800	NA	NA	0.1696793	NA	NA	0.1696793
environmental	0.0607566	NA	0.0912404	0.0626662	0.0942854	0.0980524	0.0716558	0.0706691
ethical	NA	NA	0.4542826	0.7718126	NA	NA	NA	0.4542826
healthcare	0.2310120	0.1579292	0.1793045	0.1213878	NA	0.2970552	0.2062343	NA
integrity	NA	0.2857701	0.2730447	0.2090910	0.1607119	NA	NA	NA
misconduct	0.2274438	0.4548876	0.2959112	NA	0.4881235	NA	NA	0.5073018
renewable	NA	NA	NA	NA	0.3873895	NA	0.3873895	NA
safety	0.0005246	0.0005012	0.0004726	0.0006237	0.0006134	0.0006332	0.0005246	0.0005615
social	0.0853817	0.1367866	0.1367866	0.1367866	0.1110841	NA	NA	0.1575376
transparent	NA	NA	NA	NA	NA	NA	NA	NA

5.3 Restructuring Results

After terms are weighted, results are restructured back to a fund level.

First, the Document-term Matrix is separated by terms according to the E, S or G pillars. Within each pillar, the weighted terms are summed for every 10-K report. Meaning each annual report receives an environmental, social and governance score from the sum of the weighted term frequencies in the report. A common characteristic of the annual report scores is that the governance score is higher than the other two, and the environment score often lower. To avoid overweighting one score, but rather make them comparable across all pillars, we standardize the results. Importantly, this does not alter the distribution of the scores as seen in figure 2. This standardization method uses a mean of 0 and a standard deviation of 1.

Figure 4: Distribution of governance score before and after standardization



To construct the scores on a fund level we use the holdings. Holdings are stated as weights, and these weights are multiplied with the E, S and G scores of the annual 10-K filing. This way the final score is proportional to how much of a fund's total assets are allocated to each company each time period. However, due to some missing holding data or holdings not having a 10-K filing, weights do not always add up to 100%. The weighted sums can vary across funds and time periods. To adjust for this inconsistency, we drop funds or time periods where we are missing more than 50% of the weighted holdings and normalize the remaining weightings so that they always add up to 100% using the formula (7) shown below. Lastly, the scores from each of the three pillars are summed within each fund to get the full ESG score for that fund. The result is an E, S, G, and full ESG score for each fund over 12 quarterly time periods from Q1 2016 to Q4 2018.

$$w_i' = \frac{w_i}{\sum_{j=1}^{n} w_j} \tag{7}$$

5.4 Second Iteration: By Category

For the second iteration of the code, some modifications are made. The funds are now separated according to the investment category they are defined under. Our fund sample consists of funds in 8 different investment categories. Four of these are variations of U.S. Equity funds: U.S Equity Large Cap Blend, Large Cap Growth, Large Cap Value, and Mid Cap. The other four are investment categories based on sectors: Consumer Goods and Services Sector, Healthcare Financials Sector, Sector and Technology sector.

Eight separate document-term matrices are constructed in the second iteration, one for each category. When weighting terms, it is done within each of these DTMs separately. The effect of this is that terms in the original DTM which are moderately infrequent across the whole sample, but which appear in most of the documents in one specific category are adjusted for and weighted less. For example, terms such as "medicaid", "healthcare" and "medicines" appear more frequent in companies held by healthcare funds due to the nature of the sector. The steps in structuring the scores to fund level are identical as previously. The following standardization of the results is performed to make sure the scores can be compared across categories.

5.5 Third Iteration: Green labelled Versus Conventional

In the third iteration of the code, an additional sample of funds is included. These are funds that we define as “green labelled”. This simply means that the name of the fund suggests that it has a stated focus on following a sustainable investment strategy. We search the data for funds that contain certain words in the fund name. The search code and the following output are shown in figure 5. The terms used to find “green labelled” funds include abbreviations such as “ESG” and “SRI”, as well as the stemmed version of words like “Sustainable” and “Responsible”. Stemming these words makes sure we include alterations of the words such as both “Sustainable” as well as “Sustainability”. 17 funds are identified as “green labelled”, some were already included in previous iterations, but holdings for the new ones were downloaded and included in this iteration. At the end of the code, a dummy variable called “green label” is created to distinguish the “green labelled” funds from the control group. The 17 “green labelled” funds are given a value of 1 and the remaining a value of 0.

Figure 5: Extract from code, "Green labelled"

```r
df$Name[which(grepl("Sustain.|Green.|ESG.|Responsib.|SRI.|CSR.|Social.|Carbon.|Env
iron.|Renew.", df$Name))]

##  [1] "AB Sustainable US Tmtc A AUDH Acc"
##  [2] "American Century Sustainable Equity A"
##  [3] "BNPP Easy MSCI KLD 400 US SRI Track CC"
##  [4] "Dreyfus Sustainable US Equity A"
##  [5] "ERSTE Responsible Stock Ameri CZK D02 VA"
##  [6] "Green Century Equity Individual Investor"
##  [7] "JPMorgan Intrepid Sustainable Equity A"
##  [8] "Neuberger Berman Sustainable Eq A"
##  [9] "Northern Trust NA Val ESG E EUR Inc"
## [10] "Pax ESG Beta Quality A"
## [11] "Putnam Sustainable Future A"
## [12] "Putnam Sustainable Leaders A"
## [13] "Russell Inv Sustainable Equity A"
## [14] "Sustainable North America Index Fund"
## [15] "TIAA-CREF Social Choice Eq Advisor"
## [16] "Touchstone Sustainability & Imp Eq A"
## [17] "Vanguard FTSE Social Index I"
```

Table 5 shows an example of the modified dataset. The panel includes the textual and Morningstar score, the fund ID, whether the fund is "green labelled", as well as the time period of the observation.

Table 5: Dataset, "Green labelled"

	ID	Time	Green Label dummy	Morningstar Rating	Textual Score
624	GB0006061740	9	0	40.61	-1.01029466
614	GB0006061740	10	0	40.43	-0.98886044
615	GB0006061740	11	0	40.08	NA
616	GB0006061740	12	0	40.16	-0.96759455
625	GCEQX	1	1	52.25	0.25461292
629	GCEQX	2	1	51.69	0.26565687
630	GCEQX	3	1	51.07	0.2511175
631	GCEQX	4	1	50.42	0.24360042
632	GCEQX	5	1	50.21	0.49469824

5.6 Models and Tests

The quarterly MSR-scores are retrieved from Morningstar Direct for the funds in our sample for 2016, 2017 and 2018. For the last quarter in 2018, individual environmental (E), social (S) and governance (G) scores are available, whereas the other time periods only include the combined ESG-score. The MSR-scores from these time periods also includes the controversy score[2].

5.6.1 Transparency and Sustainability

Test for correlation

The first tests performed to identify a potential relationship between the two scores are two correlation tests. These are meant to determine whether there is a significant relationship between sustainability and disclosure, and how strong this relationship is. Kendall's τ (tau) and Spearman's ρ (rho) are two rank correlation tests used to measure correlation in rank ordering between two variables. These tests are particularly useful if the distribution is not

[2] As the ESG dictionary is divided into E, S and G sections, it would have been favourable to have accessed the corresponding data, excluding controversy score, from Morningstar for 2016 and 2017. However, when we requested these, Morningstar confirms that this data is not available in Morningstar Direct.

normal, or if the measure itself is of less interest than the rank of each observation, as it often is in Information Retrieval systems (Carterette, 2009).

Regression

Regression is used to best fit a line and estimate one variable based on the other. In this thesis, we use a linear regression model, or more specifically an ordinary least squares (OLS) regression model. In the model, we use the Morningstar rating (MSR) as the dependent variable, and our textual ESG score as the independent variable. We use the Morningstar Sustainability Rating as a proxy for sustainability performance when we ask: Can the textual ESG disclosure score, based on the disclosures in underlying holdings, predict the sustainability performance of a fund? The formula for the first regression can be expressed as the following:

$$MSR = \alpha + \beta x_{TextualESG} + \varepsilon \qquad \textbf{(8)}$$

The next regressions are separated by the three "pillars" in each score, i.e. environmental, social and governance. The formulas for these regressions are expressed as:

$$p \in Pillar\text{: } Set\ of\ ESG\ pillars\ (Environmental, Social\ and\ Governance)$$
$$MSR_p = \alpha_p + \beta TextualESG_p + \varepsilon_p, \quad \forall p \in Pillar \qquad \textbf{(9)}$$

Lastly, regressions are separated by category. By doing this we can see if the relationship between sustainability performance and disclosure in underlying companies differ for each investment category in our sample. The regressions are expressed as:

$$c \in Category\text{: } set\ of\ different\ investment\ categories$$
$$MSR_c = \alpha_c + \beta TextualESG_c + \varepsilon_c, \quad \forall c \in Category \qquad \textbf{(10)}$$

The regressions formulated by expression **(8)** and **(10)** are also run again with observations from all time periods. When running over all time periods the MSR represents the full Morningstar sustainability score **(1)** including the controversy score, and no longer just the Morningstar ESG score.

5.6.2 Difference in scores

When creating the difference between the textual ESG score and Morningstar ESG score ($\Delta Score$) we first standardize both scores to a mean of 0 and a standard deviation of 1. Then we can compute the absolute difference between the scores. This allows us to examine distances between the scores while ignoring the direction of the difference. The explanatory variables in the regression include characteristics such as fund size, number of holdings and days since inception date. Additionally, fund beta is included as a measure of volatility compared to the market. S&P 500 is included as the market benchmark in this case. Lastly, two measures of investment style are included: Each fund's investments into small versus big size companies, and their investments into value (high book-to-market) versus growth (low book-to-market) companies. In the three-factor model (Fama & French, 1993) these are referred to as size risk factors and value risk factors. Measures of investment styles are taken from Morningstar's Style Box (Morningstar, 2008). Both measures range from -100 to 400. A low score in the size risk measure represent higher investments in smaller companies, and a low score in the value risk measure represent higher investments in value companies. The regression is formulated as:

$$\Delta Score = \alpha + \beta Size + \beta NumHoldings + \beta Age + \beta Beta + \beta SmallBig^{3} + \beta ValueGrowth + \varepsilon \quad \textbf{(11)}$$

5.6.3 Green labelled funds

T-test

To determine if the means of two populations are significantly different, we will use a Welch two-sample t-test. One population consists of all (17) identified "green labelled" funds among the 1201 qualified funds. The other is the population of conventional funds in our final sample. This method is more reliable when dealing with different sample sizes, as we do here (Ruxton, 2006). As a nonparametric alternative to the two-sample t-test, we use a Wilcox rank sum test where assumptions of normally distributed samples are violated.

5.6.4 Time trend

Time trend

To test the development of disclosure over time, we run a time series regression with a time

[3] There are three Large cap fund categories, one Mid cap category, and no Small cap fund categories in our sample. This is due to an insufficient sample size of Small cap funds. As a result, this measure likely to suffer from a selection bias. Interpreting the results from this coefficient is therefore done with caution.

trend as the independent variable. The dependent variable is the textual ESG score, and the independent variable is the time trend measured in quarters. The regression is set up as:

$$Textual\ ESG\ score = \alpha + \beta t + \varepsilon_t \tag{12}$$

As a comparison, the development of the Morningstar Sustainability Rating over time is tested in an identical matter, estimating the time trend in a time series regression.

$$Morningstar\ Sustainability\ Rating = \alpha + \beta t + \varepsilon_t \tag{13}$$

The regressions are intended to reflect the explanatory power of written disclosure in underlying companies, represented by the textual score, on sustainability, represented by the MSR score.

5.7 Model Requirements

To ensure valid results from the ordinary least squares (OLS) model, there are several conditions that must be fulfilled. A violation of these conditions means that the OLS will either be biased or inefficient and thus no longer BLUE (Best Linear Unbiased Estimator) (Wooldridge, 2016). The relevant tests for these conditions are covered in the following section.

Test for heteroskedasticity

The homoskedasticity assumption is also called the constant variance assumption, denoted by; $Var(u|x) = \sigma^2$. If the variance of the error term is not constant, we have a heteroskedasticity issue. This can be examined by conducting a Breusch-Pagan test. If detected, this can be corrected by using robust standard deviations.

t-test assumptions:

For a valid interpretation of the t-test results, the following assumption must be fulfilled. The normality assumption implies that variables are distributed according to a normal (Gaussian) distribution with a mean of zero. To investigate if this assumption is met, we have used a Shapiro-Wilk test. In cases where the variables are not normally distributed, the Wilcox rank sum test is implemented as a nonparametric alternative to the t-test.

5.8 Limitations of Design

As already stated, quantification of sustainability is not a straightforward process and there is divergence among the competing ESG-ratings. In this thesis, we are assuming that the Morningstar Sustainability Rating can act as a proxy for sustainability. This is partly due to Morningstar's position as a prominent and experienced actor in the industry, and partly because we could access their data at our institution. Our relatively short timeframe from Q1 2016 to Q4 2018 is also due to restrictions in Morningstar's dataset. As our sample consists of funds that were available from Q1 2016, any funds launched after this date is omitted from the analysis. Thus, it is possible that our sample does not truly reflect today's US mutual funds, as newer funds are omitted.

As for our textual analysis, we decided to use the tf-idf-method instead of sentiment analysis. The reason for this is that we were unable to find an ESG-specific dictionary that enabled sentiment analysis. Hence, we find it preferable to base our analysis on specific ESG-related keywords rather than performing a sentiment analysis. The main argument behind the choice of method, besides that it consumes less computational resources, is the lack of a relevant sentiment dictionary. Even though the existing Harvard sentiment dictionary is often used in textual analysis, this has proven to be unreliable in a financial context by other studies based on 10-K filings (Loughran & McDonald, 2011).

Moreover, one could argue that we ideally should retrieve ESG-disclosure information not only from annual 10-K filings, but also from other channels such as companies' webpages, stand-alone sustainability reports and public statements. It is likely that relevant ESG-disclosures are also publicized in alternative channels. Our issue is that these do not have central databases and/or standardised structure of text, which in turn makes our textual analysis unfeasible. It can also be argued that voluntary, unassured reports may contain a favourable language and/or be biased (Cho, Roberts, & Patten, 2010; Wen, 2014).

When comparing the "green labelled" funds versus the conventional funds, we introduce a possible selection bias as we bring in new funds from the sample of the 1201 funds that were qualified for further research. All the 17 identified "green funds" from the qualified 1201 funds are included. Note that five of the 17 "green labelled" funds are already included in our final sample. These are therefore removed from the final sample in the third research question. The remaining sample of 113 conventional funds are compared to the 17 "green labelled" funds.

6. Results and Discussion

The following section presents the results and interpretation of our analysis. The first two research questions examine whether transparency can predict sustainability performance and if the relationship varies across fund categories and characteristics. In the next question, we investigate if there exist any differences in measured transparency and sustainability between the identified green funds and conventional funds. Lastly, we study the development of transparency and sustainability over the sample period from 2016-2018.

6.1 Research Question 1: Transparency and Sustainability

In our first research question, we aim to answer whether sustainable funds hold more ESG-transparent firms, if our textual ESG score is a possible predictor for sustainability performance, and if there are any differences across the sample categories.

Research question 1: *Are there more ESG-disclosures in the underlying companies of sustainable funds, and is a measure of ESG-disclosures in underlying companies a possible predictor for sustainability performance of a fund? Are there any differences across investment categories?*

The first tests are based solely on data from the latest observation, Q4 2018. From this period, we have access to the full set of individual E, S and G scores from Morningstar. Furthermore, we can exclude the controversy score that Morningstar adds to their portfolio score, which is not the case in the rest of the sample period. As both our computed textual score and the MSR score from this period consists of three individual E, S and G components, this allows for further examination of these variables.

6.1.1 Correlation

Table 6 shows that both rank correlation methods result in statistically significant p-values. From this, we can reject the assumption of no correlation between the two scores, i.e. $\tau/\rho = 0$. In other words, there is a significant rank correlation between the Morningstar ESG score and the textual ESG score in the last period (Q4, 2018) of the data. The rank correlation tests show there is a relationship between the two scores and that they to some degree move together in a positive direction.

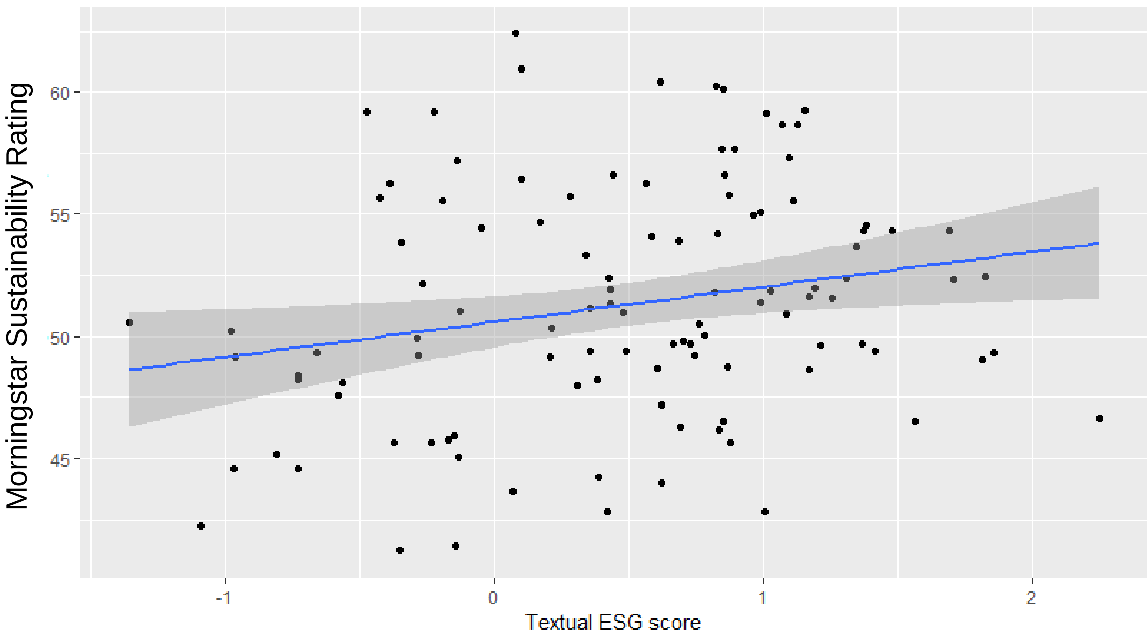
Table 6: Textual and MSR correlation, Q4 2018

Rank correlation			
Kendall's τ		Spearman's ρ	
tau	0.1597575	rho	0.2535298
p-value	0.01297	p-value	0.007256
z	2.4847	S	170140

6.1.2 Regression

A linear regression is performed of the computed textual ESG score on the Morningstar ESG score. Figure 6 shows the regression plot. Table 7 contains the results from the Morningstar ESG score regressed on our textual score, based on observations from Q4 2018. The regression is performed with the full sample of 120 funds, however, 9 of these are dropped due to missing observations in Q4 2018. This first regression estimates the textual disclosure score's impact on sustainability performance. As denoted in table 7, the coefficient of 1.435 is statistically significant on a 5% level. The interpretation of this is that an increase of 1 in the textual ESG score results in an estimated 1.435 unit increase in the Morningstar ESG score. This can indicate that the level of sustainability disclosure in underlying firms is a possible predictor of sustainability performance. However, the explanatory power of our variable will tell more about the precision of any predictions.

Figure 6: Regression, ESG – MSR regressed on Textual



Combined environmental, social and governance scores from Q4 2018

Table 7: MSR regressed on Textual, Q4 2018

Regression Results

	Dependent variable:
	Morningstar ESG score
Textual ESG score	1.435**
	(0.597)
Constant	50.595***
	(0.527)
Observations	111
R^2	0.050
Adjusted R^2	0.042
Note:	**p<0.1; **p<0.05; ***p<0.01*

As seen from the regression plot in figure 6, the residuals around the regression line are large. Consequently, the R-squared value is also small. In other words, the explanatory power of our independent variable is low even though it provides significance proof that the coefficient is not equal to 0, and that there is a positive trend. This is a problem for the purpose of precise predictions using the regression. To further evaluate precision and the consequence of a low R-squared value, we look at prediction intervals. Table 8 shows the prediction intervals of the regression when the input textual ESG score is 0.47 which is the average score in our sample. The fit is 51.27, close to the average Morningstar ESG score in our sample. However, the prediction interval ranges from 41.91 to 60.63, which covers almost the entire range of Morningstar ESG scores in the sample apart from the upper tail.

Table 8: Prediction Intervals

1: Summary statistics - Textual ESG score

	Min	1st Quarter	Mean	3rd Quarter	Max
Textual ESG	-1.358	-0.142	0.470	0.998	2.254

2: Prediction interval - 95 % confidence interval: Input = 0.470

Fit	Lower	Upper
51.27	41.91	60.63

3: Summary statistics - Morningstar ESG score

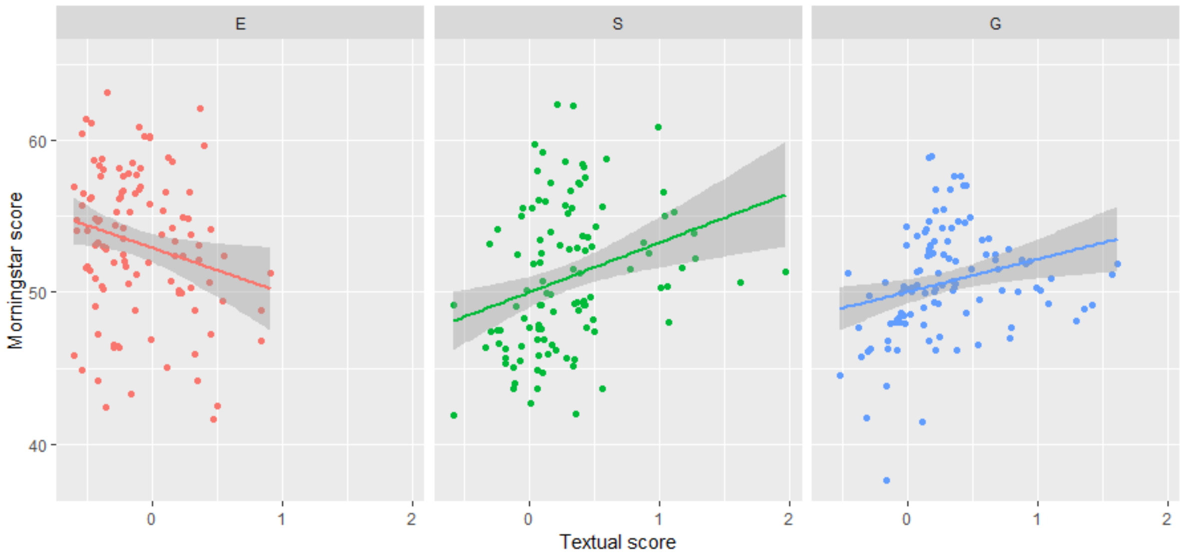
	Min	1st Quarter	Mean	3rd Quarter	Max
Morningstar ESG	41.28	48.25	51.39	54.90	65.10

Large predictions intervals and a low R-squared can partly be explained by the numerous aspects and industry-specific indicators that form the Morningstar ESG scores. Furthermore, our textual analysis of the 10-Ks might not capture all relevant sustainability disclosures. With this in mind, a positive coefficient which is significantly different from zero as seen in table 7, implies a positive relationship between our textual disclosure score and performance a higher Morningstar ESG score. However, for the purpose of using the textual disclosure score created in this thesis to predict ESG performance measured by Morningstar, low precision makes it unreliable.

6.1.3 Isolated E, S, and G scores

To further examine the previous regression, it is separated into three separate regressions based on the environmental, social and governance pillars. Figure 7 and table 9 exhibit the plot and results from the regression when divided into E, S and G scores, from the same period (Q4, 2018) as the previous regression. One interesting observation here is that the coefficient of the environmental MSR on the textual environmental score is now negative (-2.951), and still significant on a 5% level. Social and governance scores are still positive and significant.

Figure 7: Regression E, S and G - Morningstar vs Textual



Isolated E, S and G scores from 2018

Table 9: MSR regressed on Textual, Q4 2018

Regression Results

	Dependent variable:		
	Textual score		
	Environment	Social	Governance
	(1)	(2)	(3)
MS Environmental	-2.951**		
	(1.274)		
MS Social		3.267***	
		(1.005)	
MS Governance			2.147***
			(0.786)
Constant	52.923***	50.009***	50.021***
	(0.465)	(0.514)	(0.405)
Observations	111	111	111
R^2	0.047	0.088	0.064
Adjusted R^2	0.038	0.080	0.055
Note:			**p<0.1; ***p<0.05;*** ***p<0.01

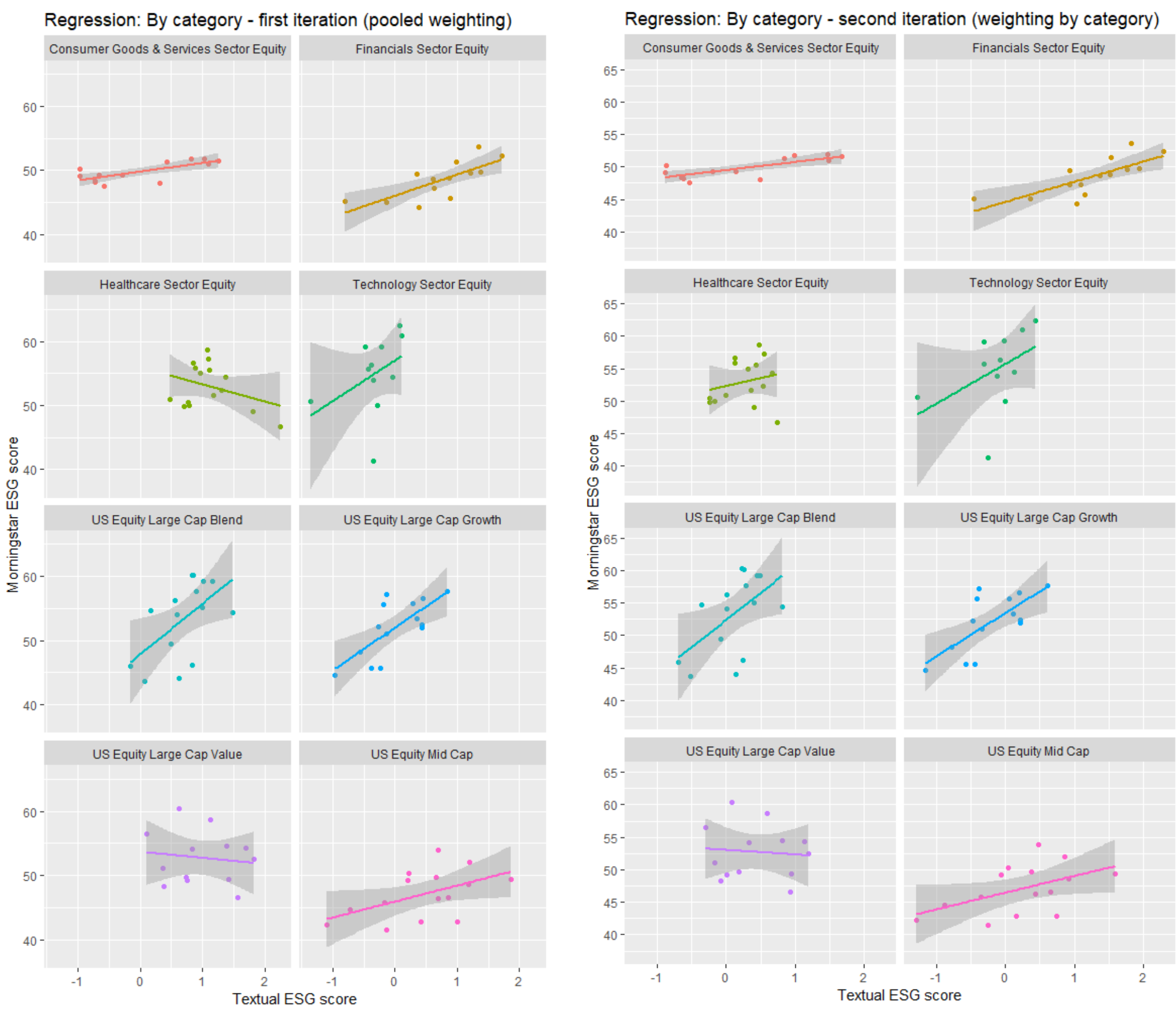
The negative environmental coefficient indicates that there is a tendency that funds performing better regarding environmental investment criteria hold companies that receive a lower textual ESG disclosure score. As previously mentioned, this might mean there are other environmental criteria that affect sustainability ratings, like carbon emission rates. It could also mean that there are some characteristics to environmental disclosure that makes a weighted frequency of environmental terms unsuited as a measure of true disclosure. Companies that underperform on Morningstar's environmental score could be exposed to more environmental risks, and thus report extensively on this topic. In these cases, it could be that the context around which environmental issues are reported in annual reports is more important than frequency. For the social and governance pillar of the scores, on the other hand, the coefficients are significant and positive. This could mean that ratings rely more on information found in company disclosure regarding these topics, and that the information retrieval process used to construct the textual ESG scores in this thesis are more effective in measuring true disclosure within these topics.

Some of the inaccuracy in the prediction of the overall ESG regression can probably be attributed to the opposing direction of the coefficient related to environmental topics. However, there is still a lot of unexplained variation in these regressions as well, which makes using them as predictors of sustainability performance within each pillar nearly as unreliable as the overall regression.

Next, we apply the same regression for each sector present in our sample of funds. This is done to reveal possible characteristics of sustainability and reporting between different investment categories and industries.

6.1.4 Results separated by categories

Figure 8: Before and after reweighting of textual scores, Q4 2018



The plots in figure 8 illustrate how our textual scores change when the "values" of terms are reweighted according to the specific sector in which the funds place their holdings. In other words, it shows how the results change between the first and second iteration of the code. This is obvious in for instance the healthcare sector, where all funds scored positive due to the frequent use of "social terms" before the adjustment. As a result, the coefficient in this category has flipped. Following the second iteration, the funds are centred closer to zero.

Table 10 shows the textual ESG score regressed on the Morningstar ESG score within each investment category. There are some interesting variations in results between categories. Significance levels, R-squared values and coefficients are different for each category. Some regressions show a higher significance than the overall regression (table 7), while others show no significance at all. Additionally, we observe a negative coefficient (Large cap value). However, this coefficient is not significant. R-square values also vary considerably but are generally larger than the one we observed in the overall regression.

Table 10: MSR regressed on Textual, Q4 2018, separated by category

Regression Results

	Dependent variable:							
	Morningstar Sustainability score							
	Consumer Goods and Services	Financial	Healthcare	Technology	US Large Cap Blend	US Large Cap Growth	US Large Cap Value	US Mid Cap
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Textual ESG score	1.242***	3.106***	2.426	6.116	8.356**	6.658***	-0.699	2.575*
	(0.299)	(0.771)	(2.935)	(4.041)	(3.442)	(1.927)	(2.403)	(1.240)
Constant	49.478***	44.603***	52.297***	55.743***	52.410***	53.469***	53.014***	46.457***
	(0.279)	(1.084)	(1.212)	(1.798)	(1.380)	(0.982)	(1.596)	(0.923)
Observations	14	14	15	11	15	14	13	15
R^2	0.590	0.575	0.050	0.203	0.312	0.499	0.008	0.249
Adjusted R^2	0.556	0.539	-0.023	0.114	0.259	0.457	-0.083	0.192
Note:								*p<0.1; **p<0.05; ***p<0.01

These differences can yet again imply that the information retrieval process we are using to construct the textual ESG is more effective within certain industries. Companies that operate within these sectors might disclose sustainability performance in a unique way or use specific terminology to address sustainability issues that are unique for that sector, i.e. use words that are not in our initial ESG dictionary. Some of the regressions have a much higher R-squared value than what we have seen previously. This could indicate that we might find some predictive power in the textual ESG score on sustainability performance within categories such as consumer goods and services or financial sector.

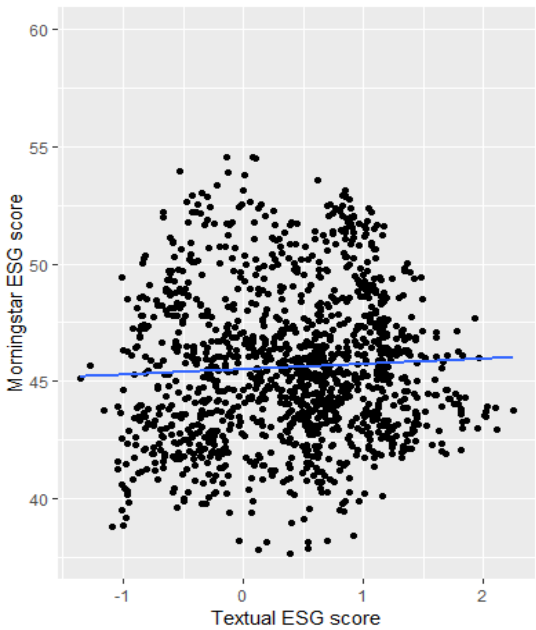
However, before interpreting these results further, we look to address the number of observations in each regression. As we have divided our observations into 8 different categories, we are only left with 11 to 15 observations for each regression. Therefore, to increase the number of observations and run a more robust regression, we include observations from all 12 time periods in the next regression.

6.1.5 Results pooled over time

When assessing the results from all 12 time periods in our sample, we can no longer use the Morningstar ESG score, but instead the full Morningstar sustainability score. The difference between the two is that the full sustainability score is adjusted for the controversy score that Morningstar gives each fund. The reason for this change is that historical data were only available for the full sustainability score.

Furthermore, when looking at the results from all time periods, the results are pooled together instead of using panel data methods. The reason for not using panel data is that plots show very little difference in coefficients over time, meaning the time period by itself has little to none effect on the coefficient results. These plots can be found in Appendix 2. We further explore the development of the results over time in research question 4.

Figure 9 and Table 11: MSR regressed on Textual - Pooled over time



Regression Results

	Dependent variable:
	Morningstar Sustainability score
Textual ESG score	0.227*
	(0.124)
Constant	45.502***
	(0.099)
Observations	1,343
R^2	0.002
Adjusted R^2	0.002
Note:	**p<0.1; **p<0.05; ***p<0.01*

Table 12: Rank correlation tests of textual and MSR, pooled over time

Rank correlation			
Kendall's τ		**Spearman's ρ**	
tau	0.039	rho	0.063
p-value	0.032	p-value	0.021
z	2.141	S	378240000

Figure 9 shows the regression plot when looking at all 12 time periods. With all 1343 observations included it is hard to visually identify any pattern or trend in the plot. Statistical tests also find weaker relationships than before. Observations fall even further from the regression line, and the R-squared demonstrates that the textual disclosure score yields practically no explanatory power. The coefficient is also only significant at the 10% level, which in many cases are regarded as insignificant. The results of the two correlation tests from table 12 are still significant, but the strength of the correlations are very weak. Our interpretation of this is that for the full data sample across all categories and ESG aspects there is only a weak, positive relationship between sustainability disclosure in underlying companies and sustainability performance. Additionally, the textual ESG score has virtually no prediction power on the sustainability performance of a fund. Note that heteroscedasticity was detected in the regression and corrected for using robust standard deviations, with little change in the results.

As a comparison, a related study by Wen (2014) found no significant relation between the unweighted frequency of sustainability words and the sustainability score. This study, however, reviews companies rather than funds and it does not examine the correlation by category. Therefore, the results are most comparable with our results in table 7 and 11 – where we only found a marginal correlation between our textual score and the Morningstar Sustainability Rating.

One reason for the lack of noteworthy results could be explained by variation in results across categories. Therefore, in the following regression, the results are again sorted by categories. Separation by the E, S and G pillars is not done because the historical Morningstar scores for each pillar is unavailable.

6.1.6 Results separated by categories, pooled over time

In the last regression of this research question, the funds are separated by category while observations are pooled over time. Figure 10 and the regression found in table 13 demonstrates that the assumed link between the level of sustainability of funds and the level of ESG-reporting among their holdings seem to vary according to which category they invest in. When the Morningstar sustainability score is regressed on the textual ESG diclosure score, the coefficients for the financial sector, large cap, large cap growth, and mid cap are positive and significant, with a slope varying from 0.768 to 4.549. Yet, we can observe from the p-values that the coefficients of consumer goods and services, tech and healthcare are insignificant. Only finance of the sector-specific categories exhibits a significant relationship.

Figure 10: Regression plot, by category

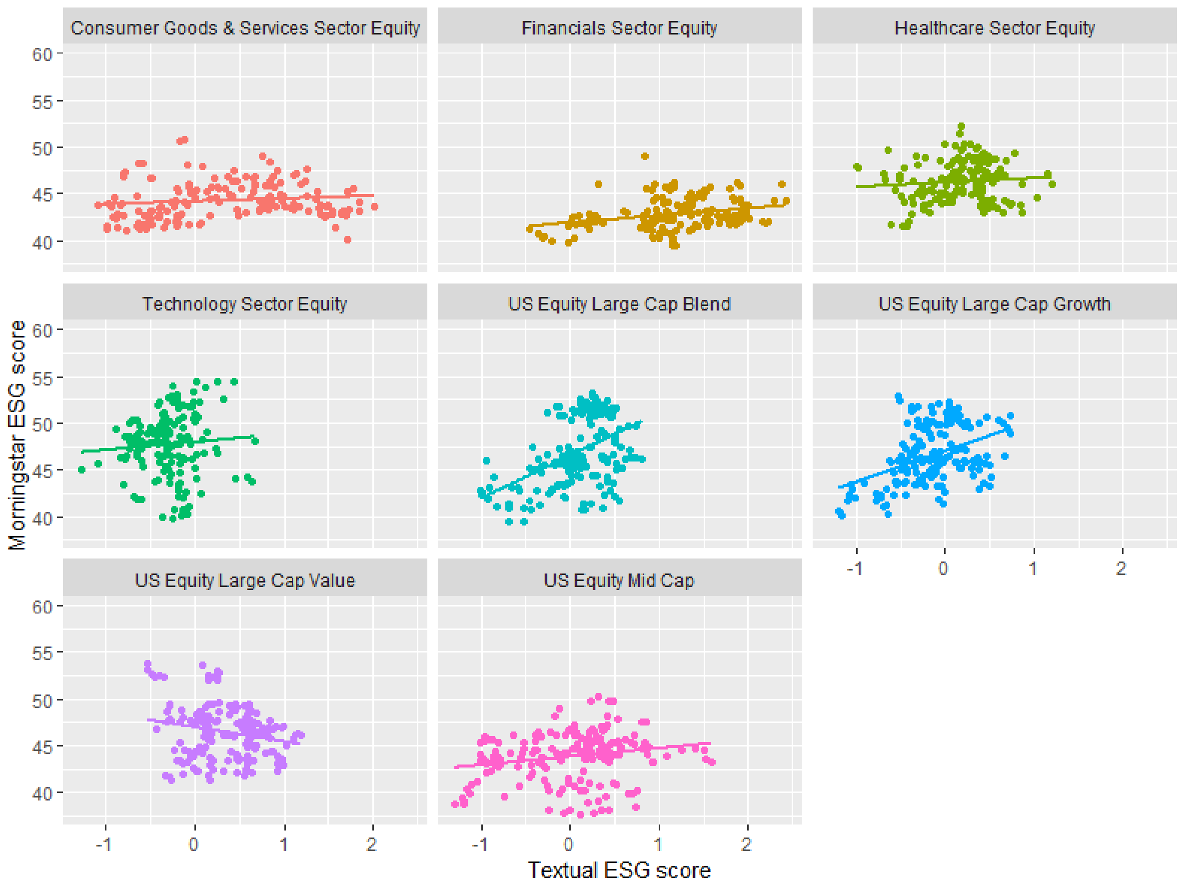


Table 13: MSR regressed on Textual, full sample, by category

Regression Results

	Dependent variable:							
	Morningstar Sustainability score							
	Consumer Goods and Services	Financial	Healthcare	Technology	US Large Cap Blend	US Large Cap Growth	US Large Cap Value	US Mid Cap
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Textual ESG score	0.290	0.768***	0.464	0.844	4.549***	3.263***	-1.534***	0.890***
	(0.188)	(0.198)	(0.421)	(0.846)	(0.612)	(0.523)	(0.543)	(0.318)
Constant	44.178***	41.898***	46.159***	47.955***	46.560***	47.027***	46.951***	43.830***
	(0.162)	(0.268)	(0.180)	(0.368)	(0.241)	(0.220)	(0.288)	(0.193)
Observations	167	159	165	154	180	174	165	179
R^2	0.014	0.087	0.007	0.007	0.237	0.185	0.047	0.042
Adjusted R^2	0.008	0.082	0.001	-0.00003	0.233	0.180	0.041	0.037
Note:								**p<0.1; ****p<0.05;*** ***p<0.01

Considering these results, it appears that that sustainability disclosure is a sector-specific concept. It might be that highly specific terms are missing from our ESG dictionary, or that terms in the dictionary are frequently used by companies in certain sectors to simply describe its operations and are in no way connected to its sustainability performance. Alternative measures could also be a source of variation. For companies that are more industrial, measures such as “percentage of materials recycled” or “percentage of energy consumption supplied by renewable sources” might be more important and better indicators of sustainability performance than disclosures in annual reports. For all four US large/mid cap categories, however, results are more robust. Funds in these categories tend to be more diversified across multiple sectors than the sector-specific funds. This could indicate that our textual ESG score benefits from diversification across sectors because it will diversify away some of the variability caused by specific terminology. In other words, using the textual ESG score as a predictor for sustainability performance within these categories yields a higher precision than what we see for more sector-specific categories. This can imply that the prediction precision of our computed ESG disclosure scores can be further improved by using dictionaries that are related to more specific industries.

In the regression in table 13, we observe that US Large Cap Value is the only category with a significant negative coefficient when the textual score is regressed on the MSR. We interpret this slope to imply that more disclosure in underlyings of this category, suggests a lower sustainability score, ceteris paribus.

This finding is consistent with a study performed on Australian mining and production companies that finds a positive correlation between the level of emissions and environmental disclosures (Clarkson et.al., 2011). The essence of value investing is to identify companies with low price-to-book ratios compared to their peers, as this is a potential indication of undervalued companies. Hence, judging by the nature of value investing, one would expect funds that operate in the US Large Cap Value-category to hold a relatively greater proportion of large companies with low price-to-book ratios. Basic minerals and energy were regarded as the biggest sectors of interest for value investors in 2018 (Morningstar, 2018b). When we looked into the underlyings among funds in this category, we found that Exxon, Chevron, AT&T, CVS Health and Pfizer are the five most common holdings.

One possible explanation for the negative correlation could be related to the legitimacy theory referred to previously. It could be that prominent underlyings in this category voluntary disclosure more than their peers, as a way to deflect attention and shift public perceptions. One could also argue that Exxon, Chevron, AT&T, CVS Health and Pfizer are all large companies exposed to political risk and/or with high consumer visibility – which in turn are more likely to disclose CSR information (Cuganesan et al., 2007). Furthermore, it could be that the activities of these companies are more exposed to ESG risks, meaning that they are obligated to disclose ESG-issues that are likely to have a material effect on their operations. All these elements can contribute to a higher score in our textual disclosure measure, without a corresponding change in the MSR. As our scores are computed on fund level, we are not able to examine each holding or to conclude on this matter.

6.2 Research Question 2: Difference in Scores

The purpose of this research question is to get a better understanding of what might cause the difference observed between the two scores.

Research question 2: *How does the difference (represented by delta) between textual disclosure score and sustainability score vary according to funds' characteristics?*

Table 14 contains the results from the regression fund characteristics regressed on the difference between our textual disclosure score and sustainability score. Again, both a regression for Q4 2018 and one for all time periods is performed. Delta represents the difference between Textual ESG score and Morningstar ESG score in Q4 2018, and the difference between Textual ESG score and Morningstar Sustainability score over all time periods.

Table 14: Fund characteristics regressed on delta of textual disclosure score and sustainability score.

Regression Results

	Dependent variable:	
	Delta	
	Q4 2018	All Time Periods
	(1)	(2)
Fund.Size (in million USD)	0.0001*	0.00001
	(0.00003)	(0.00001)
Num.of.Holdings	-0.0003	0.0002
	(0.001)	(0.0003)
Age (in days)	-0.00001	-0.00001
	(0.00002)	(0.00001)
Beta	1.021***	0.959***
	(0.386)	(0.120)
SmallBig	0.001	0.002***
	(0.001)	(0.0004)
ValueGrowth	-0.001	-0.0002
	(0.001)	(0.0002)
Constant	-0.161	-0.235
	(0.622)	(0.171)
Observations	83	1,278
R^2	0.127	0.063
Adjusted R^2	0.059	0.058
Note:	**p<0.1; *****p<0.05;*** ***p<0.01	

For both regressions the Beta is positive and statistically significant. This implies that funds with high volatility compared to the market index tend to have a larger difference between the disclosure score and sustainability score. Ergo, disclosure score as a predictor for sustainability performance is less efficient when dealing with more volatile funds, ceteris paribus. Additionally, our measure of small versus big investment style is positive statistically significant over all time periods. This would imply that funds with larger parts of their portfolio invested in bigger companies have a larger difference between the two scores. However, as previously mentioned this score is subject to selection bias with little representation of funds that score low. Thus, we are cautious in drawing any conclusion on this matter. It is worth mentioning that though the coefficient of Beta is significant, the explanatory power of the variables is low, with an R-squared of only 0.06 – implying that most of the delta is explained by other variables than these fund characteristics.

6.3 Research Question 3: Green labelled Funds

In the third research question, we examine if “green labelled” funds exhibit superior performance in the MSR and our textual disclosure score when compared with conventional funds.

Research question 3: *Is there a statistically significant difference between the identified “green labelled” funds and the conventional funds?*

To examine the third research question, results from the third iteration of the code are used. In this iteration of the code, 17 “green labelled” funds are identified, and their scores are compared to the scores of the rest of the funds in the sample. The first test, the t-test, compares the means of the two groups in one time period, Q4 2018. The results in table 15 suggest that there is a highly significant difference in measured sustainability between “green labelled” funds, with a mean Morningstar ESG score of 55.73 and conventional funds, with a mean score of 51.13. There is also a difference in means between “green labelled” funds and conventional funds in the disclosure score, but the difference is not statistically significant.

Table 15: T-test of conventional vs green labelled funds, Q4, 2018.

	Test statistic	df	P value	Mean Conventional Funds	Mean Green Label
Morningstar ESG score	-6.99	51.24	6.52e-09 * * *	51.13	55.73
Textual ESG score	-1.39	65.27	0.17	0.46	0.60

*Note: *p<0.1;* ***p<0.05;*** ****p<0.01*

Table 16: T-test of conventional vs green labelled funds, pooled.

	Test statistic	df	P value	Mean Conventional Funds	Mean Green Label
Morningstar sustainability score	-18.43	382.20	1.02e-54 * * *	45.52	48.62
Textual ESG score	-4.00	595.60	7.02e-05 * * *	0.39	0.50

*Note: *p<0.1;* ***p<0.05;*** ****p<0.01*

When a t-test is run for the whole sample with pooled time periods, the test finds evidence of a highly significant difference in both disclosure score and sustainability score between the groups. Note that the underlying company ESG-scores used by Morningstar are already industry adjusted. Therefore, it should not be the case that this difference is solely attributed to the industry in which these green funds are operating in.

In addition to pooling results from all time periods, we can compare the two groups in each time period separately. The outputs are exhibited in table 17 and 18 and visualised over time in figure 11. For some of the time periods we found violations of the normal distribution assumption. Therefore, a Wilcox rank sum test is used instead of e.g. t-test or ANOVA. The output shows the interaction between the two groups in each time periods. Similar to the results from the t-tests, we see from the Wilcox test that there is a significant difference in Morningstar Sustainability Rating between the "green labelled" and conventional funds across the time periods. This means that the MSR of "green labelled" funds is significantly higher than that of conventional funds over time. This can be visually confirmed in figure 9 by looking at the distance between the mean and error bars of each group.

These results indicate that funds in our sample that are labelled as green do in fact outperform conventional fund on ESG aspects, and thus they generally "practise what they preach". In other words, it looks like funds who take on the name of being green also tend to adopt a more sustainable investment strategy.

Table 17: Wilcox Rank Sum test of conventional vs green labelled funds: Morningstar Sustainability Rating

	Time	Test statistic	P value	Mean Conventional Funds	Mean Green Label
Morningstar Sustainability Rating	1	446	0.020 *	47.76	49.45
	2	379	0.004 * *	46.65	49.01
	3	329	0.001 * * *	45.68	48.46
	4	266	0.000 * * *	45.47	48.62
	5	249	0.000 * * *	45.13	48.43
	6	268	0.000 * * *	44.47	47.54
	7	265	0.000 * * *	44.49	47.58
	8	262	0.000 * * *	45.19	48.33
	9	229	0.000 * * *	45.16	48.86
	10	219	0.000 * * *	45.11	49.06
	11	222	0.000 * * *	44.94	48.99
	12	206	0.000 * * *	45.00	49.20

*Note: *p<0.1;* *****p<0.05;*** ****p<0.01*

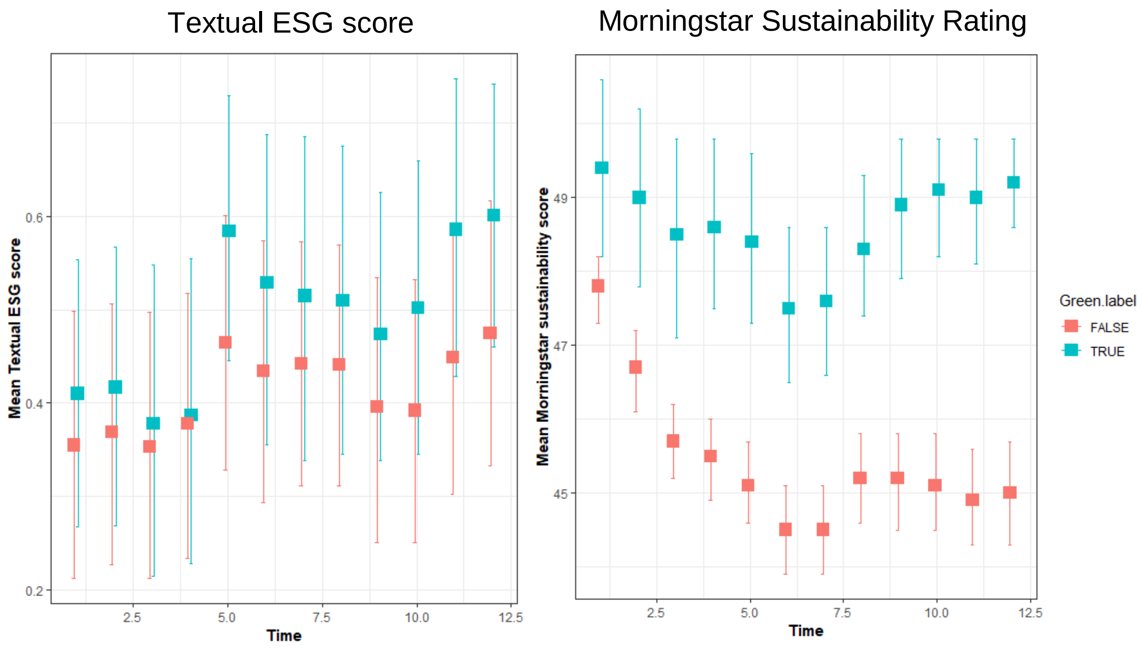
Table 18: Wilcox Rank Sum test of conventional vs green labelled funds: Textual ESG Score

	Time	Test statistic	P value	Mean Conventional Funds	Mean Green Label
Textual ESG Score	1	684	0.8073	0.3550	0.4103
	2	713	1.0000	0.3688	0.4171
	3	717	0.9722	0.3529	0.3784
	4	742	0.8006	0.3777	0.3866
	5	689	0.8412	0.4653	0.5844
	6	694	0.8754	0.4341	0.5293
	7	700	0.9168	0.4423	0.5150
	8	709	0.9792	0.4414	0.5100
	9	720	0.9514	0.3961	0.4745
	10	682	0.7939	0.3923	0.5022
	11	656	0.6257	0.4488	0.5864
	12	673	0.7341	0.4754	0.6006

*Note: *p<0.1;* *****p<0.05;*** ****p<0.01*

The differences in textual disclosure score between the groups are smaller, and only significant in the pooled t-test, which has more observations. From a visual analysis in figure 11, the mean textual score for "green labelled" funds appear to be consistently higher than that of the conventional funds. However, the difference is not substantial, and the error bars are consistently overlapping. The Wilcox test, exhibited in table 18, confirms that there is no significant difference in textual ESG score between "green labelled" and conventional funds.

Figure 11: Conventional vs green labelled funds



6.4 Research Question 4: Trend in Disclosures

The last research question focuses on the trend of written ESG-disclosure in annual reports.

Research question 4: *Has there been an uptake of ESG-disclosures in 10-Ks in recent years?*

Figure 12 visualises the development in the mean textual score and mean MSR-score over the 12 quartiles included in the sample period, from Q1 2016 to Q4 2018. By visual inspection, there seems to be a moderate uptake in the mean textual score during the sample period. However, this increase is not statistically significant and the R-squared is negligible. On the contrary, the plot and regression to the right confirm that funds in our sample on average have a declining MSR performance over the sample period. The coefficient of -0.162 is significant on a 1% level, while the R-squared is still only 0.031.

Figure 12: Textual score vs Morningstar Sustainability Rating, quarterly.

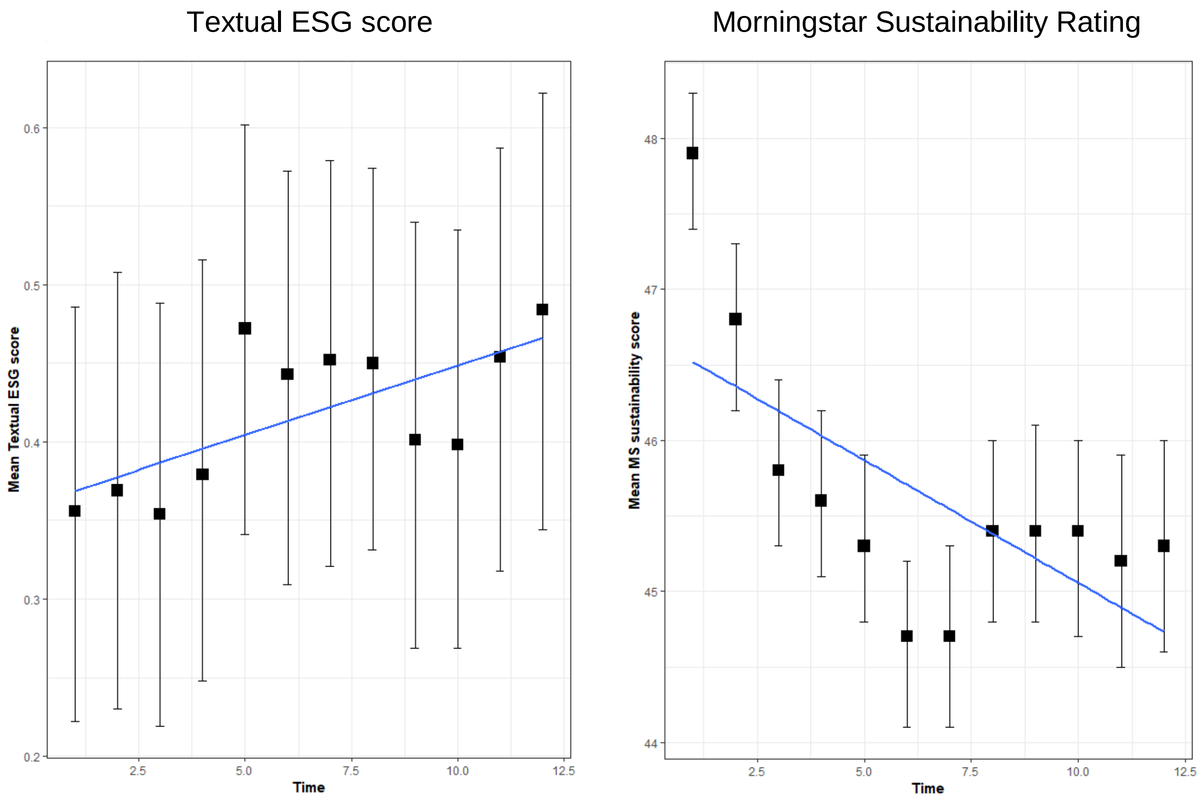


Table 19: Textual Score and MSR regressed on time

Regression Results

	Dependent variable:	
	Textual ESG score	MS sustainability score
	(1)	(2)
Time	0.009	-0.162^{***}
	(0.006)	(0.026)
Constant	0.360^{***}	46.678^{***}
	(0.042)	(0.193)
Observations	1,200	1,200
R^2	0.002	0.031
Adjusted R^2	0.001	0.030
Note:		*p<0.1; **p<0.05; ***p<0.01

Since we in this research question are more interested in the trend of ESG-disclosure in annual 10-K reports rather than comparisons of funds, we also examined the individual 10-K filings. By observing the filings directly, we avoid overlooking potential trends that might not be reflected in the textual score due to shifts in holdings or varying term weighting. Therefore, we performed an examination of each report to see if the results aligned with the textual score.

Table 20 exhibit the average absolute frequency of ESG terms in all 5 556 annual 10-Ks included in the sample. The average ESG term frequencies seem to be reflected also in the textual ESG score. Ultimately, the conclusion remains that an uptake in ESG-disclosure from 2016-2018 is not identified.

Table 20: Average ESG term frequency in 10-Ks

Average ESG term frequency in 10-Ks – by year			
	2016	**2017**	**2018**
Environment	119	113	112
Social	264	272	271
Governance	2 133	2 201	2 147
ESG	2 516	2 586	2 530

Lastly, we look for differences in disclosure development over time between categories. The plot in figure 13 shows some moderate differences in the direction of the trends. Results from the regressions shown in table 21 tell us there is just a few that show a significant trend. All significant trends are positive.

Figure 13: Textual disclosure score over time, separated by category

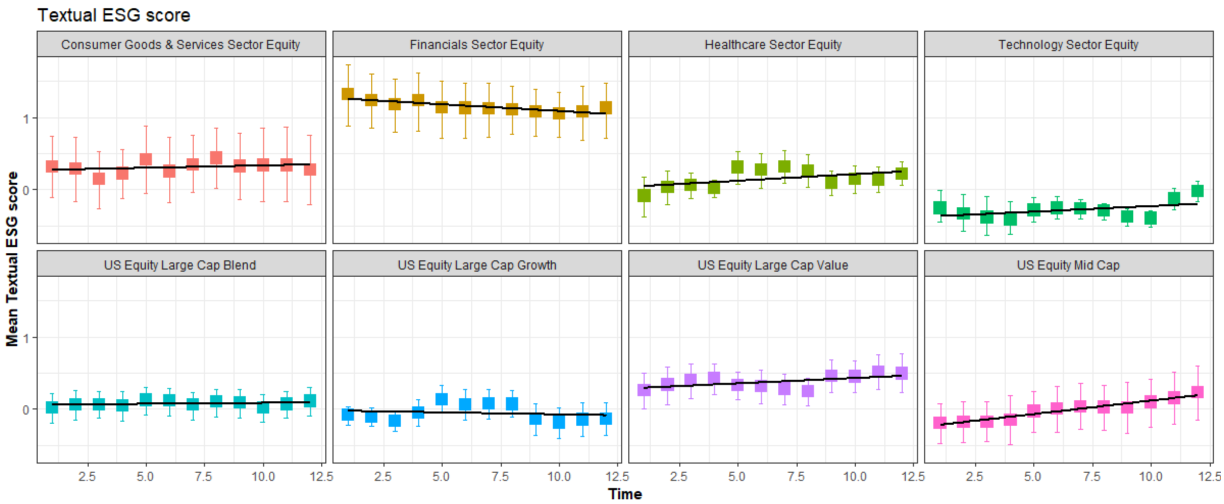


Table 21: Textual Disclosure score Regressed on time

Regression Results

	Dependent variable:							
	Textual ESG score							
	Consumer Goods and Services	Financial	Healthcare	Technology	US Large Cap Blend	US Large Cap Growth	US Large Cap Value	US Mid Cap
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Time	0.007	-0.019	0.018**	0.015*	0.003	-0.005	0.015	0.037***
	(0.019)	(0.016)	(0.009)	(0.008)	(0.008)	(0.009)	(0.010)	(0.014)
Constant	0.266*	1.271***	0.030	-0.385***	0.058	-0.020	0.281***	-0.254**
	(0.142)	(0.118)	(0.067)	(0.060)	(0.062)	(0.063)	(0.074)	(0.101)
Observations	156	132	156	120	180	144	144	168
R^2	0.001	0.010	0.025	0.029	0.001	0.003	0.015	0.042
Adjusted R^2	-0.006	0.003	0.019	0.021	-0.005	-0.004	0.008	0.036
Note:								**p<0.1; ****p<0.05; ******p<0.01

Seeing the current attention around ESG-issues and the surge of SRI-investing, we expected to find a certain uptake of disclosure during our sample period. Still, there may be numerous explanations why such a trend is not identified in our sample. One possible factor is that companies try to accommodate investors increasing disclosure demand by using other channels than the 10-Ks, such as stand-alone sustainability reports. The KPMG study that indicated more disclosure among the largest US firms were based on financial reports rather than 10-Ks. Financial reports are less comprehensive than 10-Ks as it is intended to contain only the most relevant information for investors. Therefore, companies might disclose more on ESG-issues in annual financial reports now than before because they consider the topic to be more relevant. When our results indicate a stable level of disclosure in 10-K filings, this might reflect that the regulation of non-financial reporting in the US has remained unchanged in this period. It is also possible that most US companies already in 2016 addressed ESG issues in their 10-Ks, and therefore a trend is not identified in our relatively short time frame.

The cause of the significant declining MSR performance remains uncertain. From the beginning, our sample holds funds that with an average MSR score of around 48, while the stated mean of all funds according to Morningstar's methodology is 50 (Morningstar, 2018a). However, judging by the summary statistics in table 2, our final sample is a close representation of the qualified funds in this regard. Figure 11 demonstrates that the downward trend is mainly driven by declining performance among the conventional funds. One potential explanation for this decline in sustainability performance among conventional funds their relative performance is affected when new funds are launched. Our sample only holds funds that exist before 2016, thus any introduction of new, more sustainable funds will alter the MSR of existing funds, as the score is partly relative to other funds. Further, we observe that the declining performance was mostly related to Q1 and Q2 2016. The oil price dropped by 50% in the second half of 2015, before it started to climb from January 2016 (Bloomberg, 2019). This event has a potential influence on the holdings of funds in this sample.

Regarding the time trends of the textual ESG score across each category, there is a significant positive trend in the healthcare and US mid cap categories, as well as a weak, significant (<10%) positive trend for the technology category. In previous sections, we have deliberated on potential ineffectiveness in the information retrieval process in both the healthcare and technology sector investment categories. Any such ineffectiveness does in that case also affect the result of the time trend for these categories. Thus, we are cautious in interpreting these results as significant time trends. For the US mid cap category however, results have been more robust, and we interpret this as a rise in sustainability disclosure among companies with a market capitalization between 2 billion and 10 billion USD. An explanation for this could be that in the past decade sustainability disclosure has become common for large and multinational companies, while small and medium-sized companies have been playing catch up the last three years and are focusing more on becoming sustainable, or at least disclosing this activity more in their annual reports (Global Reporting Initiative, 2016) .

7. Conclusion

The main purpose of this thesis is to examine if there exists a relationship between transparency and sustainable performance in US mutual funds. To do this, we create a tool that can retrieve, structure and analyse textual data from several hundred thousand pages of annual 10-K reports. The information retrieved from this process is used to construct scores for funds that are intended to represent the overall disclosure of sustainability aspects in underlying companies. Building on this tool, we thereafter compare the sustainability performance of conventional funds versus “green labelled” funds. In our final research question, we alter the code in an effort to observe potential trends in disclosure levels throughout the sample period from Q1 2016 to Q4 2018.

From our results, we conclude that there is a relationship between the level of disclosure in annual 10-Ks and sustainability performance. Some results, with statistical significance, suggest that funds that perform better on the textual disclosure score, also tend to receive a higher sustainability score. The results vary depending on the time frame and whether we look at the environmental, social or governance aspect of sustainability. While the textual disclosure score exhibits explanatory power for certain categories, it is not a precise predictor of sustainability performance. For consumer goods and services, technology, and healthcare, no relation between the variables is identified. We identify two possible explanations for this lack of relationship. (1) In many cases, there is no relationship between sustainability performance of funds and the disclosure rates of underlying companies. This would suggest that disclosure in annual reports is not reflected in sustainability performance as measured by Morningstar. (2) Aspects of the information retrieval process used to construct textual ESG disclosure scores in this thesis are limited and inefficient under certain circumstances. Consequently, the tool might not identify and capture sector-specific disclosure standards, or even the nature of environmental disclosure methods used by companies. Other limitations to consider can be that the Morningstar Sustainability Rating is not an accurate representation of true sustainability performance, or that sustainability activities are by many companies disclosed more extensively through other channels than their annual report. As a result, the textual disclosure score is an unprecise predictor of a fund’s sustainability performance in most cases. However, future improvements to the information retrieval process may lead to more precise predictions.

In an effort to explain the difference between the two scores, we find that riskier funds, measured by volatility against the market return, tend to have larger differences in scores, and therefore also less accurate predictions. Further, the findings suggest that "green labelled" funds practice what they preach concerning sustainability performance. However, we cannot conclude that "green labelled" have a higher level of disclosures than conventional funds. Additionally, contrary to our own expectations, we were not able to identify a positive time trend suggesting more disclosure of sustainability aspects through annual reports in later years. These results are subject to the same possible explanations as our first research question. If disclosure in annual reports does not influence ratings, there would naturally be no real difference between "green labelled" and conventional funds. Additionally, any difference in disclosure rates or time trend would be hard to identify if the information retrieval process is insufficient.

Through this thesis, we contribute to the topic by bringing aspects of information retrieval systems and more specifically textual mining methods together with sustainability research to create a tool for analysing funds. This tool builds on concepts in earlier literature and brings them together to lay the grounds for a standard quantifiable measure for largely unstructured bodies of text. It allows the user to process the large amounts of data needed to analyse funds consisting of hundreds of companies and thousands of pages of information in annual reports.

The tool is likely subject to some major limitations relating to sector-specific disclosure, as well as some aspects regarding environmental disclosure. For further research, we suggest improving the tool by looking into these limitations. One suggestion is to include a sentiment analysis aspect. Another is to investigate sustainability disclosure specifically related to environmental issues, or to create a dictionary that takes the characteristics of each category into consideration when constructing the textual ESG score. We hope to see more studies on these issues in the future.

8. References

Allen, K. (2018). Lies, damned lies and ESG rating methodologies. *Financial Time*. Available at: https://ftalphaville.ft.com/2018/12/06/1544076001000/Lies--damned-lies-and-ESG-rating-methodologies/

Aureli, S. (2017). A comparison of content analysis usage and text mining in CSR corporate disclosure. *The International journal of digital accounting research, 2017*(V17).

Aureli, S., Medei, R., Supino, E., & Travaglini, C. (2016). Sustainability Disclosure after a Crisis: A Text Mining Approach. *International Journal of Social Ecology and Sustainable Development (IJSESD), 7*(1), 35-49.

Baier, P., Berninger, M., & Kiesel, F. (2018). Environmental, Social and Governance Reporting in Annual Reports: A Textual Analysis. doi: http://dx.doi.org/10.2139/ssrn.3206751

Blasco, J., & King, A. (2017). The road ahead. *KPMG*. Retrieved from https://assets.kpmg/content/dam/kpmg/xx/pdf/2017/10/kpmg-survey-of-corporate-responsibility-reporting-2017.pdf

Bloomberg. (2019). CO1:COM, Generic 1st 'CO' Future. Available at: https://www.bloomberg.com/quote/CO1:COM

Breedt, A., Ciliberti, S., Gualdi, S., & Seager, P. (2018). Is ESG an Equity Factor or Just an Investment Guide? Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3207372

Carterette, B. (2009). On rank correlation and the distance between rankings. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval* (pp. 436-443). ACM. doi: https://doi.org/10.1145/1571941.1572017

Ching, H. Y., Gerab, F., & Toste, T. H. (2014). Scoring Sustainability Reports using GRI indicators: A Study based on ISE and FTSE4Good Price Indexes. *Journal of Management Research, 6*(3), 27-48.

Cho, C. H., Roberts, R. W., & Patten, D. M. (2010). The language of US corporate environmental disclosure. *Accounting, Organizations and Society, 35*(4), 431-443. doi: https://doi.org/10.1016/j.aos.2009.10.002

Chvatalová, Z., Kocmanová, A., & Dočekalová, M. (2011). Corporate sustainability reporting and measuring corporate performance. *International Symposium on Environmental Software Systems* (pp. 245-254). Springer, Berlin, Heidelberg. doi: https://doi.org/10.1007/978-3-642-22285-6_27

Ciciretti, R., Dalò, A., & Dam, L. (2017). The Price of Taste for Socially Responsible Investment. *CEIS Working Paper No. 413.* Available at SSRN: https://ssrn.com/abstract=3010234

Clarkson, P., Overell, M., & Chapple, L. (2011). Environmental Reporting and Its Relation to Corporate Environmental Performance. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1751767

Cuganesan, S., Ward, L., & Guthrie, J. (2007). Legitimacy theory: A story of reporting social and environmental matters within the Australian food and beverage industry. *5th Asian Pacific Interdisciplinary Research in Accounting (APIRA) Conference*, (8-10).

Dolvin, S. D., Fulkerson, J. A., & Krukover, A. (2017). Do 'Good Guys' Finish Last? The Relationship between Morningstar Sustainability Ratings and Mutual Fund Performance. Available at SSRN: https://ssrn.com/abstract=3019403

Doyle, T. M. (2018) The Big Problem With 'Environmental, Social And Governance' Investment Ratings? They're Subjective. *American Council for Capital Formation (ACCF).* Available at: http://accf.org/2018/08/12/the-big-problem-with-environmental-social-and-governance-investment-ratings-theyre-subjective/

Dubbink, W., Graafland, J., & van Liedekerke, L. (2008). CSR, transparency and the role of intermediate organisations. *Journal of Business Ethics, 82*(2), 391-406. doi: http://dx.doi.org/10.1007/s10551-008-9893-y

Eccles, R. G., Krzus, M. P., Rogers, J., & Serafeim, G. (2012). The need for sector-specific materiality and sustainability reporting standards. *Journal of Applied Corporate Finance, 24*(2), 65-71. doi: https://doi.org/10.1111/j.1745-6622.2012.00380.x

Eccles, R.G., Ioannou, I., and Serafeim, G. 2014. The Impact of Corporate Sustainability on Organizational Processes and Performance. *Management Science*, 60(11), 2835-2857. doi: https://doi.org/10.1287/mnsc.2014.1984

European Commission. (n.d.) Non-financial reporting. Available at: https://ec.europa.eu/info/business-economy-euro/company-reporting-and-auditing/company-reporting/non-financial-reporting_en

Eurosif. (n.d.). Responsible Investment Strategies. Available at: http://www.eurosif.org/responsible-investment-strategies/

EY. (2017). *2017 SEC annual reports — Form 10-K*. Available at: https://www.ey.com/Publication/vwLUAssets/SECAnnualReports10K_06546-171US_21November2017/$FILE/SECAnnualReports10K_06546-171US_21November2017.pdf

Ezeokoli, O., Layne, C., Statman, M., & Urdapilleta, O. (2017). ESG Investment Tools: A Review of the Current Field (pp. 36, 65, 89-91). Summit Consulting. Available at: https://www.dol.gov/asp/evaluation/completed-studies/ESG-Investment-Tools-Review-of-the-Current-Field.pdf

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics, 33*(1), 3-56. doi: https://doi.org/10.1016/0304-405X(93)90023-5

Fernandez, G. C. (1992). Residual analysis and data transformations: important tools in statistical analysis. *HortScience, 27*(4), 297-300. Available at: https://journals.ashs.org/downloadpdf/journals/hortsci/27/4/article-p297.pdf

Fernandez-Feijoo, B., Romero, S., & Ruiz, S. (2014). Effect of stakeholders' pressure on transparency of sustainability reports within the GRI framework. *Journal of business ethics, 122*(1), 53-63. doi: https://doi.org/10.1007/s10551-013-1748-5

Frey, B. (2018). *The SAGE encyclopedia of educational research, measurement, and evaluation (Vols. 1-4).* Thousand Oaks, CA: SAGE Publications, Inc. doi: https://dx.doi.org/10.4135/9781506326139

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, *5*(4), 210-233., doi: https://doi.org/10.1080/20430795.2015.1118917

Global Reporting Initiative. (2016). Making the case for SME Sustainability Reporting. Available at: https://www.globalreporting.org/information/news-and-press-center/Pages/Small-Business,-Big-Impact-Making-the-case-for-SME-Sustainability-Reporting.aspx

Global Sustainable Investment Alliance. (2016). Global Sustainable Investment Review 2016 (pp. 8-9). Available at: http://www.gsi-alliance.org/wp-content/uploads/2017/03/GSIR_Review2016.F.pdf

Hale, J. (2016). How Sustainalytics Does Company ESG Research. *Yahoo Finance.* Available at: https://ca.finance.yahoo.com/news/sustainalytics-does-company-esg-research-100000652.html?guccounter=1

Hamilton, S. Jo, H., & Statman, M. (1993). Doing Well while Doing Good? The Investment Performance of Socially Responsible Mutual Funds. *Financial Analysts Journal, 49*(6), 62-66. doi: https://doi.org/10.2469/faj.v49.n6.62

Henke, H. M. (2016). The effect of social screening on bond mutual fund performance. *Journal of Banking & Finance, 67*, 69-84. doi: https://doi.org/10.1016/j.jbankfin.2016.01.010

Hooks, J., & van Staden, C. J. (2011). Evaluating environmental disclosures: The relationship between quality and extent measures. *The British Accounting Review, 43*(3), 200-213. doi: https://doi.org/10.1016/j.bar.2011.06.005

Horuckova, M., & Baudasse, T. (2017). Content analysis applied to social and environmental reporting. *Acta academica karviniensia*, 32-45.

Humphrey, J. E., & Tan, D. T. (2014). Does it really hurt to be responsible?. *Journal of business ethics, 122*(3), 375-386. doi: 10.1007/s10551-013-1741-z

IRRCI. (2018). State of Sustainability and Integrated Reporting. Available at: https://www.weinberg.udel.edu/IIRCiResearchDocuments/2018/11/2018-SP-500-Integrated-Reporting-FINAL-November-2018-1.pdf

J.P. Morgan. (2018). *Sustainable Investing is Moving Mainstream*. Available at: https://www.jpmorgan.com/global/research/esg

Jurafsky, D., & Martin, J. (2009). *Speech and language processing* (p. 801-806). Upper Saddle River (New Jersey): Prentice Hall.

Koellner, T., Suh, S., Weber, O., Moser, C., & Scholz, R. W. (2008). Do conventional and sustainability investment funds differ in their environmental impacts? – A comparison by means of Input–Output Life Cycle Assessment. *J. Indust. Ecol*, *11*(3), 41-60. doi: https://doi.org/10.1162/jiec.2007.1147

Kolstad, I., & Wiig, A. (2009). Is transparency the key to reducing corruption in resource-rich countries?. *World development*, *37*(3), 521-532. doi: https://doi.org/10.1016/j.worlddev.2008.07.002

Lonare, G. & Patil, B. (2017). edgar: Platform for EDGAR Filing Management. R package version 1.0.9. Available at: https://CRAN.R-project.org/package=edgar

Louche, C., & Hebb, T. (2014). Socially responsible investment in the 21st century : does it make a difference for society?. Bingley, UK: Emerald Group Publishing Limited: Available at: https://ebookcentral.proquest.com/lib/nhh-ebooks/reader.action?docID=1712212

Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35-65. doi: https://doi.org/10.1111/j.1540-6261.2010.01625.x

Marriage, M. (2017). Morningstar intensifies competition with asset managers. *Financial Times.* Available at: https://www.ft.com/content/b9dcfc38-d355-11e6-b06b-680c49b4b4c0

Matloff, R., & Chaillou, J. (2013). Nonprofit investment and development solutions: a guide to thriving in today's economy (Vol. 247). John Wiley & Sons. Available at: https://ebookcentral.proquest.com

Mooney, A. (2018). Rising investor interest pushes ESG funds past $1tn. *Financial Times.* Available at: https://www.ft.com/content/f1e98ec7-083e-3b95-8c6b-ecc4810b988e

Morgan Stanley Institute for Sustainable Investing. (2017). *Sustainable Signals* (pp. 2-10). Available at: https://www.morganstanley.com/pub/content/dam/msdotcom/ideas/sustainable-signals/pdf/Sustainable_Signals_Whitepaper.pdf

Morningstar. (2008). Morningstar Style Box™ Methodology. *Morningstar Research.* Available at: https://hk.morningstar.com/ODS_Images/Morningstar_StyleBox_Methodology.pdf

Morningstar. (2016a). Sustainable Investing Research Suggests No Performance Penalty. *Morningstar Research.* Available at: https://video.morningstar.com/ca/170717_SustainableInvesting.pdf

Morningstar. (2016b). Morningstar Sustainability Rating. Available at: https://www.morningstar.com/articles/745467/morningstar-sustainability-rating.html

Morningstar. (2018a). Morningstar Sustainability Rating – Methodology. *Morningstar Research*. Available at: https://www.morningstar.com/content/dam/marketing/shared/research/methodology/744156_Morningstar_Sustainability_Rating_for_Funds_Methodology.pdf

Morningstar. (2018b). Which Sectors are Value Investors Favouring? Available at: http://www.morningstar.co.uk/uk/news/166642/which-sectors-are-value-investors-favouring.aspx

MSCI. (2016). MSCI Global Sustainability Indexes Methodology. *MSCI*. Available at: https://www.msci.com/eqb/methodology/meth_docs/MSCI_Global%20Sustainability_Indexes_Methodology_June2016.pdf

Nelson, M. (2019). Why it's time to take a standard approach to nonfinancial reporting. *EY* Available at: https://www.ey.com/en_gl/assurance/time-to-take-a-standard-approach-to-nonfinancial-reporting

O'Donovan, G. (2000). *Legitimacy theory as an explanation for corporate environmental disclosures* (Doctoral dissertation, Victoria University of Technology). Available at: http://vuir.vu.edu.au/15372/1/O'Donovan_2000compressed.pdf

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology, 17*(4), 688-690. doi: https://doi.org/10.1093/beheco/ark016

SASB. (2016). *The State of Disclosure*. Available at: http://library.sasb.org/wp-content/uploads/2016/12/StateofDisclosure-Report-112916-EXCERPT.pdf?hsCtaTracking=5abd902f-74eb-41d1-b96e-36fa237d2bd1%7C8ff674de-3d9c-4ce6-9f69-ff8ea278a919

Shahnaz, F., Berry, M. W., Pauca, V. P., & Plemmons, R. J. (2006). Document clustering using nonnegative matrix factorization. *Information Processing & Management, 42*(2), 373-386. doi: https://doi.org/10.1016/j.ipm.2004.11.005

Sustainalytics. (n.d.). ESG Ratings & Research. Available at: https://www.sustainalytics.com/esg-ratings/#1530569132662-3e9e8929-5bee

Te Liew, W., Adhitya, A., & Srinivasan, R. (2014). Sustainability trends in the process industries: A text mining-based analysis. *Computers in Industry, 65*(3), 393-400. doi: https://doi.org/10.1016/j.compind.2014.01.004

Thompson, J. (2019) ESG rating agencies fulfil the need for knowhow. *Financial Times*. Available at: https://www.ft.com/content/2cd37df8-a973-3f94-b498-09ee1a6ba53b

Umemura, K., & Church, K. W. (2000). Empirical term weighting and expansion frequency. In *Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora: held in conjunction with the*

38th Annual Meeting of the Association for Computational Linguistics-Volume 13 (pp. 117-123). Association for Computational Linguistics. doi: https://doi.org/10.3115/1117794.1117809

US Securities and Exchange Commission (SEC). (2011). *How to Read a 10-K*. Available at: https://www.sec.gov/fast-answers/answersreada10khtm.html

US Securities and Exchange Commission (SEC). (2019). *Information for EDGAR Filers.* Available at: https://www.sec.gov/page/infoedgarshtml

US SIF Foundation. (2018). Report on US Sustainable, Responsible and Impact Investing Trends. Available at: https://www.ussif.org/files/Trends/Trends%202018%20executive%20summary%20FINAL.pdf

Van den Bogaerd, M., & Aerts, W. (2011). Applying machine learning in accounting research. *Expert Systems with Applications, 38*(10), 13414-13424. doi: https://doi.org/10.1016/j.eswa.2011.04.172

van Steenis, H. (2019). Defective data is a big problem for sustainable investing. *Financial Times*. Available at: https://www.ft.com/content/c742edfa-30be-328e-8bd2-a7f8870171e4

Wen, J. (2014). A Business Analytics Approach to Corporate Sustainability Analysis. Available at: https://repository.upenn.edu/mes_capstones/62/

Wigglesworth, R. (2018). Rating agencies using green criteria suffer from ‘inherent biases’. *Financial Times.* Available at: https://www.ft.com/content/a5e02050-8ac6-11e8-bf9e-8771d5404543

Wooldridge, J. (2016). Introductory econometrics. Boston: Cengage Learning.

Xu, W., Liu, X., & Gong, Y. (2003, July). Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 267-273). ACM. doi: https://doi.org/10.1145/860435.860485

Zsolnai, L. (Ed.). (2012). Handbook of business ethics: Ethics in the new economy. Available at: https://ebookcentral.proquest.com

9. Appendix

9.1 Appendix 1: Code

```
rm(list = ls())
setwd()

require(tidyr)
require(tm)
require(edgar)
# Code surgery edgar #
x <- capture.output(dput(edgar::getFilings))
x <- gsub("choice <- .*", "cat(paste(msg3, '\n')); choice <- 'yes'", x)
x <- gsub("^function", "my_getFilings <- function", x)
writeLines(x, con = tmp <- tempfile())
source(tmp)
# 'Not in' function
'%!in%' <- function(x,y)!('%in%'(x,y))

###############   Read fund data    #################
# Import all parts of dataset
data.first <- read.csv(file = "OpenEnd_first_CSV.csv",
                       sep = ";", header = T, dec = ",", na.strings = c(""),
                       stringsAsFactors = F)

data.second <- read.csv(file = "OpenEnd_second_CSV.csv",
                        sep = ";", header = T, dec = ",", na.strings = c(""),
                        stringsAsFactors = F)

data.third <- read.csv(file = "OpenEnd_third_CSV.csv",
                       sep = ";", header = T, dec = ",", na.strings = c(""),
                       stringsAsFactors = F)
# Combine data
data <- rbind(data.first, data.second, data.third)
# Clear Environment
rm(data.first, data.second, data.third)
#  Extract data selection
df <- data.frame(Name = data$Name,
                 ISIN = data$ISIN,
                 Ticker = data$Ticker,
                 Category = data$Global.Category,
                 Date = data$Portfolio.Date,
                 Sustainability.rating = data$Morningstar.Sustainability.Rating,
                 Portfolio.sustainability.score = data$Portfolio.Sustainability.Score,
                 Historical.sustainability.score = data$Historical.Sustainability.Score,
                 ESG.score = data$Portfolio.ESG.Score,
                 Environmental.score = data$Portfolio.Environmental.Score,
                 Social.score = data$Portfolio.Social.Score,
                 Governance.score = data$Portfolio.Governance.Score,
                 Num.of.Holdings = data$X..of.Holdings..Long.,
                 Score.2016.Q1 = data$Portfolio.Sustainability.Score.2016.03,
                 Score.2016.Q2 = data$Portfolio.Sustainability.Score.2016.06,
                 Score.2016.Q3 = data$Portfolio.Sustainability.Score.2016.09,
                 Score.2016.Q4 = data$Portfolio.Sustainability.Score.2016.12,
                 Score.2017.Q1 = data$Portfolio.Sustainability.Score.2017.03,
                 Score.2017.Q2 = data$Portfolio.Sustainability.Score.2017.06,
                 Score.2017.Q3 = data$Portfolio.Sustainability.Score.2017.09,
                 Score.2017.Q4 = data$Portfolio.Sustainability.Score.2017.12,
                 Score.2018.Q1 = data$Portfolio.Sustainability.Score.2018.03,
                 Score.2018.Q2 = data$Portfolio.Sustainability.Score.2018.06,
                 Score.2018.Q3 = data$Portfolio.Sustainability.Score.2018.09,
                 Score.2018.Q4 = data$Portfolio.Sustainability.Score.2018.12,
                 stringsAsFactors = F)

###############   Screening process   #############
# Complete historical data
```

```
df <- df[complete.cases(df[,c(14:25)]),]
# Remove duplicates
df <- df[!duplicated(df[c(7,13)]),]
# Convert holding count to numeric
df$Num.of.Holdings <- gsub(" ", "", df$Num.of.Holdings)
df$Num.of.Holdings <- as.numeric(df$Num.of.Holdings)
# Unite Ticker and ISIN for fund ID
df <- unite(df, "ID", c(2,3), sep = "-", remove = F)
df$ID <- gsub("-NA|NA-", "", df$ID)
# Create table of most common categories
category.table <- as.data.frame(table(df$Category), stringsAsFactors = F)
category.table <- category.table[order(category.table$Freq, decreasing = T),]

###############   Find selection of funds   ##############
# select top categories
top.categories <- category.table$Var1[1:8]

selection <- character() # Pre-allocation for results
# In each broad category identify top/bottom/mid 5 funds
for (i in top.categories){
  df.selection <- df[which(grepl(i, df$Category)),]
  df.selection <- df.selection[order(df.selection$Portfolio.sustainability.score,
                                     decreasing = T),]
  selection <- append(selection,
                      df.selection$ID[1:5])
  selection <- append(selection,
                      df.selection$ID[(nrow(df.selection)-4):(nrow(df.selection))])
  selection <- append(selection, df.selection$ID[(floor((nrow(df.selection)/2)-2)):
                                                   (floor((nrow(df.selection)/2))+2)])
  rm(df.selection)
}
# Extract fund selection
selection.df <- df[which(grepl(paste0(selection, collapse = "$|^"), df$ID)),]

#Sustainable-named funds:
green.label <- df[which(grepl("Sustain.|Green.|ESG.|Responsib.|SRI.|CSR.|Social.|Carbon.|Enviro
n.|Renew.", df$Name)),]


###############   Setup for term frequencies   ########################

getMasterIndex(2018) # Download master index file 2018
load(file = "Master Index/2018master.Rda") # Load index file
index.10k <- grep("^10-K$", year.master$FORM_TYPE) # Extract 10-Ks
master.2018 <- as.data.frame(year.master[index.10k,], stringAsFactors = F) # 10-K index

getMasterIndex(2017) # Download master index file 2017
load(file = "Master Index/2017master.Rda") # Load index file
index.10k <- grep("^10-K$", year.master$FORM_TYPE) # Extract 10-Ks
master.2017 <- as.data.frame(year.master[index.10k,], stringAsFactors = F) # 10-K index

getMasterIndex(2016) # Download master index file 2016
load(file = "Master Index/2016master.Rda") # Load index file
index.10k <- grep("^10-K$", year.master$FORM_TYPE) # Extract 10-Ks
master.2016 <- as.data.frame(year.master[index.10k,], stringAsFactors = F) # 10-K index

getMasterIndex(2015) # Download master index file 2015
load(file = "Master Index/2015master.Rda") # Load index file
index.10k <- grep("^10-K$", year.master$FORM_TYPE) # Extract 10-Ks
master.2015 <- as.data.frame(year.master[index.10k,], stringAsFactors = F) # 10-K index
rm(year.master)

# Load CIK to Ticker convertion file
cik.ticker <- read.csv(file = "cik_ticker.csv", header = T, sep = "|", stringsAsFactors = F)
# Load dictionary file
full.dictionary <-  read.csv(file = "Dictonary/Dictonary_full.csv",
                             header = T, sep = ",", stringsAsFactors = F)
# Pre-allocation for Term Document Matrix (DTM)
global.corpus <- as.data.frame(matrix(data = NA, nrow = 501, ncol = 10000),
                               stringsAsFactors = F)
```

```r
rownames(global.corpus) <- full.dictionary$value # Name rows by terms in dictionary
global.corpus <- global.corpus[order(rownames(global.corpus)),] # Sort alphabetically
# Pre-allocation for discriptive information
info.df <- as.data.frame(matrix(data = NA, nrow = 10000, ncol = 4), stringsAsFactors = F)
colnames(info.df) <- c("cik", "num.terms", "num.words", "date.filed")
# Create archive DTM for all 10-Ks processed for easy copy of result in later iterations of cod
e
corpus.archive <- as.data.frame(matrix(data = NA, nrow = 501, ncol = 10000),
                                stringsAsFactors = F)
rownames(corpus.archive) <- full.dictionary$value # Name rows by terms in dictionary
corpus.archive <- corpus.archive[order(rownames(corpus.archive)),] # Sort alphabetically
info.archive <- as.data.frame(matrix(data = NA, nrow = 10000, ncol = 4), stringsAsFactors = F)
colnames(info.archive) <- c("cik", "num.terms", "num.words", "date.filed")
# ID for fund selection
selection.df <- selection.df[-which(grepl("TW00", selection.df$ID)),] # remove fund with missin
g holdings

h <- 1 # Position in corpus archive
p <- 1 # Position to insert results in pre-allcated data frame

# Load corpus archive if in directory
if ("corpus_archive.Rda" %in% list.files()) {
  load(file = "corpus_archive.Rda")
}

###############   Structuring holding data    ###################

Holdings.check <- as.list(rep(NA, 3)) # Allocate list for checking if holdings have data
names(Holdings.check) <- c("2016", "2017", "2018") # Name check list
# For-loop structuring holding data
for (ID in all.ID) {
  # Skip holding if already processed. In case for-loop needs restarting.
  if (paste0(ID, ".Rda") %!in% list.files("Holdings/")) {
    # Read Holding csv file
    Holdings <- read.csv(file = paste0("Holdings/", ID, ".csv"),
                         header = T, sep = ";", dec = ",",
                         stringsAsFactors = F, na.strings = "")
    # Remove summary statistics
    Holdings <- Holdings[-((nrow(Holdings)-11):nrow(Holdings)),-c(3:5)]

    ##### 2016 #####
    Holdings.2016 <- Holdings[,c(1,2,11:14)] # Seperate 2016 holdings
    # Remove all companies not held in 2016
    Holdings.2016 <- Holdings.2016[rowSums(is.na(Holdings.2016[,3:6])) != 4,]
    # Match CIK and Ticker
    Holdings.2016 <- merge(cik.ticker[,1:2], Holdings.2016,
                           by.x = "Ticker", by.y = "Ticker", all.y = T)
    # Find 10-K filing location using master index
    Holdings.2016 <- merge(Holdings.2016, master.2016,
                           by.x = "CIK", by.y = "CIK")

    Holdings.check[[1]] <- nrow(Holdings.2016) > 0 #Check if there is holding data

    # Process for replacing 10-K filing with previous year's 10-K if
    # report is filed late. Until new 10-K is filed:
    if (Holdings.check[[1]]) { # Check for holding data
      # Seperate holdnings where 10-K were filed later than the first quarter and
      # the company is held by the fund before new 10-K is filed
      previous.hold <- Holdings.2016[Holdings.2016$QUARTER > 1 &
                                     sapply(1:nrow(Holdings.2016),
                                            function(x) any(!is.na(Holdings.2016[x,(9-Holding
s.2016$QUARTER[x]):7]))),]
      # Set NA values in both dataframes before merging to prevent
      # holdings being counted twice for one company
      if (nrow(previous.hold) > 0) {
        # Set NA values for periods after the new 10-K is filed
        for (row in 1:nrow(previous.hold)) {
          previous.hold[row, 4:(8-previous.hold$QUARTER[row])] <- NA
        }
      }
```

```
    # Set NA values for periods before the new 10-K is filed
    for (row in which(Holdings.2016$QUARTER > 1)){
      Holdings.2016[row, (9-Holdings.2016$QUARTER[row]):7] <- NA
    }
    # Merge holdings before and after newest 10-K filing
    previous.hold <- previous.hold[,-(8:12)]
    previous.hold <- merge(previous.hold, master.2015, by.x = "CIK", by.y = "CIK")
    Holdings.2016 <- rbind(Holdings.2016, previous.hold)
  }

  ##### 2017 #####
  # Same procedure only for 2017
  Holdings.2017 <- Holdings[,c(1,2,7:10)]
  Holdings.2017 <- Holdings.2017[rowSums(is.na(Holdings.2017[,3:6])) != 4,]
  Holdings.2017 <- merge(cik.ticker[,1:2], Holdings.2017,
                         by.x = "Ticker", by.y = "Ticker", all.y = T)
  Holdings.2017 <- merge(Holdings.2017, master.2017, by.x = "CIK", by.y = "CIK")

  Holdings.check[[2]] <- nrow(Holdings.2017) > 0

  if (Holdings.check[[2]]) {
    previous.hold <- Holdings.2017[Holdings.2017$QUARTER > 1 &
                                   sapply(1:nrow(Holdings.2017),
                                          function(x) any(!is.na(Holdings.2017[x,(9-Holding
s.2017$QUARTER[x]):7]))),]

    if (nrow(previous.hold) > 0) {
      for (row in 1:nrow(previous.hold)) {
        previous.hold[row, 4:(8-previous.hold$QUARTER[row])] <- NA
      }
    }
    for (row in which(Holdings.2017$QUARTER > 1)){
      Holdings.2017[row, (9-Holdings.2017$QUARTER[row]):7] <- NA
    }

    previous.hold <- previous.hold[,-(8:12)]
    previous.hold <- merge(previous.hold, master.2016, by.x = "CIK", by.y = "CIK")
    Holdings.2017 <- rbind(Holdings.2017, previous.hold)
  }

  ##### 2018 #####
  # Same procedure only for 2018
  Holdings.2018 <- Holdings[,c(1,2,3:6)]
  Holdings.2018 <- Holdings.2018[rowSums(is.na(Holdings.2018[,3:6])) != 4,]
  Holdings.2018 <- merge(cik.ticker[,1:2], Holdings.2018,
                         by.x = "Ticker", by.y = "Ticker", all.y = T)
  Holdings.2018 <- merge(Holdings.2018, master.2018, by.x = "CIK", by.y = "CIK")

  Holdings.check[[3]] <- nrow(Holdings.2018) > 0

  if (Holdings.check[[3]]) {
    previous.hold <- Holdings.2018[Holdings.2018$QUARTER > 1 &
                                   sapply(1:nrow(Holdings.2018),
                                          function(x) any(!is.na(Holdings.2018[x,(9-Holding
s.2018$QUARTER[x]):7]))),]

    if (nrow(previous.hold) > 0) {
      for (row in 1:nrow(previous.hold)) {
        previous.hold[row, 4:(8-previous.hold$QUARTER[row])] <- NA
      }
    }
    for (row in which(Holdings.2018$QUARTER > 1)){
      Holdings.2018[row, (9-Holdings.2018$QUARTER[row]):7] <- NA
    }

    previous.hold <- previous.hold[,-(8:12)]
    previous.hold <- merge(previous.hold, master.2017, by.x = "CIK", by.y = "CIK")
    Holdings.2018 <- rbind(Holdings.2018, previous.hold)
  }
```

```
    ##### Save Results #####
    Holdings.all <- list(Holdings.2016, Holdings.2017, Holdings.2018)
    names(Holdings.all) <- c(2016, 2017, 2018)

    save(Holdings.2016, Holdings.2017, Holdings.2018, Holdings.all, Holdings.check,
         file = paste0("Holdings/", ID, ".Rda"))
    cat("file processed:", ID, "\n") # Display progress
  }
}

###############   Constructing term frequencies   ################

# For-loop downloads 10-K reports based on fund selection, holding data and year.
# Then constructs term frequencies and separates the ESG words using the dictionary.
# Results are saved in a Document-term matrix
# An additional Document-term matrix is constructing containing all 10-Ks ever processed.
# This additional DTM serves as an archive for all 10-Ks processed and results can be retrieved
# from this archive rather than processing the file again if the loop is ran with a different s
election of funds
for (ID in all.ID) {
  load(file = paste0("Holdings/", ID, ".Rda")) # Load holding file
  # Save CIK for identification
  company.cik <- list(c(Holdings.2016$CIK), c(Holdings.2017$CIK), c(Holdings.2018$CIK))
  names(company.cik) <- c(2016, 2017, 2018)
  for (j in c(2016, 2017, 2018)) {
    if (Holdings.check[[paste(j)]]) { # Check for holding data
      for (i in 1:nrow(Holdings.all[[paste(j)]])) {
        # Identify filing year for 10-Ks
        filing.year <- substr(as.character(Holdings.all[[paste(j)]][i,10]), 1, 4)
        # Skip if file is already processed
        if (paste0(company.cik[[paste(j)]][i], "_",
                   filing.year) %!in% colnames(global.corpus)) {
          # Check if file is not in archive DTM
          if (paste0(company.cik[[paste(j)]][i], "_",
                     filing.year) %!in% colnames(corpus.archive)) {

            my_getFilings(filing.year, company.cik[[paste(j)]][i], "10-K") # Download 10-K

            filing <- paste0("Edgar filings/", company.cik[[paste(j)]][i],
                             "_10-K_", filing.year, "/", company.cik[[paste(j)]][i],
                             "_10-K_", as.character(Holdings.all[[paste(j)]][i,10]),
                             ".txt") # Directory where file is located locally
            word.freq <- getWordfrquency(filing) # Construct term frequency
             # Remove unnatrually long words
            word.freq <- word.freq[which(nchar(as.character(word.freq$WORD)) < 15),]
            word.freq <- word.freq[!grepl("font|style|colspan|rowspan|^type$|^new$|valign",
                                          word.freq$WORD),] # Additional cleaning
            n.term <- nrow(word.freq) # Save number of terms
            n.words <- sum(word.freq$FREQUENCY) # save number of words
            # Only keep term frequency of ESG related terms
            ESG.freq <- merge(full.dictionary, word.freq,
                              by.x = "value", by.y = "WORD", all.x = T)
            ESG.freq <- ESG.freq[order(ESG.freq$value),] # sort by alphabetically

            corpus.archive[,h] <- ESG.freq$FREQUENCY # Insert result in archive
            colnames(corpus.archive)[h] <- paste0(company.cik[[paste(j)]][i], "_",
                                                  filing.year) # Tag result with CIK and year
            # Save additional information in archive
            info.archive[h,1] <- paste0(company.cik[[paste(j)]][i], "_", filing.year)
            info.archive[h,2] <- n.term
            info.archive[h,3] <- n.words
            info.archive[h,4] <- as.character(Holdings.all[[paste(j)]][i,10])

            global.corpus[,p] <- ESG.freq$FREQUENCY # Insert results in Document-term matrix
            colnames(global.corpus)[p] <- paste0(company.cik[[paste(j)]][i], "_",
                                                 filing.year) # Tag result with CIK and year
            # Save additional information
            info.df[p,1] <- paste0(company.cik[[paste(j)]][i], "_", filing.year)
            info.df[p,2] <- n.term
            info.df[p,3] <- n.words
```

```
            info.df[p,4] <- as.character(Holdings.all[[paste(j)]][i,10])
            # Display progression
            cat("File processed:",
                "\n fund:", ID,
                "\n year:", j,
                "\n company:", i, "/", nrow(Holdings.all[[paste(j)]]) ,"\n")
            # Delete 10-K from local directory
            unlink(filing, recursive = T)
            h <- h+1 # Next position in archive
            p <- p+1 # Next position in Document-term matrix
            # Save results
            save(corpus.archive, info.archive, h, file = "corpus_archive.Rda")
          } else { # If file is in archive but not in current DTM. Copy results from archive
            # Copy results
            global.corpus[,p] <- corpus.archive[,paste0(company.cik[[paste(j)]][i], "_",
                                                         filing.year)]
            colnames(global.corpus)[p] <- paste0(company.cik[[paste(j)]][i], "_",
                                                  filing.year) # Tag with CIK and year
            # Copy additional information
            info.df[p,] <- info.archive[which(grepl(paste0(company.cik[[paste(j)]][i], "_",
                                                            filing.year),info.archive$cik)),]
            # Display progression
            cat("File processed:",
                "\n fund:", ID,
                "\n year:", j,
                "\n company:", i, "/", nrow(Holdings.all[[paste(j)]]) ,"\n")

            p <- p+1 # Next position in Document-term matrix
          }
          # Save results
          save(global.corpus, info.df, p, file = "corpus.greenlabel.Rda")
        }
      }
    }
  }

}

###############   TF IDF    ###############

# Remove unused space from preallocation in Document-term matrix
global.corpus <- global.corpus[,!apply(is.na(global.corpus), 2, all)]
info.df <- info.df[!apply(is.na(info.df), 1, all),]

# Create function for term-frequency inverse-document-frequency weighting scheme
TF.IDF <- function(corpus) {
  tf.t <- apply(corpus, 1, function(x) sum(x, na.rm = T))
  df.t <- apply(corpus, 1, function(x) length(which(!is.na(x))))
  avtf <- tf.t/df.t
  tf <- apply(corpus, 2, function(x) ((1+log10(x))/(1+log10(avtf))))
  tf <- as.data.frame(tf, stringsAsFactors = F)
  D <- length(corpus)
  idf <- apply(corpus, 1, function(x) log10(D/(length(which(!is.na(x))))))
  tf.idf <- apply(tf, 2, function(x) x*idf)
  tf.idf <- as.data.frame(tf.idf, stringsAsFactors = F)
  return(tf.idf)
}

tf.idf <- TF.IDF(corpus = global.corpus) # Construct weighted Document-term matrix

###############   Structure Results on Company Level  ##############
# Load dictionaries separated by E, S, and G
Environment <- read.csv(file = "Dictonary/Dictonary.environment.csv", header = T, sep = ",", st
ringsAsFactors = F)
Governance <- read.csv(file = "Dictonary/Dictonary.governance.csv", header = T, sep = ",", stri
ngsAsFactors = F)
Social <- read.csv(file = "Dictonary/Dictonary.social.csv", header = T, sep = ",", stringsAsFac
tors = F)

# Seperate Document-term matrix based on E, S, and G
```

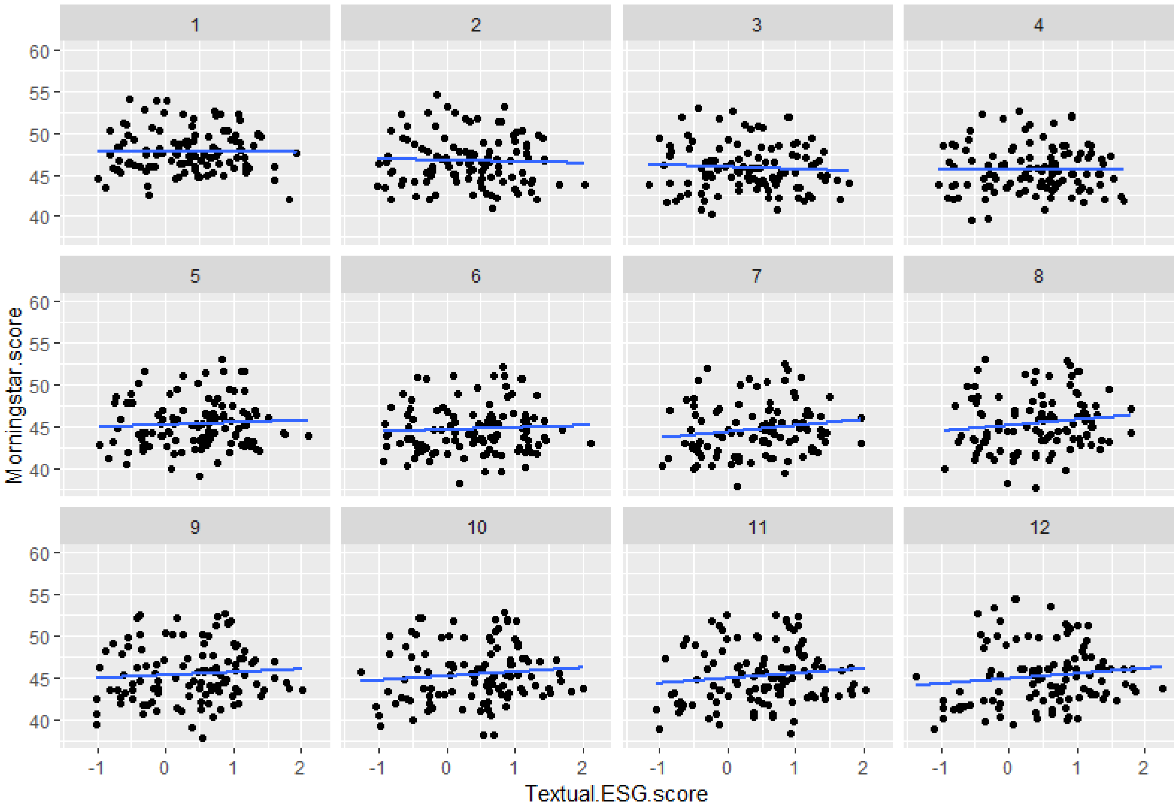
```
Environment <- tf.idf[which(grepl(paste(Environment$value, collapse = "$|^"), rownames(tf.idf))
),]
Social <- tf.idf[which(grepl(paste(Social$value, collapse = "$|^"), rownames(tf.idf))),]
Governance <- tf.idf[which(grepl(paste(Governance$value, collapse = "$|^"), rownames(tf.idf))),
]

# Sum term frequency in each ESG pillar
results <- data.frame(CIK = colnames(tf.idf),
                      Environment = colSums(Environment, na.rm = T),
                      Social = colSums(Social, na.rm = T),
                      Governance = colSums(Governance, na.rm = T),
                      stringsAsFactors = F)

#  Standardization
results$Environment <- scale(results$Environment)
results$Social <- scale(results$Social)
results$Governance <- scale(results$Governance)

###############   Structure score to fund level   ################
# Allocate new colums for results
new.columns <- c("E_Q1_2016", "S_Q1_2016", "G_Q1_2016", "ESG_Q1_2016",
                 "E_Q2_2016", "S_Q2_2016", "G_Q2_2016", "ESG_Q2_2016",
                 "E_Q3_2016", "S_Q3_2016", "G_Q3_2016", "ESG_Q3_2016",
                 "E_Q4_2016", "S_Q4_2016", "G_Q4_2016", "ESG_Q4_2016",
                 "E_Q1_2017", "S_Q1_2017", "G_Q1_2017", "ESG_Q1_2017",
                 "E_Q2_2017", "S_Q2_2017", "G_Q2_2017", "ESG_Q2_2017",
                 "E_Q3_2017", "S_Q3_2017", "G_Q3_2017", "ESG_Q3_2017",
                 "E_Q4_2017", "S_Q4_2017", "G_Q4_2017", "ESG_Q4_2017",
                 "E_Q1_2018", "S_Q1_2018", "G_Q1_2018", "ESG_Q1_2018",
                 "E_Q2_2018", "S_Q2_2018", "G_Q2_2018", "ESG_Q2_2018",
                 "E_Q3_2018", "S_Q3_2018", "G_Q3_2018", "ESG_Q3_2018",
                 "E_Q4_2018", "S_Q4_2018", "G_Q4_2018", "ESG_Q4_2018")
selection.df[new.columns] <- NA

# For-loop structuring results
for (ID in all.ID) {
  # Load Holdings for fund
  load(file = paste0("Holdings/", ID, ".Rda"))
  ##### 2016 #####
  if (nrow(Holdings.2016) > 0) {
    # Extract cik and tag with year of 10-K filing date
    Holdings.2016$CIK <- paste0(Holdings.2016$CIK, "_", substr(Holdings.2016$DATE_FILED,1, 4))
    # Merge results and holdings by cik and year
    Holdings.2016 <- merge(Holdings.2016, results, by.x = "CIK", by.y = "CIK")
    # Convert weightings to numeric
    Holdings.2016[,4:7] <- apply(Holdings.2016[,4:7], 2,
                                 function(x) as.numeric(x))
    # Check if there are results for a weighted 50% of the holdings in the fund
    Holdings.check <- as.vector(apply(Holdings.2016[4:7], 2,
                                      function(x) sum(x, na.rm = T)) > 50)
    # For-loop constructing results for each quarter in 2016
    for (i in 1:4) {
      # Check 50% results
      if (Holdings.check[(5-i)]) {
        # Extract results and holdings for current quarter
        H <- data.frame(CIK = Holdings.2016$CIK,
                        # Weighted holdings for current quarter in loop
                        weight = Holdings.2016[,(8-i)],
                        # Environental score
                        E = as.vector(Holdings.2016$Environment),
                        # Soial score
                        S = as.vector(Holdings.2016$Social),
                        # Governance score
                        G = as.vector(Holdings.2016$Governance), stringsAsFactors = F)
        H[,2] <- apply(H[2], 2, function(x) x/sum(x, na.rm = T)) # Adjust weightings
        H[,3:5] <- H[,3:5]*H[,2] # Multiply results with corresponding weight in fund
        # Add Environmental, Social and Governance score to create full ESG score
        H$ESG <- apply(H[,3:5], 1, sum)
        # Sum the weighted results to a single score for current
        # fund in current quarter and insert into result data frame
```

```
      selection.df[which(grepl(ID, selection.df$ID)),
                   (23+(i*4):(i*4+3))] <- apply(H[3:6], 2, function(x) sum(x, na.rm = T))
    } else {
      # If results < 50% set as NA
      selection.df[which(grepl(ID, selection.df$ID)),(23+(i*4):(i*4+3))] <- NA
    }
  }
}

##### 2017 #####
# Same procedure only for 2017
if (nrow(Holdings.2017) > 0) {
  Holdings.2017$CIK <- paste0(Holdings.2017$CIK, "_", substr(Holdings.2017$DATE_FILED,1, 4))
  Holdings.2017 <- merge(Holdings.2017, results, by.x = "CIK", by.y = "CIK")

  Holdings.2017[,4:7] <- apply(Holdings.2017[,4:7], 2,
                               function(x) as.numeric(x))
  Holdings.check <- as.vector(apply(Holdings.2017[4:7], 2,
                                    function(x) sum(x, na.rm = T)) > 50)
  for (i in 1:4) {
    if (Holdings.check[(5-i)]) {
      H <- data.frame(CIK = Holdings.2017$CIK,
                      weight = Holdings.2017[,(8-i)],
                      E = as.vector(Holdings.2017$Environment),
                      S = as.vector(Holdings.2017$Social),
                      G = as.vector(Holdings.2017$Governance), stringsAsFactors = F)
      H[,2] <- apply(H[2], 2, function(x) x/sum(x, na.rm = T))
      H[,3:5] <- H[,3:5]*H[,2]
      H$ESG <- apply(H[,3:5], 1, sum)
      selection.df[which(grepl(ID, selection.df$ID)),
                   (39+(i*4):(i*4+3))] <- apply(H[3:6], 2, function(x) sum(x, na.rm = T))
    } else {
      selection.df[which(grepl(ID, selection.df$ID)),(39+(i*4):(i*4+3))] <- NA
    }
  }
}

##### 2018 #####
# Same procedure only for 2018
if (nrow(Holdings.2018) > 0) {
  Holdings.2018$CIK <- paste0(Holdings.2018$CIK, "_", substr(Holdings.2018$DATE_FILED,1, 4))
  Holdings.2018 <- merge(Holdings.2018, results, by.x = "CIK", by.y = "CIK")

  Holdings.2018[,4:7] <- apply(Holdings.2018[,4:7], 2,
                               function(x) as.numeric(x))
  Holdings.check <- as.vector(apply(Holdings.2018[4:7], 2,
                                    function(x) sum(x, na.rm = T)) > 50)
  for (i in 1:4) {
    if (Holdings.check[(5-i)]) {
      H <- data.frame(CIK = Holdings.2018$CIK,
                      weight = Holdings.2018[,(8-i)],
                      E = as.vector(Holdings.2018$Environment),
                      S = as.vector(Holdings.2018$Social),
                      G = as.vector(Holdings.2018$Governance), stringsAsFactors = F)
      H[,2] <- apply(H[2], 2, function(x) x/sum(x, na.rm = T))
      H[,3:5] <- H[,3:5]*H[,2]
      H$ESG <- apply(H[,3:5], 1, sum)
      selection.df[which(grepl(ID, selection.df$ID)),
                   (55+(i*4):(i*4+3))] <- apply(H[3:6], 2, function(x) sum(x, na.rm = T))
    } else {
      selection.df[which(grepl(ID, selection.df$ID)),(55+(i*4):(i*4+3))] <- NA
    }
  }
}
cat("File processed:", ID, "\n")
}

# Save final results
save(global.corpus, selection.df, results, info.df, tf.idf, full.dictionary, df,
     file = "results.Rda")
```

9.2 Appendix 2: Time Periodes

Results by time period

As the plots demonstrates, the relation between disclosure score and sustainability are relatively consistent over time, thus we choose not to use panel data.



9.3 Appendix 3: ESG-dictionary

The following ESG-dictionary is used the textual analysis. It is based on the dictionary by Baier, Berninger & Kiesel (2018), with a few added words related to the environmental pillar.

Environmental	clean, environmental, epa, sustainability, ecological, environment, environmentalist, esg, gri, preservation, preserve, climate, warming, ozone, biofuels, biofuel, green, renewable, solar, stewardship, wind, emission, emissions, ghg, ghgs, greenhouse, atmosphere, emit, co2, dioxide, agriculture, deforestation, pesticide, pesticides, wetlands, agricultural, rainforest, zoning, farmland, biodiversity, species, wilderness, wildlife, freshwater, groundwater, water, cleaner, cleanup, coal, contamination, fossil, resource, air, carbon, nitrogen, pollution, superfund, carbons, biphenyls, hazardous, householding, pollutants, printing, recycling, toxic, waste, wastes, weee, recycle
Social	citizen, citizens, csr, disabilities, disability, disabled, human, nations, social, un, veteran, veterans, vulnerable, kld, foodbank, orphan, children, epidemic, health, healthy, ill, illness, pandemic, childbirth, drug, medicaid, medicare, medicine, medicines, healthcare, hiv, aids, alcohol, drinking, bugs, conformance, defects, fda, inspection, inspections, minerals, standardization, warranty, dignity, discriminate, discriminated, discriminating, discrimination, equality, freedom, humanity, nondiscrimination, sexual, communities, community, expression, marriage, privacy, peace, bargaining, eeo, fairness, fla, harassment, injury, labor, overtime, ruggie, sick, wage, wages, workplace, bisexual, diversity, ethnic, ethnically, ethnicities, indigenous, ethnicity, female, females, gay, gays, gender, genders, homosexual, immigration, lesbian, lesbians, lgbt, minorities, minority, ms, race, racial, religion, religious, sex, transgender, woman, women, occupational, safe, safely, safety, ilo, labour, eicc, endowment, endowments, people, philanthropic, philanthropy, socially, societal, society, welfare, charitable, charities, charity, donate, donated, donates, donating, donation, donations, donors, foundation, foundations, gift, gifts, nonprofit, poverty, courses, educate, educated, educates, educating, education, educational, learning, mentoring, scholarships, teach, teacher, teachers, teaching, training, employ, employment, headcount, hire, hired, hires, hiring, staffing, unemployment
Governance	align, aligned, aligning, alignment, aligns, bylaw, bylaws, charter, charters, culture, death, duly, parents, independent, compliance, conduct, conformity, governance, misconduct, parachute, parachutes, perquisites, plane, planes, poison, retirement, approval, approvals,

	approve, approved, approves, approving, assess, assessed, assesses, assessing, assessment, assessments, audit, audited, auditing, auditor, auditors, audits, control, controls, coso, detect, detected, detecting, detection, evaluate, evaluated, evaluates, evaluating, evaluation, evaluations, examination, examinations, examine, examined, examines, examining, irs, oversee, overseeing, oversees, oversight, review, reviewed, reviewing, reviews, rotation, test, tested, testing, tests, treadway, backgrounds, independence, leadership, nomination, nominations, nominee, nominees, perspectives, qualifications, refreshment, skill, skills, succession, tenure, vacancies, vacancy, appreciation, award, awarded, awarding, awards, bonus, bonuses, cd, compensate, compensated, compensates, compensating, compensation, eip, iso, isos, payout, payouts, pension, prsu, prsus, recoupment, remuneration, reward, rewarding, rewards, rsu, rsus, salaries, salary, severance, vest, vested, vesting, vests, ballot, ballots, cast, consent, elect, elected, electing, election, elections, elects, nominate, nominated, plurality, proponent, proponents, proposal, proposals, proxies, quorum, vote, voted, votes, voting, brother, clicking, conflict, conflicts, family, grandchildren, grandparent, grandparents, inform, insider, insiders, inspector, inspectors, interlocks, nephews, nieces, posting, relatives, siblings, sister, son, spousal, spouse, spouses, stepchildren, stepparents, transparency, transparent, visit, visiting, visits, webpage, website, attract, attracting, attracts, incentive, incentives, interview, interviews, motivate, motivated, motivates, motivating, motivation, recruit, recruiting, recruitment, retain, retainer, retainers, retaining, retention, talent, talented, talents, cobc, ethic, ethical, ethically, ethics, honesty, bribery, corrupt, corruption, crimes, embezzlement, grassroots, influence, influences, influencing, lobbied, lobbies, lobby, lobbying, lobbyist, lobbyists, whistleblower, announce, announced, announcement, announcements, announces, announcing, communicate, communicated, communicates, communicating, erm, fairly, integrity, liaison, presentation, presentations, sustainable, asc, disclose, disclosed, discloses, disclosing, disclosure, disclosures, fasb, gaap, objectivity, press, sarbanes, engagement, engagements, feedback, hotline, investor, invite, invited, mail, mailed, mailing, mailings, notice, relations, stakeholder, stakeholders, compact, ungc