Norwegian School of Economics
Bergen, Spring, 2022

Material ESG and Stock Performance

A Textual Analysis Approach to Investigate the Relationship Between Material ESG Performance and Stock Performance

Martin Uthaug and Petter Kårstad

Supervisor: Krisztina Molnar

Master thesis, MSC in economics and business administration,

Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgments

We want to express our sincere gratitude to our supervisor Krizstina Molnar for guidance and engagement throughout our work with this thesis. Her inputs were of great use for the quality of our thesis. We would also like to thank NHH for access to Refinitiv Eikon's database. Lastly, we thank each other for a fulfilling partnership.

Abstract

This thesis investigates the relationship between material environmental, social and governance (ESG) performance and stock performance. We construct a new ESG score based on textual analysis of annual reports of companies listed on the Oslo Stock Exchange. The ESG score is a product of the presence of material ESG-related terms in the company's annual report. We use a custom ESG dictionary to identify material ESG-related words. Furthermore, we construct equal and value-weighted zero-investment portfolios, best-in-class portfolios, and negative screening portfolios to investigate if portfolios consisting of high ESG-scoring firms achieve abnormal stock returns. The ESG portfolio's excess returns are estimated using the Fama-French five-factor model + momentum.

We find that the equal-weighted zero-investment portfolio consisting of a long position in the top quintile ESG-scoring firms and a short position in the bottom quintile ESG-scoring firms achieve significant negative abnormal returns in the period 2008-2014. Furthermore, we find that an equal-weighted best-in-class portfolio that is consisting of the top quintile ESG-scoring companies, is also associated with negative abnormal returns in the same period. We do not find significant abnormal returns after 2014. We argue that the market has mispriced the risk associated with ESG companies in 2008-2014, and that a learning effect has led to ESG companies being correctly priced in recent years. We argue that the learning effect is due to an increased supply of material ESG metrics among investors.

Contents

List of Figures

List of Tables

1.0 Introduction

Companies have one function, and that is to maximize profits for their owners. At least that is what Milton Friedman thought in 1970. He believed that corporate social responsibility was to satisfy the owners, and that companies which included sustainability in their operations would underperform in terms of stock returns (Friedman, 1970). These opinions have historically received broad support from investors, and still have broad support among many to this day. Nevertheless, environment, social and governance (ESG) is facing seemingly more attention in recent years. The acronym ESG is used especially for sustainability in business and describes how companies incorporate environmental, social, and governance issues in their operations. The increased attention to sustainability has led investors to incorporate ESG measures into their investment strategies. Some researchers even claim that it is possible to achieve positive abnormal returns by investing in companies which perform well in ESG. We investigate this claim, and formulate the following research question:

Can companies which perform well on material ESG issues expect positive abnormal returns?

To answer this question, we construct portfolios and regress the portfolio excess returns to the excess market returns and risk factors accounting for size, value, profitability, investments, and past stock returns. The portfolios consist of companies on the Oslo Stock Exchange and the portfolio composition is based on the companies' ESG scores. Finally, we see whether these portfolios have historically yielded abnormal risk adjusted returns.

We use textual analysis on companies' annual report to measure ESG performance. The ESG scores will depend on textual presence of ESG related terms in the annual reports, and the ESG terms will be industry dependent. Only the ESG issues that are categorized as material for the company’s industry will have an impact on the company's score. By material issues, we mean issues that has been categorized as value-driving for the respective industry. We follow the guidelines of Sustainability Accounting Standards Boards (SASB) when categorizing material issues.

This paper is unique because we investigate the relationship between ESG performance and stock performance for the Norwegian market, more specifically companies on the Oslo Stock Exchange, while existing literature in the ESG field examine the US markets. But what distinguishes this paper the most from existing literature is that we construct a materiality ESG score based on textual analysis. Existing literature is based on ESG scores provided by third party agencies such as Bloomberg, Sustainalytics and Refinitiv. Furthermore, the concept of materiality is relatively new and was not incorporated into research prior to 2016.

It has been difficult to conduct studies that investigate the relationship between ESG performance and stock performance on the Oslo Stock Exchange due to lack of ESG data on Norwegian companies. The advantage of the score that we construct is that it covers most companies on the Oslo Stock Exchange.

Another advantage of using our ESG score is that established ESG scores has selection bias. Established ESG scores are product of many different quantitative and qualitative metrics which companies can select to report on. This means that only companies which disclose ESG metrics will receive a rating. It is plausible that companies that choose to report on nonfinancial ESG metrics are systematically different from companies that keep this information undisclosed. Drempetic, Klein, & Zwergel (2020) find that Refinitiv's database of companies' ESG scores has a size bias, more specifically, they find that companies that get scores from Refinitiv are large companies. We tackle this by giving a score to all companies that have an English annual report, which is most of the listed companies. If the company's annual report keeps ESG issues undisclosed, they will receive a low score from us.

We add value to existing literature by introducing a seemingly untested method on a market with limited prior ESG research. The method used is innovative and can inspire to new ways of measuring ESG. We thus formulate the secondary research question as follows:

Is it possible to measure a company's material ESG performance based on textual analysis of the company's annual report?

A large proportion of this paper will focus on validating the constructed ESG score and the extent to which this score gives the same outcome as an established ESG score.

The rest of the paper proceeds as follows. In chapter 2 we discuss relevant literature within textual analysis in finance and the relationship between ESG and stock performance. In chapter 3 we define and explain our hypothesis. In chapter 4 we describe data sources, the data retrieval approaches and the transformation from textual data to quantitative measures. In chapter 5 we explain the methodology of creating the portfolios and regression models. In chapter 6 we present the results of the regression models. In chapter 7 we discuss the results and in chapter 8 we conclude the thesis.

2.0 Literature Review

In this section, we look at existing literature within textual analysis in finance, the relationship between ESG and stock performance, and disagreement between the established score providers.

2.1 Textual Analysis in Finance

Textual analysis is a relatively new field in finance and the literature on textual analysis in an ESG context is limited. Nevertheless, a review of existing literature in textual analysis in a finance context can help us make good methodological choices. Loughran & McDonald (2016) wrote in a literature review on the subject and concluded that "The words selected by managers to describe their operations and the language used by media to report on firms and markets have been shown to be correlated with future stock returns, earnings, and even future fraudulent activities of management". It is thus conceivable that the frequency of ESG related terms in the annual reports may also correlate with stock returns.

Sentiment analysis is the most common form of textual analysis in finance. Sentiment analysis measures the "tone" of a text document. The tone is a product of the word-count of positive and negatively charged words. To categorize which words are positively or negatively charged, one uses dictionaries specifically made for sentiment.

Price, Doran, Peterson, & Bliss (2012) use sentiment-analysis and find that companies with transcripts of quarterly earnings conference calls with negative tone are associated with negative abnormal returns. Davis, Piger, & Sedor (2012) also use sentiment analysis and find that earnings press releases with positive tone are associated with higher subsequent return on assets. Sentiment analysis has been applied to newspaper articles and García (2013) find that the tone of news articles from financial times in the period 1905 to 2005 plays a role in predicting future returns, especially in recessionary periods. Mayew & Venkatachalam (2012) find that a positive tone in the transcript of earnings conference call audio files is associated with higher returns and negative tone is associated with negative returns.

Sentiment analysis is the most common but not the only way to do textual analysis on financial text documents. Li (2008) find that companies with low reported earnings tend to have annual reports that are difficult to read. The Fog-Index[1] was used to measure readability in annual reports. Loughran, McDonald, & Yun (2009) find that companies with a high word-count of the words «ethic», «corporate responsibility», «social responsibility» and «socially responsible» in 10-k filings are more often categorized as sin[2]- stocks and exercise bad corporate governance. This finding is of great relevance to our paper as we give ESG scores to companies based on word-count of ESG words.

2.2 ESG and Financial Performance

Literature which investigates the relationship between ESG performance and financial performance shows mixed results. Older literature has focused on socially responsible investing (SRI). SRI means that investors pick stocks after a screening process where stocks that doesn't meet the investors criteria are filtered. These criteria can for instance be the level of shareholder engagement, community investing, sustainability, or exclusion of sin stocks. The ESG term was first used at the Who Cares Wins (2005) conference and has gradually replaced SRI as a sustainability measure for investors. ESG-investing is a broader term and look at how environmental, social and governance factors impact performance and risk. For instance, an alcohol producer can be considered responsible in ESG investing if it is working continuously with water resource management but could be excluded by an SRI investor due to the nature of its business.

Kempf & Osthoff (2007) find a positive relationship between SRI and stock performance and Statman & Glushkov (2008) find that a best-in-class method where tilting towards social responsible firms can produce superior returns.

More recent studies have focused on the relationship between ESG and financial performance. Halbritter & Dorfleitner (2015) find results which indicate that investors should no longer expect abnormal returns by trading a difference portfolio of high and low rated firms regarding ESG aspects. Sargis & Wang (2020) also find no risk/reward trade-off to investing in ESG on a global level.

[1] The Fog index = $0.4 * [\left(\frac{Total\ words}{Total\ sentences}\right) + 100\left(\frac{Complex\ words}{Total\ words)}\right)]$

[2] Exclusion of companies involved in activities which is considered unethical or immoral, such as alcohol, tobacco, gambling, and pornography

Friede, Busch, & Bassen (2015) conduct a meta study of 2200 unique studies and conclude that ESG outperformance opportunities exist in North-America and that capital markets demonstrated no consistent learning-effects[3].

The main inspiration for this paper comes from the research conducted by Khan, Serafeim, & Yoon (2016). They find that firms that do well on material sustainability issues tend to outperform in terms of stock price, while those that do well on immaterial sustainability issues do not. They develop a data set to measure engagement in material sustainability issues by hand-mapping recently available industry-specific guidance on materiality from the Sustainability Accounting Standards Board (SASB) to companies on the MSCI KLD index.

Our main difference from their research is that we use textual analysis of annual reports to gather ESG information, while they do manual retrieval of available ESG metrics. Their sample consists of shares from U.S. in the period 1991-2012. Our sample exclusively consists of shares on the Oslo Stock Exchange from 2008 to the end of 2020.

Danielsen & Johansen (2021) find results that substantiated the research by Khan et al. (2016). They find that an investment strategy based on taking a long position in companies with high ESG score when ESG is value relevant and a short position in stocks with a high ESG score when ESG is not value relevant generates superior performance.

2.3 Score Disagreement

The research on ESG performance and Stock performance gives contradictory results. Dorfleitner, Halbritter, & Nguyen (2015) argue that the different results come from researchers using different ESG scores from different rating providers. Established ESG scores generally have a low correlation, especially on the social and governance pillars. Gibson, Krueger, & Schmidt (2019) find an average correlation of 0.46 between 6 different score providers. The providers appear to be more in agreement on the Environment pillar score. Berg, Kölbel, & Rigobon (2019) suggest that the differences in ESG scores come from the use of different categories (scope divergence), measuring the same categories differently (measurement divergence), and using different weights in measuring the categories (weight divergence).

[3] Learning effect refers to the market learning to price correctly over time.

ESG is complex and there is no consensus on how to measure it. This inspires us to measure ESG using a new and innovative method. We do not intend to replace existing ESG scores, but rather contribute to the field of ESG research by adding a new method of scoring.

3.0 Hypothesis

Previous research in the ESG field has contradictory results. It is plausible that our results will suffer from low external validity as different methods and datasets tend to show different results. We therefore formulate a hypothesis that applies to our specific situation. Our hypothesis therefore reads as follows:

Companies on the Oslo Stock Exchange which perform well on material ESG issues have positive abnormal returns.

Based on the literature review, we expect a positive link between material ESG performance and abnormal returns. Most of the research supports this. We expect performance on material ESG to be rewarded with superior stock performance because companies which prioritizes ESG issues that contribute positively to the company financials will allocate resources effectively.

The hypothesis can only be tested if we are successfully able to measure material ESG performance. The process of validating the constructed ESG score will indicate whether we are able to measure ESG performance or not.

4.0 Data and Variable Construction

In this section we present our data, which includes data sources and descriptive statistics. Furthermore, we present the construction of our dictionary and the ESG score. This section will also aim to we validate the constructed ESG score.

4.1 Data Sources

Annual reports are retrieved manually from the respective companies' websites. Existing ESG scores and market values are retrieved from Refinitiv's API solution. Industry-specific ESG data is retrieved manually from SASB's website. Price data is retrieved from Yahoo finance's API solution. Risk factors used in the regressions are retrieved from Kenneth R. French's data library.

4.1.1 Annual Reports and Sample Construction

We obtain 1572 annual reports manually from the companies' webpages. Laws regarding how and when annual reports are to be published differ from country to country. In Norway, annual reports must contain a management report, introduction to the business and key figures, the year's activities and results, management and control of the business, assessment of future prospects and annual accounts. The parts must follow each other chronologically in the given order and the annual report must be published by May 1st in the following year (DFØ, 2022).

Some of the annual reports we download are dated back to 2000 and some has their first year on Oslo Stock Exchange. We exclude annual reports from 2000 to 2007 from our sample because the availability of annual reports for these years are minimal, which means that the sample would be very low during these years. The sample is thus limited to the period 2008-2020.

Our sample consists of companies that were listed as of 2022 May 14th[4], which means we are losing out on companies that have either gone bankrupt or private in the period under examination. The sample is thus weighted towards companies that do not go bankrupt or

[4] The last price-data retrieval happened 2022.05.14 which restricts the sample to companies that were listed this date.

private during the period. It is difficult to eliminate this sample bias because Yahoo Finance do not offer pricing information for companies that are no longer listed. In addition, it is difficult to retrieve annual reports for companies that no longer exist.

Figure 1: Sample industries

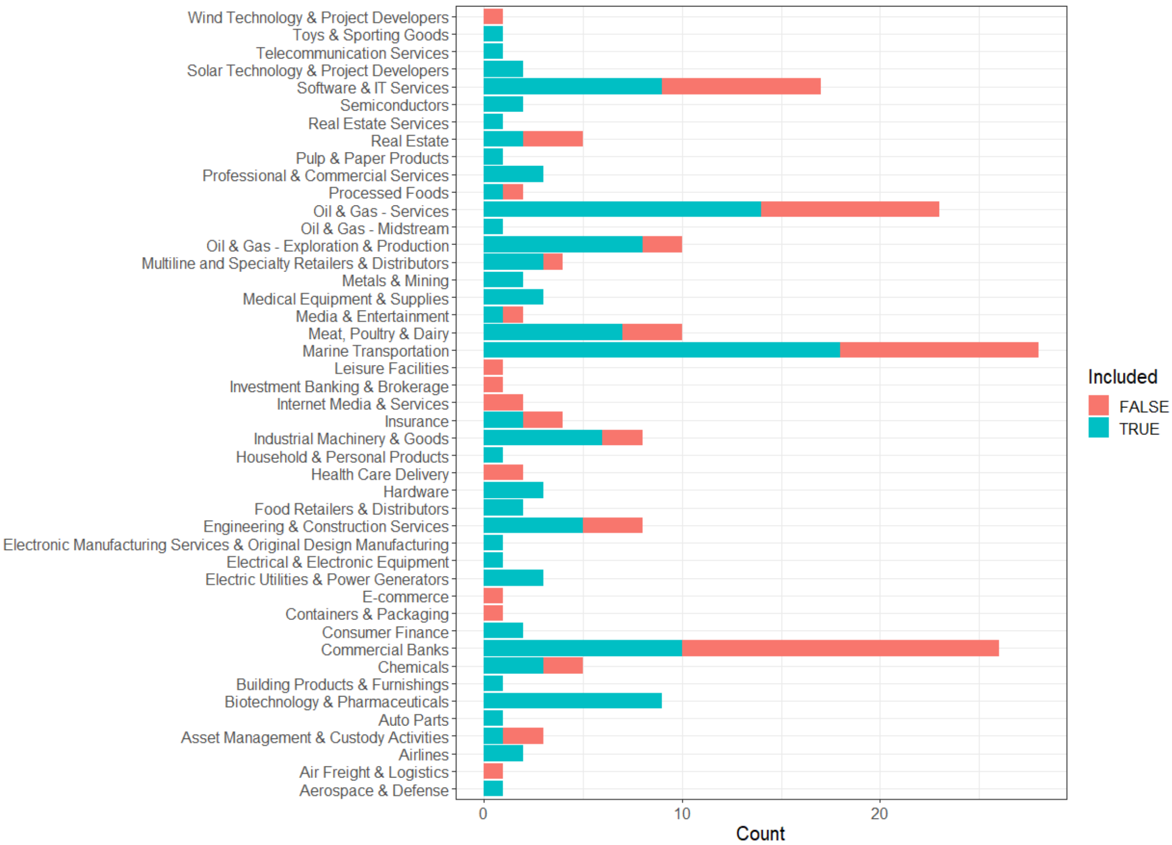


Figure 1 is a stacked bar plot and illustrates which companies are represented on the Oslo Stock Exchange. The colours indicate whether the companies in the industry are present in our sample or not. Red bar means that the companies are represented at Oslo Stock Exchange, but not in our sample, and blue bar means that the companies are represented in our sample.

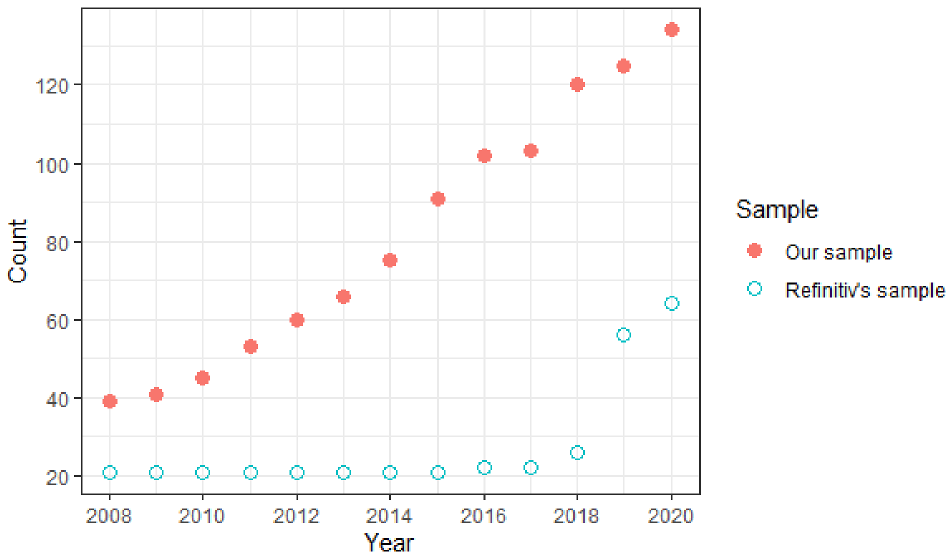
Figure 1 illustrate which industries are represented on the Oslo Stock Exchange and which industries are represented in our sample. Out of the 77 possible industries[5], only 45 are represented on Oslo Stock Exchange. Commercial banks, marine transportation, oil and gas services, and software and IT services are represented most frequently on the Oslo Stock Exchange. Those 4 industries are represented by 15 companies or more. There are 18 industries only represented by one company.

[5] As classified by Sustainable Industry Classification System (SICS)

By manual retrieval of annual reports, we notice that many commercial banks publish their annual reports written in Norwegian. We do not include these annual reports in our sample. Figure 1 display that our sample is missing more than half of the commercial banks listed on the Oslo Stock Exchange. However, commercial banks are well represented with over 10 companies in our sample. Apart from that, we do not see any systematic differences in our sample compared to full sample.

The portfolios are rebalanced annually, which means that sample size change over time. Figure 2 display our sample size as a function of year. The figure also includes how many of the companies on the Oslo Stock Exchange for which Refinitiv provide ESG scores.

Figure 2: Sample size over time



The red solid dots show our sample size in 2008-2020, while the blue hollow dots show how many companies there are in Refinitiv's ESG database in the same period. Both samples are of the Oslo Stock Exchange.

We have more annual reports in recent years because our data sample consists of companies that existed in 2022. If a company went off the Oslo Stock Exchange in 2019, the company will not be included in our sample. Many of the companies at the Oslo Stock Exchange are relatively new, hence we have fewer annual reports from 2008. Refinitiv provide ESG score for a total of 21 companies in 2008 and 64 companies 2020, while we provide ESG score for 39 companies in 2008 and for 134 companies in 2020. The limited presence of Norwegian companies in ESG databases makes it problematic to conduct ESG studies at the Norwegian market. The score we construct tackle this issue by increasing the sample size.

Refinitiv’s sample rose sharply in 2019 and 2020. The increase in their ESG database could be a result of the implementation of EU Taxonomy. In 2018, the EU stated that an ESG taxonomy with associated regulations would be introduced. The taxonomy was fully implemented in 2020 and various regulations for reporting gradually enter into force from 2021 (European Union, 2020). It is plausible that some companies integrated ESG reporting already in 2019 and to a larger extent in 2020 to be precautionary of the new regulations. Refinitiv’s score is based on disclosed ESG information, and an increase in disclosure would increase Refinitiv’s ESG database.

4.1.2 Refinitiv

Refinitiv is one of the world’s largest providers of financial market data and infrastructure. In this thesis we use the platform, Refinitiv Eikon to retrieve market capitalizations and ESG-scores of companies on the Oslo Stock Exchange. Refinitiv Eikon offer one of the most comprehensive ESG databases in the industry with a history that goes back to 2002. They provide ESG data coverage for more than 10,000 global companies across 76 countries (Refinitiv, 2022). Even though Refinitiv is one of the largest providers of ESG information, they only have ESG scores for 64 Norwegian companies as of 2020.

Refinitiv collects data on 630 ESG variables in total. Each company will receive an ESG score which is a composition of 186 of these variables. Which variables make up a company's score depend on which industry the companies belong to. Refinitiv also uses the term «material» about its ESG measures, but it has a different meaning than the definition we use. We use SASB's definition where material issues are the issues that drive value for the company from an investor's perspective. Refinitiv's definition of materiality is based on industry norms. Refinitiv’s ESG score is thus weighted heavier by variables if it is common industry practice to report on this variable compared with other variables. The ESG score is composed of 10 different subcategories within ESG and each of these subcategories has additional subcategories that are reported on. See appendix 4 for which subcategories the score consists of.

4.1.3 SASB

Sustainability Accounting Standard Boards (SASB) provide a common language for sustainability and identifies which issues are material for which industries (SASB, 2022). SASB is maintained by the Value Reporting Foundation, a non-profit organization whose purpose is to provide a mutual understanding of the valuation of enterprises, which include ESG.

SASB has identified 77 industries, and for each of these industries they have identified a set of sustainable issues that affects the financial performance of companies in the respective industry. SASB aims to help investors, which means that the issues are categorized as material because they have an impact on the financial performance of the company and thereby the investor. SASB do not focus on other stakeholders, but some material issues do overlap between shareholders and other stakeholders. See appendix 2 for a material mapping example for the Oil & Gas – Exploration & Production industry.

SASB use 5 categories within sustainability which is Environment, Social Capital, Human Capital, Business Model & Innovation, and Leadership & Governance. Each category is further divided into 3 - 7 issues that can be categorized as material or immaterial. For example, the Oil and Gas Exploration and Production industry has defined "Employee health and Safety" as a material issue, but "data security" as an immaterial issue. This is probably because companies in that industry do not have huge databases with sensitive personal customer information, thus focusing on data privacy would lead to ineffective resource allocation. Explanations for the materiality mapping is reported in a 25-40 paged standard report. SASB have a unique standard report for each of the 77 different industries. These reports are the foundation for the dictionary[6] we create.

We also used SASB to categorize each company in an industry. We use EuroNext to extract International Securities Identification Number (ISIN) from all the companies noted at Oslo Stock Exchange. EuroNext uses Industry Classification Benchmark (ICB) sector classifying while SASB uses the Sustainable Industry Classification System (SICS) to classify companies to their respective industry. For example, "MOWI ASA" is classified as "Food producers" by EuroNext and as "Meat, Poultry & Dairy" by SASB. We find it most beneficial to use SASB's industry classification for all the companies, as SASB provides the 77 sector-reports that we use for materiality. We use the R-Package rvest (Wickham, 2022) to write a web-

[6] A dictionary in textual data analysis is a collection of words and associated attributes.

scraping script that gather all the SASB industry-classifications for each company, as those are not available in any structured database or API.

We use guidelines from SASB for ESG materiality classification because it has been validated by Khan et al. (2016) to have predictive power of future returns. They find that companies that focus on material issues outperformed in terms of stock returns while companies that focused on immaterial issues did not. The researchers used SASB's standard reports as a guideline for determining which issues are material and immaterial for a given industry. "Our results serve as a way to validate whether SASB's output has any significant predictive power over future financial performance." – Khan et al. (2016).

4.1.4 Yahoo Finance

Yahoo is a global media and tech company with a worldwide user base. As part of the Yahoo network, Yahoo Finance is a provider for financial news and data.

We collect share prices from Yahoo Finance because they provide pre-calculated adjusted closing prices (Yahoo, 2022). Adjusted closing price is adjusted for dividend payments and stock splits. If a dividend-paying company pays a 4% dividend on a given date, the share price will fall accordingly. If we do not adjust for this, the dividend-paying companies will have systematically lower returns than the market in our models. Refinitiv also provides adjusted closing price, but they are only adjusted for stock-splits (not for dividends). Yahoo Finance provide price data throughout our sample from 2008 - 2020.

The daily adjusted stock prices are converted to daily returns by the formula:

$$r_t = \frac{P_t}{P_{t-1}} - 1$$

where r_t is returns at time t and P_t is adjusted closing price at time t

4.1.5 Kenneth R. French – Data Library

The Fama & French (2015) five-factor model aims to describe stock returns and the factors will act as independent variables in our regressions. We have downloaded timeseries data for the European Five-Factor + Momentum Model from Kenneth R. Frenchy’s data library

(2022). The intuition behind these factors is discussed in more detail in the methodology section.

4.2 Dictionary Creation

One of the most straightforward applications of textual analysis is to identify a word or phrase and simply tabulate the presence of this phrase in a financial document (Loughran & McDonald, 2016). We follow Loughran and McDonald's procedure and tabulate the presence of industry specific ESG-related bigrams[7] in annual reports. What is defined as "ESG-related bigrams" is determined by a dictionary that is specially designed to identify ESG-related bigrams.

4.2.1 Existing Dictionaries for Finance

Within textual analysis, there are a few established dictionaries, most of which are dictionaries that show words attributable to sentiment. The first dictionary created specifically for financial texts is The Henry (2008) Word List. This dictionary is short and has only 85 words which are attributed with negative sentiment. In comparison, the Harvard General Inquirer (GI) Dictionary has 4100 words which are attributed as negative, but the Harvard GI list is not made specifically for financial texts (Loughran & McDonald, 2016). One of the most recognized sentiment dictionaries for financial texts is the Loughran and McDonalds (2011) Word List containing 354 positive words and 2329 negative words.

Although there are established dictionaries for sentiment, there are few dictionaries for ESG. Baier & Florian (2020) provide an ESG dictionary that consist of 482 words which are all attributed as ESG related. However, there are no attributes showing which industries the words are material for. Their dictionary consists of single words called unigrams. We prefer bigrams which is a combination of two words. Bigrams is commonly used in textual analysis when it is believed that the sequencing of words in a text-document contain information. The reason why we prefer bigrams will be explained further in the pre-processing section. The lack of ESG dictionaries lead us to create a new dictionary that suits our purpose.

[7] An n-gram is a sequence of n words. A unigram is one word, bigram is two words, trigram is three words, etc.

4.2.2 The ESG Dictionary

Textual analysis is a method that quantifies qualitative data. The textual data is subjective, and the method can therefore be prone to errors. "The imprecision of textual analysis is not something that precludes its usage but is a characteristic that must be confronted in producing empirical results that are expected to have credible impact and that can be reasonably replicated" - Loughran & McDonald (2016). The methodology we use is innovative and new which calls for a detailed explanation for the research to be replicable and trustworthy. The explanation of how we create the dictionary will thus be of a detailed nature.

The dictionary's purpose is to identify an ESG-related bigram and the attributed industries which the bigram is material to. We start by creating a data frame with two columns: one column is reserved for bigrams and the other column will contain the attributable industries. We run all 77 SASB's industry materiality documents through a pre-processing script where we subsequently transform the text into bigrams. The transformation procedure will be explained in the pre-processing chapter. The script will take note of which industry each individual bigram is material to from SASB's documents. Initially our dictionary contains 217,477 bigrams, where 74,859 of them are unique. This is because some bigrams are material for several industries. For example, the bigram "Board composition" is material for all 77 industries and thus 77 of the rows will represent this bigram.

The dictionary is quite large and contains many bigrams which is not associated with ESG. This is because the dictionary is a product of all the words in the SASB reports, and a comprehensive report will also include words that cannot be tied to ESG. We remove these bigrams by manual reviewal of the dictionary. After manually removing terms that we do not see fit, we are left with 1,151 unique bigrams, each connected to one or more industry. The new dictionary consists of 15,260 rows, which means that each bigram is at average material for approximately 13 industries. Chapter 4.2.4 provides a detailed description of how we reduced the dimensionality of dictionary.

The concept of the dictionary can be clarified by the following example. The Alcoholic Beverages Report from SASB describes that companies in the alcoholic beverages industry should measure «water withdrawn» and «water consumed». Since our dictionary consist of all bigrams from all the industry reports, every company who mentions «water withdrawn» and «water consumption» in their annual report and has these bigrams as a material issue will be rewarded with a positive impact on their score. «Water consumption» will not have a positive impact on the consumer bank industry if they should mention it in their annual report because it is not a material issue for that industry.

4.2.3 Pre-Processing

Data pre-processing refers to the manipulation of data before it is used, to ensure performance. We pre-process the industry reports provided by SASB and annual reports to remove unwanted characteristics of the text that do not provide information of value to our analysis. The pre-processing is done in R using the package "tm" by Feinerer & Hornik (2020), which provides functions that allow us to automate the pre-processing.

First, we remove numbers, as our score do not consider quantitative measures in annual reports. Then remove punctuations because we do not find any informative value in punctuations. For example, we are not able to differentiate between "Energy!" and "Energy." Furthermore, we remove "/n", which means "next line" in computer language. We transform all our letters into letters with lower case as we do not differentiate between words with upper and lower case. We remove stop words which is words like "and", "is" and "the" because these words will have no informative value when we look for ESG related bigrams.

Lastly, we stem the documents. Stemming is a technique where words with different endings is shrunk into one single root. We do this because we do not differentiate between words with different endings. For example, "company" and "companies" provide similar meaning and after stemming the root word is "compani". If an annual report mentions a word that is present in the dictionary but presented in a grammatical way that is not recognized, we will lose valuable information. The downside of stemming is that in some rare cases, two words with different meaning is shrunk to the same root such as "business" and "busy" which becomes "busi".

We have exemplified the pre-processing of a paragraph from the Meat, Poultry & Dairy SASB (2020) report. The paragraph is just a small sample from a total of 36 pages.

Table 1: Pre-processing example

Initial text:	*"As water scarcity becomes an issue of growing importance due to population growth, increasing\nconsumption per capita, poor water management, and climate change, companies in the industry\nmay face higher operational costs or lost revenues due to water shortages and/or regulations\nresulting in production reduction\n"*
After removing numbers, punctuation, whitespace, and capital letters:	*"as water scarcity becomes an issue of growing importance due to population growth increasing consumption per capita poor water management and climate change companies in the industry may face higher operational costs or lost revenues due to water shortages andor regulations resulting in production reduction "*
After removing stop words:	*" water scarcity issue growing importance due population growth increasing consumption capita poor water management climate change companies industry operational costs lost revenues due water shortages andor regulations resulting production reduction "*
After stemming:	*"water scarciti issu grow import due popul growth increas consumpt capita poor water manag climat chang compani industri oper cost lost revenu due water shortag andor regul result product reduct"*
Result (Bigrams):	*'water_scarciti' 'scarciti_issu' 'issu_grow' 'grow_import' 'import_due' 'due_popul' 'popul_growth' 'growth_increas' 'increas_consumpt' 'consumpt_capita' 'capita_poor' 'poor_water' 'water_manag' 'manag_climat' 'climat_chang' 'chang_compani' 'compani_industri' 'industri_oper' 'oper_cost' 'cost_lost' 'lost_revenu' 'revenu_due' 'due_water' 'water_shortag' 'shortag_andor' 'andor_regul' 'regul_result' 'result_product' 'product_reduct'*

There is a trade-off using statistical packages to solve pre-processing automatically. It saves a lot of time, but errors will happen and sometimes they are hard to spot. In the example, the raw text "shortages and/or regulations" becomes the bigrams 'shortag_andor' and 'andor_regul'. This happens because we remove the punctuation so that “and/or” becomes “andor”, which is included as a separate word. This specific example will have low impact for the results. The word “andor” will probably not be mentioned in a single annual report, and especially not in the bigram context and thus not lead to a change in score.

Some bigrams such as "result_product" are included in the dictionary without it having anything to do with ESG. We will explain in chapter 4.2.4 how we handle the irrelevant bigrams.

We use bigrams (composition of 2 words) instead of unigrams because we assume that the word sequence is important. The advantage of bigram over unigram is that we get valuable information in the form of word context. In the example above, we see "water_scarsiti" which gives a different meaning than the words "water" and "scarsiti" would do if the words stood separately. Another sequency of 2 words which is common in annual reports is "board_structure" which gives another meaning than “board” and “structure” separately. The disadvantage of using bigrams over unigrams is that we lose valuable information where word sequence has limited or no informative value. The word "ethic" is ESG-related in itself, and if the company talks a lot about ethics in its annual report without it being in the context we have defined in our dictionary, information will be lost.

4.2.4 Reducing the Dimensionality of the Dictionary

Our raw dictionary has a dimensionality of 2 x 217,477 where many bigrams are not relevant in an ESG context. To improve the dictionary, we manually remove the unwanted bigrams. The process of manually removing unwanted bigrams is prone to some errors because we do not have perfectly control over which bigrams is actually irrelevant in an ESG context. This is an issue because these bigrams must be removed on subjective assessment from us. An example of a hard term to assess, is the stemmed bigram “hydraul_fractur” which probably comes from the unstemmed bigram “hydraulic fracturing”. This is a method of mining petroleum that is harmful for the environment and thus associated with bad ESG practice. However, after investigating the companies which mention this bigram, we find that these companies reported on yearly oil spillage due to the use of hydraulic fracturing. Disclosure of oil spillage is associated with good ESG practice. The following questions then rises. Should we give a positive score because they are open about their oil spillage? Or should they receive a low score because they use Hydraulic fracturing? In this specific case we give positive score. In most cases we include bigrams that could be both good and bad in a sustainability context. We give positive scores to these companies because we believe that mentioning words that intuitively may seem harmful for ESG, most often can be associated with ESG disclosure. Transparency about ESG issues is rewarded in our ESG score and the score we construct can thus also be considered a proxy for ESG disclosure.

The raw ESG dictionary consists of 74,859 unique bigrams, and we cannot effectively assess all the bigrams manually. We must therefore see which bigrams have the greatest impact on the companies' ESG score and prioritize these bigrams. Loughran & McDonald (2016) argues that identifying and reviewing bigrams that are mentioned most often can reduce the likelihood of misclassification.

We identify which of the bigrams from the raw ESG dictionary are mentioned most often in the annual reports. The bigrams most often mentioned from each report will, by definition, be those which give most impact on the score. We will then gather 20 % of the most mentioned bigrams from each company in our sample. This gives us 134 lists of the most mentioned bigrams for every company. The reason why we want to include the top bigrams for every single company instead of just including the top bigrams from all the companies combined is because some bigrams may be mentioned very often in a specific industry but not by our sample combined. Thus, these bigrams will not necessarily make up 20% of the most mentioned bigrams from all the reports, yet these bigrams will be very important for the specific industry. Finally, we extract all the unique bigrams from all the 134 lists. The extracted bigrams now give us a complete list of 3,225 bigrams that asses through manually.

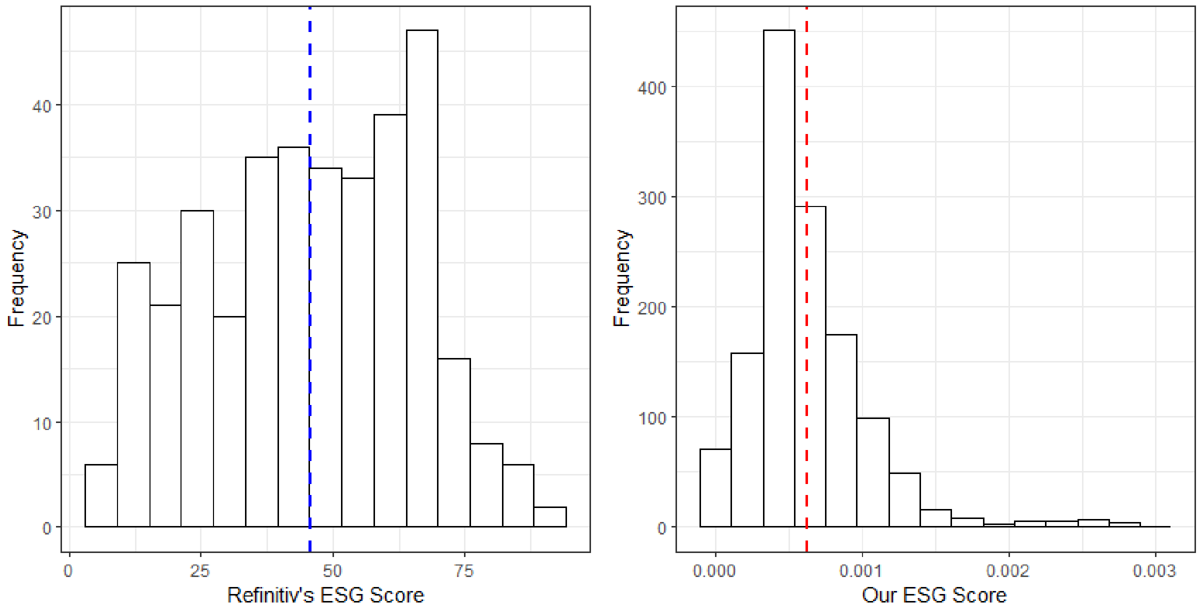
The process of gathering the most mentioned bigrams reduces the dictionary from 74,859 to 3,225 unique bigrams. The remaining 71,634 unique bigrams which we discard probably never got mentioned in an annual report, or at least got mentioned so few times that they do not have a significant effect on the score.

It is still difficult to effectively assess 3,225 bigrams manually and we want to prioritize precise assessment of the most important bigrams. We do not have the time or ESG-domain specific expertise that is needed to do a critical and precise assessment of all the 3,225 bigrams. Loughran & McDonald (2016) state that "Zipf's law documents the fact that in any non-pathological list of words, a very small number of words will dominate the frequency counts. This property of word distributions creates a research environment where seemingly innocent word misclassifications do not simply add small amounts of random noise to the results and can produce outliers that drive spurious results." Based on a study of 10-K / Q filings, they find that 1% of the words from the dictionary, account for 44% of the scoring words in the filings. Therefore, we prioritize to be extra careful with the 50-100 bigrams that is mentioned by most annual reports. The prioritizing is possible because the bigrams are sorted by how often they are mentioned. From the 3,225 bigrams, we remove 2,074 unwanted bigrams manually and we are therefore left with 1151 subjectively approved unique bigrams. (See appendix 1 for the list of the 1151 bigrams that make up the dictionary.)

4.3 The ESG Score

When researchers measure sentiment or tone of a financial document, they typically count the presence of a word associated with a particular sentiment and scale the wordcount by the total number of words in the document (Loughran & McDonald, 2016). We copy this approach and our ESG score is given by how many times material ESG related bigrams are mentioned divided by the total number of bigrams in the annual report. The advantage of a scoring type like this is that shorter annual reports doesn't necessarily underperform in terms of score.

Figure 3: ESG score distribution



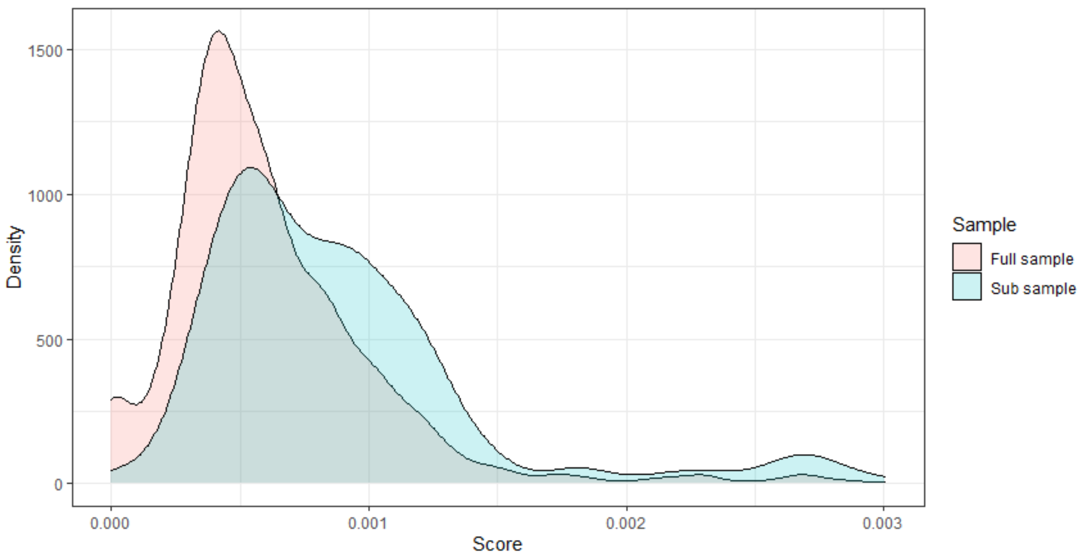
The histograms display the score distribution of Refinitiv's ESG score and our Materiality score. The blue and red dotted line shows the averages for the respective scores.

Figure 3 display the distribution of ESG scores in Refinitiv's database for Norwegian companies, and our sample respectively. Our score can go from 0 to 1 in theory, and a score of 0.002 means that for every 1000th bigram in the annual report, an ESG-relevant bigram is mentioned 2 times. The distribution show that the ESG score has a long tail on the right side where some companies score as high as 0.003. The average score is 0.0006

Refinitiv's score goes from 0 to 100 and the average score for Norwegian companies is slightly below 50. In comparison, Refinitiv's international average is 50, which means that Norwegian companies in general do not show significant differences in ESG performance compared to international companies. (Refinitiv, 2022)

Refinitiv has provided an ESG score for a total of 64 companies on Oslo Stock Exchange in 2020. Out of those 64 companies, we have provided score for 62 which means that there are 2 companies in Refinitiv's ESG database for which we could not find an English annual report. In total, we have provided a score for 134 companies. Figure 4 is a density plot of the ESG score of our sample compared to the ESG score of a subset of our sample. The subsample consists of those 62 stocks that both we and Refinitiv can provide score for.

Figure 4: Density plot of full sample and subsample



The density plot display score-density of our full sample compared to a subsample consisting of the part of our sample that Refinitiv is also able to provide a score for. This means that the sub sample consists of 62 companies while the full sample consists of 134 companies.

We can see that the subsample is skewed to the right in the density plot. In particular, the extremely high scores between 0.002 and 0.003 from our sample, seem to come from companies that are also present in Refinitiv's ESG database. Furthermore, we see that all the companies that get very low scores in our sample is not present in Refinitiv's ESG database. These findings suggest that Refinitiv's ESG database systematically perform better at ESG than our full sample. This makes intuitively sense because Refinitiv's ESG database only include companies which choose to disclose ESG metrics, and it is plausible that these companies are systematically better at ESG than the companies that keep their ESG metrics undisclosed.

Doyle (2018) find that large companies tend to get high ESG scores. To see if this also applies on the Oslo Stock Exchange, we have divided Refinitiv and our own sample into 3 equalsized parts, sorted by market capitalization. Note that the sample changes for every year and this is an image of how the sample looks in 2020. Refinitiv's sample is represented in table 2 and display the mean of the 3 pillars and the aggregated ESG score. Table 3 display the mean of our ESG score.

Table 2: Descriptive statistics of Refinitiv ESG scores

	Small cap	*Medium cap*	*Large cap*
N	*21*	*22*	*21*
Environment (Mean)	*28.02*	*40.94*	*63.4*
Social (Mean)	*41.92*	*43.94*	*69.01*
Governance (Mean)	*42.83*	*42.27*	*64.32*
ESG Score (Mean)	*37.45*	*42.9*	*62.3*
Min Market Cap (Millions)	*652*	*6,965*	*25,756*
Max Market Cap (Millions)	*6,696*	*21,842*	*594,114*

The small cap ranges from 652 m to 6,696 m, the medium cap ranges from 6,965 m to 21,842 m and the large cap firms range from 25,756 m to 594,114 m

Table 3: Descriptive statistics of our ESG score

	Small cap	*Medium cap*	*Large cap*
N	*44*	*46*	*44*
ESG Score (Mean)	*0.63*	*0.75*	*0.93*
Min Market Cap (Millions)	*31*	*1,541*	*6,965*
Max Market Cap (Millions)	*1,502*	*6,933*	*594,114*

The small cap firms range from 31 m to 1,502 m, the medium cap ranges from 1,541 to 6,933 m, and the large cap firms range from 6,965 m to 594,114 m

The tables show that the average ESG score is highest for large cap companies and lowest for low cap companies. This is consistent in our sample and in Refinitiv's ESG database. This relationship substantiates Doyle's (2018) finding that large companies score better on ESG.

Small cap, medium cap and large cap are relative sizes and constitute the 33% lowest, 34% medium and 33% highest sorted by market capitalization in relation to the sample. The companies we categorize as small cap (31 m – 1 502 m) generally have a much lower market capitalization than the companies that Refinitiv categorizes as small cap (652 m – 6 696 m). The lowest threshold (6.965 m) for being a medium cap in Refinitiv's database is the same as the lowest threshold for being a large cap in our sample.

We argue that Refintiv's ESG database is not representative of the Oslo Stock Exchange, but systematically consists of companies with sufficient ESG reporting and that these companies generally have a larger market capitalization than Oslo Stock exchange. Our findings are consistent with the research of Drempetic et al. (2020) who find a sample bias towards large companies in Refinitivs ESG database. Our sample consists of 134 of the total 209 listed companies and will probably be a better representation of Oslo Stock Exchange.

Figure 5: Average ESG score from 2008 to 2020

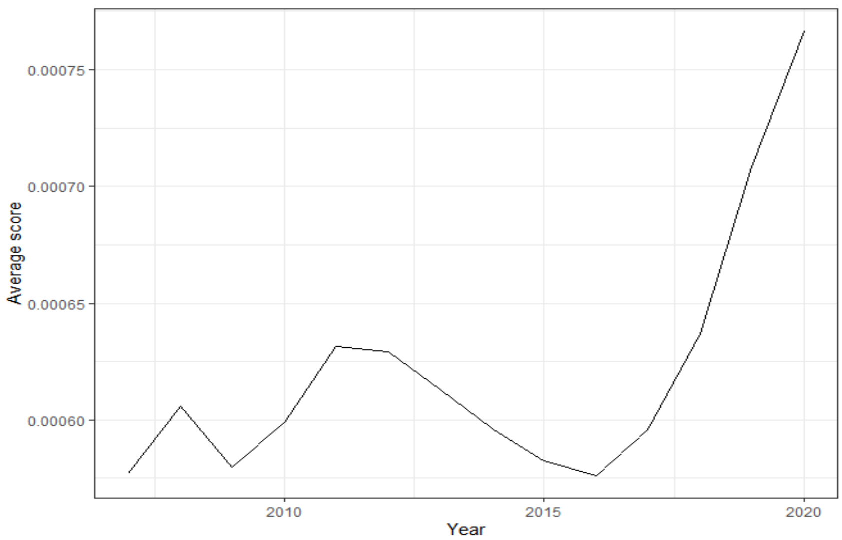


Figure 5 illustrates how the constructed average ESG score has developed in the sample period. There is spike in average ESG score in 2019-2020. This means that the companies in our sample have in average a stronger presence of ESG terminology in their annual reports in 2019 and 2020 compared to the previous years.

In figure 2, we illustrate how Refinitiv's ESG database grew significantly in 2020. We argue that regulations related to EU taxonomy required companies to disclose ESG metrics, and that companies' increased ESG disclosure led to an increase in Refinitv's ESG database. The development in our ESG score coincides with that assumption because we have previously argued that our score can be used as a proxy for ESG disclosure. We suggest that the increase in the average ESG score in 2019 and 2020 is due to increased disclosure on ESG metrics, which happened because of EU taxonomy.

4.4 Validation of our ESG Score

The use of textual analysis on annual reports to measure material ESG performance has not been done before in known literature. The ESG score is a product of the dictionary we construct and is an innovative score that has not been used in other literature and has thus not been validated. Large parts of this paper are dedicated to cover validation of the score.

Our score makes sense intuitively. It is plausible that presence of material ESG terminology in an annual report can measure ESG performance for a company. But to substantiate the claim, we will make a regression analysis where we use our score as an independent variable to explain the ESG score from an external ESG provider. If our score has predictive power over an established score, we can argue that our score capture ESG performance.

4.4.1 Score Validation Regression

We validate our score by comparing it to an external provider's sustainability score for the same companies. We use Refinitiv as the external provider and the regression model is given by,

$$REF_i = c + ESG_i + time_i + e_i$$

where REF_i is Refinitiv's sustainability score, ESG_i is our constructed ESG score, time is years until 2020, c is the intercept and e_i is a zero-mean residual.

Table 4: Score validation regression

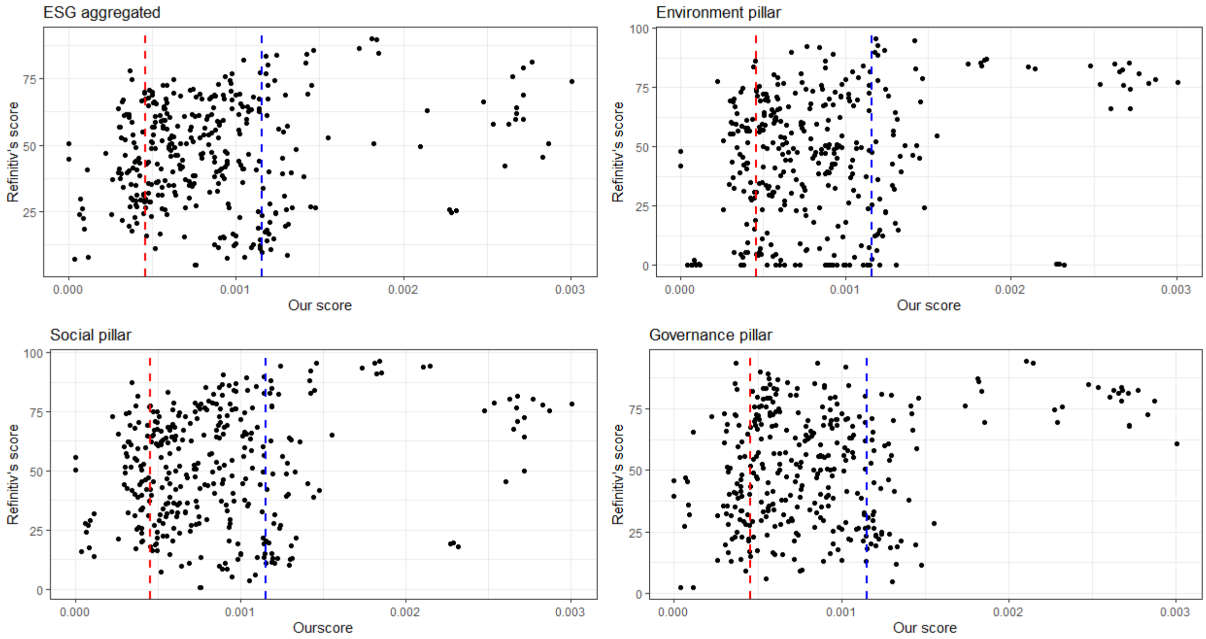
	Dependent variable: Refinitiv's Score			
	Governance Pillar (Refinitiv)	Social Pillar (Refinitiv)	Environment Pillar (Refinitiv)	Aggregated ESG (Refinitiv)
	(1)	(2)	(3)	(4)
OUR SCORE	10,851.680***	9,270.379***	12,924.000***	6,472.040***
	p = 0.00000	p = 0.0001	p = 0.00000	p = 0.001
TIME (YEARS TO 2020)	0.314	-0.562**	-0.246	-0.356*
	p = 0.180	p = 0.024	p = 0.382	p = 0.081
CONSTANT	40.008***	45.909***	35.338***	43.039***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
OBSERVATIONS	349	349	349	349
R2	0.072	0.063	0.072	0.045
ADJUSTED R2	0.066	0.058	0.067	0.039
NOTE:	*p < 0.1, **p < 0.05, ***p<0.01			

The coefficient for our score can be interpreted as follows: If a company's annual report increase its presence of ESG bigrams by 1/1000, on average Refinitiv will increase its governance score by 10.85, its social score by 9.27, its environment score by 12.92, and its ESG score by 6.47. Furthermore, all the coefficients are significant at the 1% level. These findings suggest that our score succeed to explain Refinitv's ESG measures and capture ESG performance of a company.

4.4.2 Score Correlations

Figure 6 is a graphical representation of the relationship between our score and Refinitiv's score. The dashed red line separates the lower quintile of our score to the left, and the blue dashed line separates the upper quintile of our score to the right.

Figure 6: Relationship between our score and Refinitiv's score.



Every observation on the left side of the red line represents the 20 % companies that scores lowest on ESG. Every observation on the right side of the blue line represents the 20 % of companies with highest ESG score. The dots in the plot is based on the companies present in both our sample and Refinitiv's ESG database from 2008 to 2020.

The figure suggests that the companies which we give low scores, also receive low scores from Refinitiv, and companies that we give high scores also receive high scores from Refinitiv. Still, there is some noise. We do not expect that the number of times ESG words are mentioned in an annual report coincides 100% with an ESG score given by a third-party agency.

There seems to be less correlation between our score and Refinitiv's score in the companies that get medium scores. In practice, the medium scores will not affect the results as the portfolios we construct are based on top and bottom quintiles.

Table 5 show the correlation between our score and Refinitiv's scores. We do not expect to have a particularly high correlation because the score is given by completely different metrics. Our score shows ESG materiality presence in annual reports, while Refinitiv's score is a weighted average of various ESG measures. Nevertheless, we expect a positive correlation as both scores is intended to measure ESG performance. A positive correlation will validate the score we construct.

Table 5: Correlations between our score and Refintiv’s scores

	ESG Aggregated	*Environment Pillar*	*Social Pillar*	*Governance Pillar*
ESG (Full sample)	0.20	0.28	0.23	0.27
ESG (Top 20%)	0.31	0.39	0.349	0.64
ESG (Bottom 20%)	0.26	0.33	0.24	0.09
ESG (Top and Bottom 20%)	0.32	0.40	0.33	0.49

The table display the correlation between our score (rows) and Refinitiv’s ESG score, environment score, social score, and governance score (columns), for the full sample, top quintile, bottom quintile and both quintiles combined.

Our score correlates with all ESG pillars as well as aggregated ESG score from Refinitiv. Our score has a correlation of 0.32 with Refinitiv's aggregate ESG score on top and bottom quintile. Gibson et al. (2019) find an average correlation of 0.46 between the 6 biggest ESG providers. The correlation between our score with Refinitiv is lower than the correlation between the 6 largest ESG providers, yet it is higher than we would expect considering that our methodology is fundamentally different from the methodology of the big ESG score providers. The relatively high correlation on aggregated ESG validates that we can capture ESG performance with our score.

4.4.3 ESG Portfolio Attributes

If the ESG portfolios we create share characteristics with ESG portfolios in related literature, it can be argued that the score we made has managed to capture ESG performance. In table 3 from the ESG score section we demonstrate a positive link between our score and companies' market capitalization which is consistent with Doyle (2018). We will investigate the characteristics of the ESG portfolios further in the results chapter.

5.0 Empirical Methodology

We construct portfolios based on ESG scores. The purpose is to see if ESG portfolios generates an abnormal return after controlling for other factors that may explain returns. We use the Fama French five-factor model plus momentum where daily excess returns are dependent variable.

We use three different investment strategies to construct portfolios. These strategies are zero-investment, best-in-class, and negative screening. We weigh the share allocation in these portfolios with equal weighting, and weighting based on market capitalizations. Furthermore, we construct portfolios with a restriction that the ESG score must be available at the time of investment. Lastly, we divide the sample in subperiods to test for heterogeneity.

5.1 Fama French Five-Factor Model Plus Momentum

The Fama French five-factor[8] model plus momentum is designed to explain portfolio returns. It is an extension of the traditional Capital Asset Pricing Model (CAPM) introduced by Sharpe (1964) and Lintner (1965) that was designed to explain portfolios' returns but only included the market factor as a risk factor, which in practise mean that a portfolio would only receive a premium for market risk taken. The CAPM equation is given by:

$$R_{it} = R_{ft} + b_i * (R_{mt} - R_{ft})$$

R_{it} is the return of portfolio i at time t, R_{ft} is the risk-free rate at time t, R_{mt} is the market return at time t, and b_i is the coefficient of the market premium.

The intuition of CAPM is that the return of a diversified portfolio is equal to the risk-free rate if the portfolio is risk-free (a coefficient of zero). A higher coefficient means that the security is riskier and should be compensated with the market premium. This model assumes that exposure to the market is the only driving factor of variations in returns of a diversified portfolio.

[8] The Fama French five-factor model plus momentum is developed on several markets. There are separate measured factors for European, Japanese, Asia Pacific ex Japanese, North American and developed ex North American. In this paper we use the European factors, as they most resemble the Norwegian market. The construction of the factors is explained in detail in French (2022) data library descriptions.

Fama & French (1993) argue that CAPM lack explanations of variations in returns and propose the three-factor model that also accounted for firms' size and book-to-market value. (Carhart, 2012) complement the three-factor model and introduce the Carhart four-factor model which included a momentum factor. Later, Fama & French (2015) propose the 5-factor model which include a profitability factor and an investment factor. The Five factor with momentum model is a combination of the five-factor model by Fama & French (2015) and the Momentum factor by Carhart (2012).

The reason why we use the Fama French five-factor model with momentum is two folded: 1) The model is widely used for its ability to explain variations in returns. 2) Fama French's factors can reveal portfolio characteristic such as size and profitability which we can use to validate our score by comparing characteristics we find in our portfolios with characteristics that the literature finds on their ESG portfolios. For instance, that bigger firms tend to perform best at ESG.

The Fama French five factor model plus momentum regression is given by,

$$R_{it} - R_{ft} = \alpha_i + b_i * (R_{mt} - R_{ft}) + s_i * SMB_t + h_i * HML_t + r_i * RMW_t + c_i * CMA_t + w_i * WML_t + e_{it}$$

where R_m - R_f is the market premium and captures the portfolios' exposure to market.

Small minus big (SMB) is a risk factor that capture the relationship between return and size (market capitalization) of stocks. SMB is the average returns of diversified portfolios with small companies minus the average returns of diversified portfolios of big companies. A positive SMB coefficient explains that the portfolio's returns are attributable to the small stock premium.

High minus low (HML) is a risk factor that captures the relationship between return and book/market[9] value. HML is the average returns of diversified portfolios of companies with high book/market value minus the average returns of diversified portfolios of companies with low book/market value. The book value of a company is equal to the equity value as reported in its financial statements. Stocks with high book/market value are considered "value-stocks", while stocks with low book/market value are considered "growth-stocks". This is due to the nature of how stocks are valued in the balance sheet, with a typically lower book value than market value. A low book/market value indicates that the company is expected to grow in

[9] Book-to-Market ratio is a valuation indicator that compares book value to market value.

price and thus called "growth-stock". The return of portfolios consisting of value stocks are attributable to the HML premium.

Robust minus weak (RMW) is a risk factor that capture the relationship between return and profitability. RMW is the average returns of diversified portfolios with robust profitability minus the average returns of diversified portfolios with weak profitability. A positive coefficient towards RMW means that the returns of the portfolio are attributable to the profitability premium.

Conservative minus aggressive (CMA) is a risk factor that capture the relationship between return and investments. CMA is the average returns of diversified portfolios of conservative investment firms minus the average returns of diversified portfolios of aggressive investment firms. A positive coefficient towards CMA means that the returns of the portfolio are attributable to the conservative investing premium.

Winners minus losers (WML) is a risk factor that capture the relationship between return and past stock performance. WML is the average return of diversified portfolios of winners minus the average return of diversified portfolios of losers. A positive coefficient towards WML means that the returns of the portfolio are attributable to the winner premium.

If the exposures to the five factors, b, s, h, r, c and w capture all variation in expected returns, the intercept α is zero for the portfolio (Fama & French, A five-factor asset pricing model, 2015). We apply Fama French Five Factors with Momentum to the investment strategies we present in the next subchapters.

We replace the European market risk factor with a Norwegian market risk factor. This is done to account for the risk adjusted superior performance of Norwegian companies compared to European companies (See appendix 3). If we do not adjust the market factor, we will get erroneous results with an artificially high alpha where we cannot differentiate whether the ESG portfolio receives alpha due to its ESG score, or if it is because the portfolio consists of Norwegian companies.

5.2 The Zero Investment Portfolios

A zero-investment portfolio is composed of securities that cumulatively result in a net value of zero. The portfolio consists of a short-position and a long-position of equal sizes. The model assumes no fees of buying stocks and no costs related to shorting. Excess returns for zero investment portfolios are given by.

$$RE = R_{Long} - R_{Short}$$

Where RE= excess returns, R_{Long} = returns of the long portfolio, and R_{Short}= returns of the short portfolio. We do not need to deduct risk-free rate in the calculation of excess returns because a net zero investment has a net zero alternative cost.

In our case, the zero-investment portfolios (henceforth ZI) has a long position in the top quintile material ESG scoring firms, and a short position in the bottom quintile of the material ESG scoring firms.

5.3 The Best-In-Class Portfolios

The best-in-class approach is to invest in the companies that score the highest in material ESG. The best-in-class approach is used to mitigate risks to poorly performing ESG companies and to gain the benefit of well managed companies. Our best-in-class portfolios (henceforth BIC) consist of the top quintile of material ESG companies.

5.4 The Negative Screening Portfolios

Negative screening involves excluding the companies that are categorized as the worst ESG companies. The portfolio excludes either companies that are categorized as sin-stocks, or the companies that score lowest on ESG. Negative screening is the most widespread strategy for sustainable portfolio composition (Amel-Zadeh & Serafeim, 2018). Our negative screening portfolios (henceforth NS) include 80% of companies that score highest on material ESG. That is, an exclusion of the bottom quintile.

5.5 Equal and Value-Weighted Portfolios

We construct equal and value-weighted portfolios to investigate if our results are sensitive to company weights.

An equal-weighted portfolio is constructed based on the criteria that all companies in the portfolio have the same weighting. Equal weighting offers protection to a portfolio's returns if a large company or sector experience a drop in stock price. The weighting of each firm in an equal-weighted portfolio is given by the equation:

$$w_i = \frac{1}{n}$$

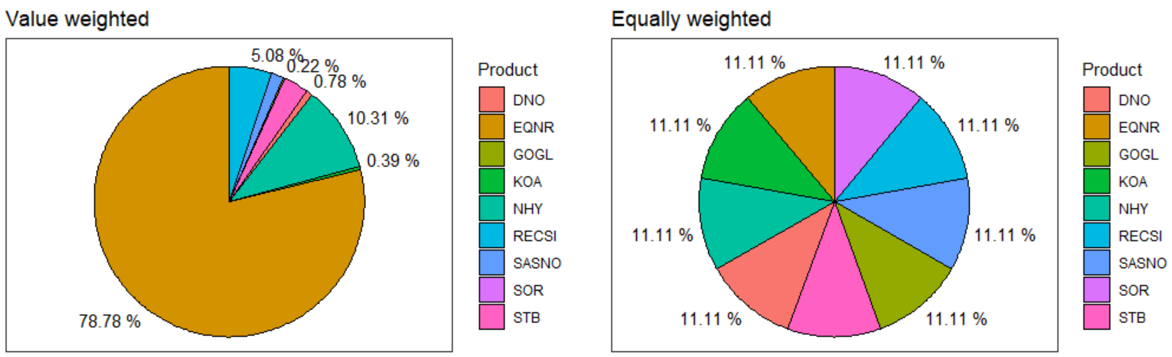
Where w_i is the weight of firm i, and n is the total number of companies in the portfolio.

Value-weighted portfolios' company weightings are based on the companies' market value. Value-weighted portfolios can thus benefit from a large company or sector's upswing. The weighting of each firm in a value-weighted portfolio is the market capitalization of the company divided by the sum of all market capitalizations for the portfolio. The weighting for a company in a value-weighted portfolio is given by the equation:

$$w_i = \ market\ capitalization_i / \sum_{i=1}^{n} market\ capitalization_i$$

Where w_i is the weighting of firm i, market capitalization is the amount of shares outstanding × stock price of firm i, and n is the total number of companies in the portfolio.

Figure 7: Portfolio weights in 2010 for the Best-In-Class portfolio



Some percentage-labels have been removed from the value-weighted chart due to overlapping. The removed percentages are: SOR = 0.04%, SASNO = 1.37%, STB = 0.39%

Figure 7 illustrates the implications of the value-weighted portfolios on the Oslo Stock Exchange. Equinor (EQNR) received a high ESG score in all years in our sample and is therefore represented in all the portfolios. This company has a large market capitalization compared to the other companies in our sample and thus value-weighted portfolios will be very much affected by how Equinor’s stock price develops. Value-weighted portfolios are subject to firm specific risk arising from Equinor's unsystematic fluctuations.

5.6 Practical Portfolios

The investment strategies and portfolio compositions presented so far are theoretical portfolios that show a connection between ESG scores in one year and performance in the same year and are not possible to implement in practice. This is because the ESG scores used to form a portfolio are not available until 1-2 years after the date of investment. For example, the annual report for 2010 is published in the time interval January 2011 - May 2011.

We construct equal-weighted portfolios that is possible to implement in practice where we use ESG scores available at the time of investment. The portfolio composition is thus based on 2-year-lagged ESG scores. The analysis from this type of portfolio will not show a connection between ESG score and stock performance in a year, but rather a connection between ESG score and future stock performance.

5.7 Rebalancing

All portfolios are rebalanced annually because the ESG scores are based on annual reports that are published annually. The rebalancing allows the portfolio composition to change over time. The number of shares in our sample increases over time and thus the quintiles on which the portfolios are based will also increase over time. Table 6 shows the percentage overlap from one year to another in the top and bottom quintile portfolio.

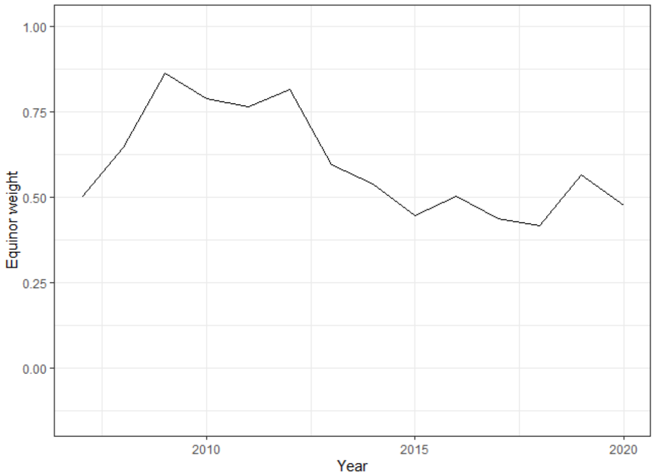
Table 6: Overlapping by time

	Overlap Long portfolio	Overlap Short portfolio
2008-2009	62.5%	62.5%
2009-2010	75%	50%
2010-2011	77.7%	77.7%
2011-2012	90.1%	54.5%
2012-2013	75%	58.3%
2013-2014	100%	76.9%
2014-2015	80%	60%
2015-2016	83%	72.2%
2016-2017	75%	70%
2017-2018	76%	80.9%
2018-2019	75%	70.8%
2019-2020	88%	64%

The percentage of portfolio overlap is generally high from one year to another. We see that 100% of the shares in the long-term portfolio from 2013 overlap with the shares in 2014. This indicates that the companies' ESG score does not change much relative to the sample from one year to the following year. Out of the 134 companies in the sample, 100 companies are at some point represented in the top or bottom quintile. This means that 34 companies remain among the middle 60% throughout the period.

There are a total of 4 companies that remain in the top ESG scoring quintile throughout the whole period. This is a substantially amount considering that the portfolio consisted of 8 companies in 2008. These companies are DNO, Equinor, Norsk Hydro and Storebrand. There were no companies that remained in the bottom quintile throughout the sample period. This suggest that high ESG performers seem to be more consistent than low ESG performers in terms of ESG presence in their annual report.

Figure 8: Equinor weighting over time



The figure illustrates how Equinor's weighting ranges from 40% to 85% throughout the period 2008-2020 in the value-weighted BIC portfolio.

We have already argued how Equnior dominates the value-weighted portfolios due to its large market capitalization. Equnior's dominant position is further strengthened by the fact that Equnior maintains its position in the top quintile over the entire sample period. This can have a major impact on the value-weighted portfolios.

6.0 Results

In this section, we present our regression results. We test in total 9 material ESG portfolios in the period 2008-2020 (three equal-weighted, three value-weighted and three practical portfolios) for abnormal returns and what risk factors these portfolios possess as according to the Fama French five factor model with momentum. We also test the three equal-weighted portfolios in the subperiods 2008-2014 and 2015-2020.

We examine the value and significance of the α as this is an expression of the portfolio's abnormal returns. Secondly, we will examine the coefficients and significance of the factors as these can express characteristics in the portfolios.

What distinguishes the regression models is how the portfolios are constructed. The first table presents regression output for equal-weighted portfolios, the second presents value-weighted portfolios, followed by practical portfolios. Finally, we show regression output of equal-weighted portfolios for subperiods of the data, i.e., from 2008 to 2014 and from 2015 to 2020.

6.1 Fama French Five-Factor Plus Momentum on Equal-Weighted Portfolios

Table 7: Fama French five-factor plus momentum on equal-weighted portfolios

	Dependent variable: Ri-Rf		
	ZI	BIC	NS
Rm-Rf	0.085***	0.546***	0.488***
	p = 0.004	p = 0.000	p = 0.000
SMB	-0.510***	-0.321***	-0.037**
	p = 0.000	p = 0.000	p = 0.013
HML	0.384***	0.191***	0.057***
	p = 0.00001	p = 0.0002	p = 0.007
RMW	0.326***	0.182***	0.011
	p = 0.002	p = 0.007	p = 0.712
CMA	-0.588***	-0.367***	-0.116**
	p = 0.00002	p = 0.0002	p = 0.036
WML	-0.139***	-0.080***	-0.014
	p = 0.004	p = 0.003	p = 0.231
α	-0.0004**	-0.0003**	-0.00004
	p = 0.048	p = 0.023	p = 0.496
Observations	3,264	3,264	3,264
R2	0.176	0.786	0.952
Adjusted R2	0.174	0.785	0.951
Note:	*p < 0.1, **p < 0.05, ***p<0.01		

This table presents the results on Fama French five-factor model with momentum. The table show risk factor loadings and daily abnormal returns (α) on three equal-weighted material ESG investing strategies constructed with ESG score. The regression has excess returns over the risk-free rate as dependent variable. R_m-R_f is the market factor, SMB is the size factor, HML is the value factor, RMW is the profitability factor, CMA is the investment factor and WML is the momentum factor. The α of the models represents any effects that ESG will have on the portfolio's abnormal returns. The p-values are based on robust standard errors.

The regression output shows significant negative abnormal returns for the ZI portfolio (α = -0.0004, p<0.05) and the BIC portfolio (α = -0.0003, p<0.05) which is contrary to what we expected to find.

The market's risk premium is expressed through the market factor R_m-R_f. The factor exposure is significant for all portfolios, and with least exposure for the ZI portfolio (β = -0.085, p<0.01). A coefficient close to 0 for the zero-investment portfolio is expected because it does not have a positive net exposure in the market. This happens because the long portfolio is financed by shorting the corresponding amount. The low market coefficient values indicate that the portfolios consist of companies with low market risk[10].

The SMB factor is significant negative for all the portfolios: ZI (β = -0.510, $p<0.01$), BIC (β = -0.321, $p<0.01$) and NS (β = -0.037, $p<0.05$). The NS portfolio's low coefficient means that the portfolio is just marginally overweighted by big companies compared to the ZI and BIC which is heavier weighted against larger companies. This negative SMB exposure is expected. Table 3 in chapter 4.5 display that companies with high market capitalization have higher average ESG scores than companies with low market capitalization. The finding is also consistent with literature in the field which unequivocally argues that companies with high ESG scores tend to be big companies (Doyle, 2018).

We find a significant positive exposure to the HML factor for all of the portfolios. Overall, the portfolios' returns covariate with the returns of portfolios of value-stocks with high book-to-market value.

We find a significant exposure to the RMW factor for the ZI portfolio (β = 0.326, $p<0.01$), indicating that our long position in companies with high ESG-score is characterized as firms with robust profitability, and our short position in companies with low ESG score is characterized as firms with weak profitability. We also find a significant exposure to the RMW factor for the BIC portfolio (β = 0.182, $p<0.01$), which also substantiates that companies with high ESG-score tend to have robust profitability. Giese, Lee, Melas, & Nishikawa (2019) find evidence supporting the assertion that high ESG-rated companies are more profitable.

Furthermore, we find a significant negative exposure to the CMA factor for all portfolios which indicates that these portfolios' returns covariate with the returns of portfolios of aggressive investing companies. Lastly, we find a significant negative exposure to the WML factor for the ZI portfolio (β = -0.139, $p<0.01$), and BIC portfolio (β = -0.080, $p<0.01$), indicating that these portfolios to some extent behaves like the past loser portfolios.

The adjusted R squared is a measure of how well the variables can explain the portfolio's excess returns. The adjusted R squared is lower for the zero-investment portfolio because the portfolio is not positively exposed to the market, thus the market risk premium will not

[10] We have isolated Rm-Rf in a CAPM model to say with certainty that the portfolios have a market beta that is lower than 1 (See Appendix 13)

explain the returns. The high adjusted R squared for the negative screening portfolios can be explained by the fact that the portfolios hold 80% of the companies in the market, and thus the market's returns will explain large parts of the portfolio's returns.

6.2 Fama French Five-Factor Plus Momentum on Value-Weighted Portfolios

Table 8: Fama French five-factor model plus momentum on value-weighted portfolios

	Dependent variable: Ri-Rf		
	ZI	BIC	NS
Rm-Rf	0.095***	0.518***	0.503***
	p = 0.0001	p = 0.000	p = 0.000
SMB	-0.442***	-0.186**	-0.090**
	p = 0.00000	p = 0.046	p = 0.039
HML	-0.014	-0.072	-0.012
	p = 0.923	p = 0.635	p = 0.858
RMW	0.236	0.069	-0.031
	p = 0.110	p = 0.694	p = 0.722
CMA	0.069	0.142	-0.005
	p = 0.619	p = 0.134	p = 0.929
WML	-0.051	-0.067	-0.044*
	p = 0.460	p = 0.233	p = 0.071
α	-0.0003	-0.0002	-0.0001
	p = 0.236	p = 0.241	p = 0.199
Observations	3,264	3,264	3,264
R2	0.115	0.816	0.942
Adjusted R2	0.114	0.816	0.942
Note:	*p < 0.1, **p < 0.05, ***p<0.01		

This table present the results on Fama French Five-Factor model with momentum. The table show risk factor loadings and daily abnormal returns (α) on three value-weighted material ESG investing strategies constructed with ESG score. The regression has excess returns over the risk-free rate as dependent variable. R_m-R_f is the market factor, SMB is the size factor, HML is the value factor, RMW is the profitability factor, CMA is the investment factor and WML is the momentum factor. The α of the models represents any effects that ESG will have on the portfolio's abnormal returns. The p-values are based on robust standard errors.

The regression output displays no significant abnormal returns for any of the value-weighted portfolios. We find a significant relationship to the market factor for all portfolios at the 1% level, showing similar results as the equal-weighted with least market factor exposure to the zero-investment portfolio.

The regression output displays a significant negative SMB exposure for all portfolios indicating that the return of these portfolios tilt more towards the returns of portfolios with big stocks. The negative SMB exposure is consistent between equal and value-weighted portfolios.

The differences between the value and equal-weighted portfolios become more apparent when looking at the HML, RMW, CMA and WML factors. While the equal-weighted portfolios show a positive and significant exposure to the HML factor, the value-weighted portfolios do not. The value-weighted portfolios are also insignificantly exposed to the RMW factor, and only the NS portfolio show a significant exposure to the WML factor, which is the opposite case for the equal-weighted portfolios.

The difference between the abnormal returns, factor magnitude and significance between the equal-weighted and value-weighted portfolios suggest that our portfolios' risk exposure are sensitive to company weights. We believe that equal-weighted portfolios give more valid results than value-weighted portfolios. Figure 8 in the data section displayed that Equinor represent between 40% and 85% of the value-weighted portfolios constructed by the top quintile of ESG scoring companies. The value-weighted portfolios may not be diversified enough to exclude unsystematic company-specific fluctuations. Based on this, we will consider the results from value-weighted portfolios with caution and emphasize results from the equal-weighted portfolios.

6.3 Fama French Five-Factor Plus Momentum on Practical Portfolios

Table 9: Fama French five-factor model plus momentum on practical portfolios

	Dependent variable: Ri-Rf		
	ZI	BIC	NS
Rm-Rf	0.098***	0.545***	0.489***
	p = 0.002	p = 0.000	p = 0.000
SMB	-0.371***	-0.300***	-0.046**
	p = 0.000	p = 0.000	p = 0.012
HML	0.235*	0.174*	0.018
	p = 0.068	p = 0.098	p = 0.491
RMW	0.129	0.136	-0.053
	p = 0.418	p = 0.320	p = 0.157
CMA	-0.180	-0.198**	-0.062
	p = 0.163	p = 0.041	p = 0.301
WML	-0.123**	-0.092**	-0.024
	p = 0.030	p = 0.012	p = 0.112
α	-0.0002	-0.0001	0.00001
	p = 0.273	p = 0.377	p = 0.926
Observations	3,264	3,264	3,264
R2	0.121	0.746	0.926
Adjusted R2	0.119	0.746	0.926
Note:			*p < 0.1, **p < 0.05, ***p<0.01

This table present the results on Fama French Five-Factor model with momentum. The table show risk factor loadings and daily abnormal returns (α) on three equal-weighted practical material ESG investing strategies constructed with ESG score. The regression has excess returns over the risk-free rate as dependent variable. R_m-R_f is the market factor, SMB is the size factor, HML is the value factor, RMW is the profitability factor, CMA is the investment factor and WML is the momentum factor. The α of the models represents any effects that ESG will have on the portfolio's abnormal returns. The p-values are based on robust standard errors. What distinguishes this regression output from the regression output in chapter 6 is that these portfolios are restricted to rely solely on ESG scores that are available at the time of investment. The portfolio's composition is a result of the scores the companies received 2 years prior to investment date. The regression thus shows the relationship between ESG score and future performance.

The regression output displays no significant abnormal returns for any practical portfolio in the full sample, however we find negative abnormal returns in the subperiod 2008-2014 for the ZI ($\alpha = -0.0005$, $p = 0.112$) and BIC ($\alpha = -0.0004$, $p<0.10$) portfolios (see appendix 5).

The market factor exposure is quite low, in line with the previous regressions. The SMB, HML and WML factor exposure is also comparable to the equal-weighted portfolios, although with generally less magnitude. This indicates that even when portfolios are constructed by lagged ESG scores, they tend to have similar attributes to theoretical constructed equal-weighted portfolios.

6.4 Fama French Five-Factor Plus Momentum on Equal-Weighted Portfolios (Subperiods)

The following regressions are based on the same portfolios as the regressions presented previously but are from the subperiods 2008-2014 and 2015-2020.

Nagy, Cogan, & Sinnreich (2013) suggest that abnormal returns will converge to 0 with time because abnormal returns may have occurred due to temporary mispricing of risk, and the market will learn to price it correctly with time. This phenomenon is referred to as the learning effect. By dividing the sample in pre/post 01.01.2015, we will be able to capture a possible learning effect.

Table 10: Fama French five-factor model plus momentum on equal-weighted portfolios (2008-2014)

	Equal-weighted portfolios from 2008 to the end of 2014		
	Dependent variable: R_i-R_f		
	ZI	BIC	NS
R_m-R_f	0.169^{***}	0.607^{***}	0.512^{***}
	p = 0.000	p = 0.000	p = 0.000
SMB	-0.490^{***}	-0.333^{***}	-0.027^{*}
	p = 0.000	p = 0.000	p = 0.069
HML	0.262^{*}	0.064	0.063^{***}
	p = 0.062	p = 0.356	p = 0.005
RMW	0.219	0.142	0.023
	p = 0.169	p = 0.114	p = 0.450
CMA	-0.768^{***}	-0.464^{***}	-0.081^{**}
	p = 0.00001	p = 0.00001	p = 0.014
WML	-0.042	-0.033	0.004
	p = 0.472	p = 0.343	p = 0.686
α	-0.001^{*}	-0.0004^{**}	-0.00004
	p = 0.078	p = 0.035	p = 0.497
Observations	1,758	1,758	1,758
R^2	0.257	0.807	0.966
Adjusted R^2	0.255	0.806	0.966
Note:			$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p<0.01$

Table 11: Fama French five-factor model plus momentum on equal-weighted portfolios (2015-2020)

	Equal-weighted portfolios from 2015 to the end of 2020		
	Dependent variable: R_i-R_f		
	ZI	BIC	NS
R_m-R_f	-0.016	0.453***	0.463***
	p = 0.469	p = 0.000	p = 0.000
SMB	-0.368***	-0.244***	-0.020
	p = 0.00001	p = 0.00000	p = 0.479
HML	0.207*	0.092	-0.064
	p = 0.082	p = 0.412	p = 0.311
RMW	0.469***	0.236*	-0.031
	p = 0.004	p = 0.053	p = 0.480
CMA	0.281**	0.105	0.011
	p = 0.036	p = 0.198	p = 0.776
WML	-0.192***	-0.129***	-0.070**
	p = 0.0005	p = 0.010	p = 0.022
α	-0.0001	0.00004	0.00001
	p = 0.652	p = 0.810	p = 0.894
Observations	1,506	1,506	1,506
R2	0.074	0.748	0.932
Adjusted R2	0.071	0.747	0.932
Note:			*p < 0.1, **p < 0.05, ***p<0.01

Table 10 and 11 present the results on Fama French Five-Factor model with momentum. The tables show risk factor loadings and daily abnormal returns (α) on three equal-weighted material ESG investing strategies constructed with ESG score in the periods 2008-2014 and 2015-2020. The regression has excess returns over the risk-free rate as dependent variable. R_m-R_f is the market factor, SMB is the size factor, HML is the value factor, RMW is the profitability factor, CMA is the investment factor and WML is the momentum factor. The α of the models represents any effects that ESG will have on the portfolio's abnormal returns.

The regression output shows significant negative abnormal returns for the ZI (α = -0.0001, p<0.10) and BIC (α = -0.0004, p<0.05) portfolios in the period 2008-2014. These abnormal returns are not significantly different from zero in the period 2015-2020. This is consistent with khan et al. (2016) and Nagy et al. (2013) which states that abnormal returns will converge towards zero as the market learns to price ESG correctly.

The portfolios are exposed to a greater market risk in the period of 2008-2014. Exposure to the SMB factor remain significantly negative throughout both periods for the ZI and the BIC portfolio, although the magnitude of the coefficient is smaller in the period 2015-2020.

The NS portfolio's exposure to the HML factor declines in significance in the 2015-2020 period, although the coefficient is initially quite low, with a value of 0.063 in the period 2008-2014.

The ZI and the BIC portfolios has insignificant exposure to the RMW factor in the first subperiod but had a positive exposure to the RMW factor in the last subperiod. This suggest that these portfolios became more profitable with time.

The CMA factor exposure is significantly negative for the ZI (β = -0.768, $p<0.01$) portfolio in the period 2008-2014. The coefficient remains significant at the 5% level in 2015-2020, although the coefficient is positive with a value of 0.281 in the period 2015-2020. The BIC has a significant coefficient value of -0.464 at the 1% level in the first subperiod which changes to a non-significant value of 0.105 in the last subperiod. This finding suggests that the companies with the highest ESG score invested more aggressively, and the companies with the lowest ESG score invested more conservatively in the first subperiod.

In the period of 2008-2014 there are no portfolios with a significant exposure to the WML factor. However, the ZI and BIC portfolios has a significant exposure at the 1% level with coefficients of -0.192 and -0.129 respectively in the period 2015-2020. The NS portfolio also has a significant exposure to the WML factor in the last subperiod, although the magnitude is small.

The ZI and BIC portfolios show the largest changes in the abnormal returns and the coefficient's magnitude and significance between the two subperiods.

6.5 Model Robustness

There are several assumptions for a multiple linear model to be robust. First, independent variables must have a linear relationship to all dependent variables. Furthermore, there must be no multicollinearity or heteroskedasticity. Lastly, the residuals must be independent and normally distributed. All the plots provided in this chapter are from equal and value-weighted ZI portfolios.

6.5.1 Linearity and Multicollinearity

We plot portfolio excess returns against independent variables and see that there is a linear relationship between the variables (See appendix 6). We thus do not need to transform any variables.

Multicollinearity is a statistical concept where one or more of the independent variables are correlated. We use the Variance Inflation Factor method (VIF) with the R-package usdm by Naimi (2017) to check the models for multicollinearity. The method indicates no multicollinearity[11].

6.5.2 Heteroscedasticity and Autocorrelation

Heteroscedasticity is a concept in timeseries data where the variance in the residuals is time dependent. If the residuals do not have a constant variance, the Standard error will be incorrect. This will in turn make the significance of the coefficients incorrect. We use the Breusch & Pagan (1979) method with the R-package lmtest by Zeileis & Hothorn (2002) to test for heteroskedasticity. The test indicates that we have heteroskedasticity in most value-weighted portfolios[12].

Figure 9: Residuals for zero-investment portfolios

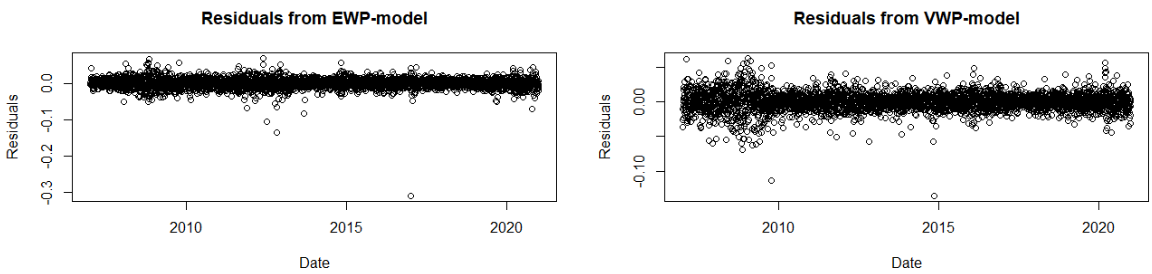


Figure 9: The plots show residuals on the y axis and time on the x axis. EWP and VWP are acronyms for equal-weighted portfolios and value-weighted portfolios respectively.

Figure 9 illustrates that the residuals in the value-weighted model have a higher variance in the years before 2010. The unexplained variance in the value-weighted portfolio was probably significantly higher these years because the portfolios was smaller combined with the fact that one of the companies in the long portfolio, EQNR, represented a very large part of the portfolio. In short, we think the value-weighted portfolios are substantially less diversified in the earlier years than in recent years which leads to firm specific fluctuations in the early

[11] We use a threshold of 10 in the VIF method when testing for multicollinearity. This is common practice and recommended by the author of the R-package we use (usdm). All the variables for all the models scored under the threshold.

[12] The value-weighted portfolios get p-values below the 5% threshold when conducting the bptest function from the lmtest package.

years. The market factor is not able to explain these fluctuations, thus the fluctuations is represented in the residuals.

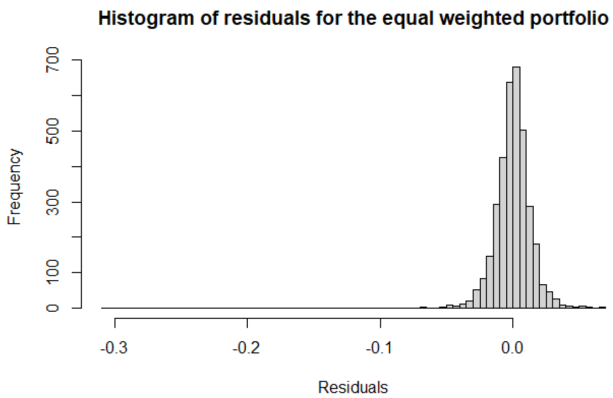
Autocorrelation is a concept in timeseries data where the residuals are dependent of each other. We use the Auto Correlation Function (ACF) with the R-package stat by Bolar (2019) to see if there is autocorrelation. The ACF Plots indicate autocorrelation with 1 lag for most models. (See appendix 7: ACF Plots).

We use Newey & West (1987) heteroscedasticity and autocorrelation consistent (HAC) covariance matrix estimators with the r-package sandwich by Zeileis & Lumley (2021) to estimate robust standard errors. The method is used for all models, and all standard errors, and thereby p-values reported in the result section are robust.

6.5.3 Multivariate Normality

Multivariate normality means that the residuals are normally distributed. We use quantile plots to check for multivariate normality. The visual check indicates some extreme values in returns more often than what a theoretical normal distribution does (See appendix 8).

Figure 10: Histogram of residuals for the equal-weighted zero-investment portfolio



We can see that the distribution of the residuals has a long tail on the left side which indicates extreme negative returns. We used the r-package robustHD by Alfons (2021) to winsorize 0.1% outliers on each side. Removing 1% and 0.5% from each tail turned out to be too much (See appendix 9: Outlier removal).

7.0 Discussion

In sum our results suggests that there is no positive relationship between material ESG performance and stock performance across different ESG investment strategies. This means that a firm's commitment to material ESG issues as stated in their annual reports is not rewarded with positive abnormal returns. However, the regression results of the full sample display that the equal-weighted ZI portfolio achieve -0.0004 daily abnormal returns significant at the 5% level, and the equal-weighted BIC portfolio achieve -0.0003 daily abnormal return significant at the 5% level, and that there are no abnormal returns for the value-weighted portfolios and no abnormal returns for the practical portfolios.

The portfolio returns regression of the period 2008-2014 shows that there is negative significant daily abnormal returns for the equal-weighted ZI portfolio ($\alpha = -0.001$, $p<0.1$) and the equal-weighted BIC portfolio ($\alpha = -0.0004$, $p<0.05$). The practical BIC portfolio also achieves daily abnormal returns of -0.0004 significant at the 10% level in the period 2008-2014. The regression of the subperiod 2015-2020 shows that there are no significant abnormal returns for any portfolio in our sample in this period. This suggests that the abnormal returns we find from our full sample originates from the period 2008-2014.

Negative screening is the investment strategy that has the least exposure to SMB and is the only one of the three investment strategies that never achieves significant abnormal returns. It is plausible that negative screening never has abnormal returns because the portfolio does not share the same characteristics as ESG portfolios. This can be explained by Nagy et al. (2013) who finds that portfolios created with an exclusion strategy do not achieve as high ESG score as other investment strategies.

The transition from negative to no abnormal returns in the subperiods of our sample can be explained by the fact that the market has gained access to material ESG metrics, which has led to the correct pricing of ESG shares. Khan et al. (2016) find abnormal returns in the sample period 1991-2012. They were the first to use ESG materiality in research and argue that abnormal returns arose because the concept of materiality had not been available to investors during this time. In the wake of this study, the concept of ESG materiality has received wider recognition and it is plausible that materiality has to a greater extent been implemented in investment strategies.

Nagy et al. (2013) argued that the market has mispriced the risk associated with ESG companies, which led to positive abnormal returns, and that the abnormal returns will diminish over time as markets learn to price ESG correctly. Our findings that abnormal returns have diminished recently for the equal-weighted ZI, BIC and the practical BIC portfolio can be explained by the fact that the learning effect has led to a correct pricing in 2015-2020.

In summary, we believe that the lack of ESG materiality measurements has led to a mispricing of risk associated with ESG companies, and that the implementation of materiality from 2015 has led to a learning effect that has removed abnormal returns.

7.1 Internal Validity

The ESG score we construct is valid. Based on economic intuition it makes sense that companies word frequency of ESG terms in their annual reports indicate ESG performance of the company. Loughran et al. (2009) find that companies which have a high word-count of the words «ethic», «corporate responsibility», «social responsibility» and «socially responsible» in 10-k filings more often is categorized as sin-stocks. Their findings may indicate that our score works against its purpose, and reward bad ESG performers. Nevertheless, we will argue against because in the validation chapter, we find a relatively high correlation between our score and Refinitiv's ESG score, especially on the companies that scored high. It is possible that the discoveries of Loughran et al. (2009) do not apply on the Oslo Stock Exchange.

Furthermore, the portfolios we construct based on our ESG score have the same characteristics as the ESG portfolios constructed in relevant literature. Our secondary research question was as follows,

Is it possible to measure a company's material ESG performance based on textual analysis of the company's annual report?

Based on the validity process, we argue that, yes, it is possible to measure a company's ESG performance based on textual analysis of the company's annual report.

However, our analysis suffers from minor sample bias. We only include stocks that were listed in 2022, which means that companies that have gone private or bankrupt are not represented in the sample. Furthermore, we have only included companies with English annual reports. Nevertheless, we argue that we have a small degree of sample bias compared to existing studies in the ESG field. The companies without English annual reports were in the minority, and we did not see any systematically exclusion of industries based on the availability of English annual reports except from the commercial bank industry. Existing studies, on the other hand, are based on score providers who have a limited ESG databases consisting of companies that have chosen to disclose their performance on ESG metrics. We find that the companies in our sample that were also in Refinitv's ESG database were systematically larger and received systematically higher ESG scores than the rest of our sample.

We tested the models for robustness and included control variables in regressions that capture factors that can explain returns. Overall, our analysis has good internal validity.

7.2 External Validity

Our results should not be generalized across time or markets without testing. We find heterogeneity in our models and argue that alpha changes over time due to the learning effect. The learning effect may occur at different times in different markets. Existing literature is contradictory and different markets may have different mechanisms. The findings on the Oslo Stock Exchange can not necessarily be generalized to other markets. Our findings are based on Norwegian annual reports. The design of annual reports in Norway are based on Norwegian norms and laws. Different norms and laws in other countries may lead to the design of annual reports being different, and that the presence of ESG words may have different patterns leading to different results.

Overall, we argue that our findings suffer from low external validity. This weakness is not mainly due to methodological choices, but because ESG studies generally have low external validity. The mechanisms within ESG seem to change a lot from market to market and from period to period.

7.3 Limitations and Suggested Further Research

We recognize that there may be a causality problem in this paper. Annual reports for 2012 were written in 2013, and returns from 2012 may have affected managers, which could be reflected in the wording of the annual report. In this case, abnormal returns could lead to change in ESG score. We tackle this problem by constructing a portfolio based on ESG scores constructed 2 years earlier. The results show that these portfolios have relatively similar results. Although the practical portfolios do not show negative abnormal returns for the entire period, the equal-weighted BIC portfolio ($\alpha = -0.0004$, $p < 0.10$) shows significant negative alpha in the subperiod 2008-2014. In addition, we see from the data section that the portfolios' companies overlap to a large extent from year to year, which indicates that the theoretically constructed portfolios are relatively similar to the practical portfolios.

For anyone who wants to replicate this method, we recommend improving the dictionary. The dictionary is based on words mentioned in SASB's industry reports. One can also expand the dictionary with words mentioned by other ESG sources such as the International Sustainability Standards Board (ISSB). We only got to review a fraction of the words manually and the way we went through the words was precise to the extent possible with limited resources. We recommend using domain-specific knowledge to manually go through the most important words in the dictionary for a more accurate score-giving.

In our dictionary, all ESG words have equal weighting on scores. Loughran & McDonald (2016) reported that "considering how the terms are weighted in the sentiment counts could improve the power of statistical tests attempting to identify sentiment patterns" We assume that this is also true for ESG words, and we believe that finding a systematic way to weight the different bigrams can contribute to a more precise score.

We recognize that an ESG score that is 100% based on textual analysis cannot be prevailing for how ESG should be measured for all regulators, institutions, and investors, as managers could easily manipulate the score. Our intention is not to replace the ESG score, but to provide a contributing factor to measure ESG. We recommend further research to consider several ESG factors and combine them with textual analysis to create an aggregate ESG score to get a more comprehensive picture of companies' ESG performance.

8.0 Conclusions

We investigate whether companies that perform well material ESG issues in the period 2008-2020 have overperformed in terms of abnormal returns. We use textual analysis with a self-constructed ESG dictionary to measure companies' material ESG performance. We perform the analysis under the hypothesis that,

Companies on the Oslo Stock Exchange which perform well on material ESG issues have positive abnormal returns.

To test the hypothesis, we create portfolios based on 3 different investment strategies; zero-investment portfolios that has a long position in the top quintile of ESG performers and a short position in the bottom quintile, best-in-class portfolios which consist of the top quintile of ESG performers, and negative screening portfolios that exclude the bottom quintile of ESG performers. We measure the excess return on these portfolios against the Fama French five-factor model plus momentum to see if the portfolios achieve abnormal returns that cannot be explained by the risk factors.

Our results suggests that significant negative abnormal returns are present in the period 2008-2014, and that these returns converged to zero in the period 2015-2020. We therefore argue that investors have mispriced risk associated with ESG companies on the Oslo Stock Exchange, and thus mispriced these shares in the period 2008 to 2014. Furthermore, we argue that the availability of material ESG data in recent years has led to a learning effect taking place. This implies that ESG companies are priced correctly and that abnormal returns are not present in recent times. The main hypothesis is rejected as we find no evidence of a positive relationship between ESG material performance and stock performance.

References

Alfons, A. (2021, 11 23). *robustHD: Robust Methods for High-Dimensional Data*. Retrieved from r-project: https://CRAN.R-project.org/package=robustHD

Amel-Zadeh, A., & Serafeim, G. (2018). Why and How Investors Use ESG Information: Evidence from a Global Survey. *Financial Analysts Journal, 74(3)*, 87-103.

Baier, P., & Florian, K. (2020). Environmental, social and governance reporting in annual reports: A textual analysis. *Financial Markets, Institutions & Instruments, 29(3)*, 91-118.

Berg, F., Kölbel, J. F., & Rigobon, R. (2019). Aggregate Confusion: The Divergence of ESG Ratings. *Forthcoming Review of Finance*.

Bolar, K. (2019, 04 01). *STAT: Interactive Document for Working with Basic Statistical Analysis*. Retrieved from r-project: https://CRAN.R-project.org/package=STAT

Breusch, T., & Pagan, A. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica, 47*, 1287-1294.

Carhart, M. M. (2012). On persistence in Mutual Fund Performance. *The Journal of The American Finance Association, 52(1)*, 57-82.

Danielsen, T. A., & Johansen, J. (2021). *Value relevant ESG scores and superior performance : when is a high ESG score indicative of better financial performance?* NHH Brage Open institutional repository.

Davis, A. K., Piger, J. M., & Sedor, L. M. (2012). Beyond the Numbers: Measuring the Information Content of Earning Press Release Language. *Journal of Business Finance & Accounting, 29(3)*, 845-868.

DFØ. (2022, 01 06). *Årsrapport*. Retrieved from Direktoratet for forvaltning og økonomistyring: https://dfo.no/fagomrader/arsrapport

Dorfleitner, G., Halbritter, G., & Nguyen, M. (2015). Measuring the level and risk of corporate responsibility : an empirical comparison of different ESG rating approaches. *The journal of asset management, 16(7)*, 450-466.

Doyle, T. (2018). *Ratings that Don't Rate: The Subjective World of ESG Ratings Agencies.* American Council for Capital Formation.

Drempetic, S., Klein, C., & Zwergel, B. (2020). The Influence of Firm Size on the ESG Score: Corporate Sustainability Ratings Under Review. *Journal of Business Ethics, 167(1)*, 333-360.

European Union. (2020). REGULATION (EU) 2020/852 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. *Official Journal of the European Union, 63*.

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.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics, 116(1)*, 1-22.

Feinerer, I., & Hornik, K. (2020, 11 18). *Text Mining Package*. Retrieved from Cran r-projects: https://cran.r-project.org/web/packages/tm/tm.pdf

French, K. R. (2022). *Data library*. Retrieved from Kenneth R. French: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

French, K. R. (2022, 04). *Description of Fama/French 5 Factors for Developed Markets*. Retrieved from Kenneth R French Data Library.

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.

Friedman, M. (1970). The Social Responsibility of Business Is to Increase Its Profits. *New York Times Magazine*, 122-126.

García, D. (2013). Sentiment during recessions. *The Journal of The American Finance Association, 68(3)*, 1267-1300.

Gibson, R., Krueger, P., & Schmidt, P. S. (2019). ESG Rating Disagreement and Stock Returns. *Financial Analyst Journal*, 19-67.

Giese, G., Lee, L.-E., Melas, D., & Nishikawa, L. (2019). Foundations of ESG Investing: How ESG Affects Equity Valuation, Risk, and Performance. *The Journal of Portfolio Management, 45(5)*, 7-28.

Halbritter, G., & Dorfleitner, G. (2015). The wages of social responsibility — where are they? A critical review of ESG investing. *Review of Financial Economics, 26(1)*, 25-35.

Henry, E. (2008). Are Investors Influenced by how Earnings Press Releases are Written? *Journal of Business Communication, 45*, 363-407.

Kempf, A., & Osthoff, P. (2007). The Effect of Socially Responsible Investing on Portfolio Performance. *European Financial Management, 13(5)*, 908-922.

Khan, M., Serafeim, G., & Yoon, A. (2016). Corporate Sustainability: First Evidence on Materiality. *The Accounting Review, 91(6)*, 1697-1724.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics, 45(2-3)*, 221-247.

Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics, 47(1)*, 13-37.

Loughran, T., & McDonald, B. (2011). When is a Liability not a Liability? Textual analysis, Dictionaries, and 10-Ks. *Journal of Finance, 66*, 35-65.

Loughran, T., & McDonald, B. (2016). Textual analysis in Accounting and Finance: A Survey. *Journal of Accounting Research, 54(4)*, 1187-1230.

Loughran, T., McDonald, B., & Yun, H. (2009). A Wolf in Sheep's Clothing: The Use of Ethics-Related Terms in 10-K Reports. *Journal of Business Ethics, 89*, 39-49.

Mayew, W. J., & Venkatachalam, M. (2012). The Power of Voice: Managerial Affective States and Future Firm Performance. *The Journal of The American Finance Association, 67(1)*, 1-43.

Nagy, Z., Cogan, D., & Sinnreich, D. (2013). *Optimizing Environmental, Social and Governance Factors in Portfolio Construction: Analysis of Three ESG-Tilted Strategies.* SRPN: Other Corporate Social Responsibility (CSR) (Topic).

Naimi, B. (2017, 06 25). *Variance Inflation Factor and test for multicollinearity*. Retrieved from r-project: https://search.r-project.org/CRAN/refmans/usdm/html/vif.html

Newey, W., & West, K. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica, 55*, 703-708.

Price, S. M., Doran, J., Peterson, D. R., & Bliss, B. A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance, 36(4)*, 992-1011.

Refinitiv. (2022, 05). *Environmental, Social And Governance Scores from Refinitiv.* Retrieved from Refinitiv: https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf

Sargis, M., & Wang, P. (2020, 02 19). *How Does Investing in ESG Companies Affect Returns?* Retrieved from Morningstar: https://www.morningstar.com/insights/2020/02/19/esg-companies

SASB. (2022). *About us*. Retrieved from Value Reporting Foundation SASB Standards: https://www.sasb.org/about/

Sharpe, W. F. (1964). CAPITAL ASSET PRICES: A THEORY OF MARKET EQUILIBRIUM UNDER CONDITIONS OF RISK. *The Journal of Finance, 19(3)*, 425-442.

Statman, M., & Glushkov, D. (2008). The Wages of Social Responsibility. *Financial Analysts Journal, 65(4)*, 33-46.

Sustainability Accounting Standards Board. (2020). *MEAT, POULTRY & DAIRY.* 10.

Who Cares Wins. (2005). *Who Cares Wins 2005 Conference Report: Investing for Long-Term Value.* International Finance Corporation (IFC).

Wickham, H. (2022). *rvest: Easily Harvest (Scrape) Web Pages*. Retrieved from Tidyverse: https://github.com/tidyverse/rvest

Yahoo. (2022). *What is adjusted close?* Retrieved from Yahoo: https://help.yahoo.com/kb/SLN28256.html

Zeileis, A., & Hothorn, T. (2002). Diagnostic Checking in Regression Relationships. *R News, 2(3)*, 7-10.

Zeileis, A., & Lumley, T. (2021, 05 18). *sandwich: Robust Covariance Matrix Estimators*. Retrieved from r-project: https://CRAN.R-project.org/package=sandwich

Appendix

Appendix 1: Positive ESG Bigrams From our Dictionary

'corpor_govern' 'risk_manag' 'law_regul' 'health_safeti' 'intern_control' 'manag_system' 'oil_gas' 'report_standard' 'sustain_report' 'manag_risk' 'climat_chang' 'develop_product' 'research_develop' 'risk_profil' 'capit_adequaci' 'risk_relat' 'energi_consumpt' 'greenhous_gas' 'explor_product' 'activ_includ' 'servic_provid' 'energi_effici' 'product_facil' 'suppli_chain' 'environment_social' 'financi_condit' 'social_secur' 'compani_manag' 'renew_energi' 'environment_impact' 'explor_develop' 'gas_emiss' 'rule_procedur' 'code_conduct' 'reduc_risk' 'local_communiti' 'natur_gas' 'sustain_develop' 'activ_invest' 'regulatori_requir' 'probabl_default' 'risk_capit' 'risk_assess' 'risk_includ' 'fuel_consumpt' 'gas_reserv' 'european_union' 'financi_risk' 'oil_product' 'stress_test' 'ghg_emiss' 'crude_oil' 'materi_impact' 'oil_natur' 'credit_risk' 'gas_product' 'food_safeti' 'safeti_manag' 'risk_exposur' 'energi_sourc' 'ballast_water' 'impact_oper' 'reduc_emiss' 'product_process' 'signific_impact' 'credit_exposur' 'gas_explor' 'human_right' 'risk_opportun' 'carbon_dioxid' 'product_oil' 'explor_activ' 'money_launder' 'ism_code' 'environment_protect' 'global_compact' 'prove_reserv' 'food_product' 'busi_ethic' 'fuel_oil' 'seismic_survey' 'custom_supplier' 'natur_resourc' 'clinic_trial' 'metric_ton' 'classif_societi' 'intern_convent' 'live_weight' 'advers_impact' 'life_insur' 'power_product' 'fresh_water' 'person_data' 'competit_advantag' 'health_insur' 'includ_employe' 'inform_secur' 'classif_system' 'estim_amount' 'materi_product' 'cubic_meter' 'develop_reserv' 'govern_structur' 'improv_product' 'natur_peril' 'wast_manag' 'environment_regul' 'water_treatment' 'human_resourc' 'petroleum_product' 'gas_produc' 'hazard_substanc' 'legal_requir' 'load_factor' 'product_tanker' 'code_ethic' 'direct_emiss' 'energi_product' 'carbon_captur' 'conting_liabil' 'fuel_effici' 'negat_affect' 'respons_audit' 'water_consumpt' 'anim_welfar' 'emiss_air' 'emiss_trade' 'unit_nation' 'cubic_feet' 'fossil_fuel' 'intellectu_properti' 'legal_regulatori' 'busi_practic' 'develop_technolog' 'emiss_reduct' 'materi_busi' 'bunker_fuel' 'civil_aviat' 'coast_guard' 'develop_produc' 'hazard_wast' 'intern_standard' 'oil_condens' 'passeng_kilomet' 'product_qualiti' 'scope_emiss' 'action_plan' 'air_emiss' 'air_travel' 'board_composit' 'captur_storag' 'correct_action' 'host_govern' 'imo_intern' 'nation_intern' 'social_govern' 'custom_relationship' 'pollut_damag' 'regulatori_framework' 'seafood_product' 'commiss_sec' 'insur_polici' 'insur_portfolio' 'qualiti_product' 'respons_invest' 'solar_energi' 'standard_cubic' 'air_transport' 'attract_retain' 'board_iasb' 'compani_custom' 'control_influenc' 'data_protect' 'gas_process' 'includ_board' 'oil_spill' 'reserv_prove' 'seat_kilomet' 'technolog_develop' 'air_pollut' 'discharg_oil' 'financi_crime' 'financi_crisi' 'harsh_environ' 'health_care' 'research_council' 'safeti_perform' 'stewardship_council' 'total_emiss' 'activ_manag' 'cloud_servic' 'complianc_requir' 'infrastructur_project' 'pharmaceut_product' 'rule_regul' 'system_risk' 'total_energi' 'water_inject' 'activ_ownership' 'compani_requir' 'data_center' 'electron_manufactur' 'intang_asset' 'manag_polici' 'manufactur_process' 'regulatori_approv' 'relev_activ' 'servic_custom' 'signific_influenc' 'softwar_servic' 'sustain_invest' 'approv_product' 'cede_reinsur' 'electr_generat' 'emiss_control' 'fix_salari' 'govern_manag' 'medic_treatment' 'meet_oblig' 'meet_requir' 'nautic_mile' 'requir_law' 'result_materi' 'right_violat' 'safeti_secur' 'sale_licens' 'sensit_analysi' 'term_employ' 'workers_compens' 'benefit_expens' 'communiti_impact' 'dioxid_emiss' 'dispos_cost' 'emiss_scope' 'employe_compani' 'employe_engag' 'fish_harvest' 'increas_competit' 'integr_report' 'kyoto_protocol' 'nitrogen_oxid' 'potenti_impact' 'record_incid' 'requir_applic' 'return_custom' 'sustain_issu' 'anticorrupt_polici' 'bauxit_mine' 'corrupt_practic' 'custom_includ' 'data_servic' 'drill_activ' 'electr_consumpt' 'emerg_prepared' 'facil_includ' 'feed_raw' 'imo_adopt' 'impact_assess' 'meet_criteria' 'power_generat' 'product_innov' 'social_impact' 'solar_industri' 'solar_modul' 'solar_power' 'anim_protein' 'clinic_research' 'comput_softwar' 'drug_administr' 'emiss_emiss' 'ensur_complianc' 'environment_law' 'food_ingredi' 'hydraul_fractur' 'injuri_rate' 'intern_energi' 'issu_effect' 'licens_oper' 'manag_materi' 'materi_sustain' 'occup_health' 'oil_sand' 'packag_materi' 'purchas_electr' 'water_discharg' 'bank_requir' 'carbon_emiss' 'cloud_comput' 'council_mine' 'custom_data' 'data_secur' 'emerg_respons' 'emiss_electr' 'employe_train' 'energi_hydro' 'equal_rate' 'food_drug' 'global_warm' 'impact_company' 'intern_council' 'pollut_ship' 'record_injuri' 'safeti_risk' 'treatment_system' 'agricultur_product' 'air_qualiti' 'benefit_paid' 'busi_council' 'clean_energi' 'compani_involv' 'contract_research' 'custom_insur' 'data_access' 'data_privaci' 'develop_program' 'energi_intens' 'energi_suppli' 'foreign_corrupt' 'freedom_associ' 'grievanc_mechan' 'life_cycl' 'mainten_aircraft' 'pollut_prevent' 'process_safeti' 'safeti_standard' 'stakehold_engag' 'averag_age' 'employe_contribut' 'energi_recoveri' 'ensur_custom' 'ensur_qualiti' 'govern_regul' 'impact_financi' 'locat_countri' 'materi_process' 'metal_icmm' 'missil_system' 'polici_procedur' 'power_purchas' 'protein_product' 'respect_human' 'scope_ghg' 'share_inform' 'solar_pv' 'total_water' 'treatment_plant' 'vertic_integr' 'basel_committe' 'complianc_risk' 'cost_solar' 'develop_insur' 'develop_market' 'discharg_water' 'due_fraud' 'electr_power' 'emiss_water' 'employe_manag' 'employe_note' 'employe_represent' 'energi_generat' 'environment_damag' 'extrem_weather' 'feed_purchas' 'financi_impact' 'fund_sourc' 'ghg_protocol' 'global_aquacultur' 'human_capit' 'inform_includ' 'inform_technolog' 'infrastructur_servic' 'monitor_program' 'popul_growth' 'produc_water' 'product_solut' 'product_sustain' 'product_weight' 'pv_system' 'receiv_govern' 'technolog_custom' 'technolog_product' 'vacanc_rate' 'wast_dispos' 'workers_right' 'workrel_injuri' 'accid_incid' 'activ_particip' 'activ_relat' 'administr_fda' 'amount_wast' 'ban_import' 'bank_sustain' 'black_sea' 'casualti_insur' 'chain_manag' 'communiti_oper' 'compani_engag' 'competit_solar' 'direct_greenhous' 'electr_vehicl' 'employe_categori' 'employe_defin' 'employe_locat' 'energi_save' 'environment_footprint' 'expect_impact' 'feed_price' 'feedin_tariff' 'food_industri' 'forc_labor' 'global_economi' 'government_regul' 'greenhous_gase' 'guid_principl' 'hydrocarbon_reserv' 'inform_compani' 'inform_sustain' 'life_sea' 'low_oil' 'manag_practic' 'market_develop' 'million_vehicl' 'oil_consumpt' 'oper_sustain' 'palm_oil' 'passeng_load' 'person_injuri' 'prevent_pollut' 'product_safeti' 'project_respons' 'pv_solar' 'reserv_oil' 'respons_sustain' 'safeti_cultur' 'satur_fat' 'social_capit' 'solar_panel' 'standard_disclosur' 'weapon_system' 'base_water' 'bilg_water' 'briberi_corrupt' 'busi_human' 'carbon_footprint' 'ceo_respons' 'compani_disclos' 'compani_disclosur' 'compens_employe' 'complianc_code' 'compris_compani' 'condit_ship' 'consum_health' 'data_breach' 'demand_food' 'develop_solar' 'direct_ec' 'effici_product' 'emiss_energi' 'emiss_includ' 'energi_market' 'environ_sustain' 'equat_principl' 'esg_factor' 'ethic_corpor' 'eu_et' 'financi_literaci' 'freedom_express' 'gas_chain' 'improv_qualiti' 'incid_rate' 'includ_energi' 'industri_transpar' 'inform_account' 'inform_inform' 'inform_requir' 'integr_product' 'intern_regul' 'manag_employe' 'manag_sustain' 'mine_process' 'obtain_regulatori' 'oil_andor' 'oper_govern' 'polici_note' 'process_packag' 'product_crude' 'product_elimin' 'properti_casualti' 'recycl_materi' 'reduc_environment' 'reduc_product' 'report_environment' 'report_integr' 'resourc_scarciti' 'risk_polici' 'shale_oil' 'sourc_sustain' 'sustain_inform' 'sustain_risk' 'access_health' 'activ_ingredi' 'agreement_ppa' 'altern_fuel' 'amount_energi' 'audit_standard' 'bank_contribut' 'chemic_product' 'chemic_specialti' 'climat_risk' 'code_ism' 'communic_corpor' 'compact_principl' 'compani_contribut' 'compani_intern' 'competit_posit' 'complianc_regul' 'condit_employe' 'control_procedur' 'corrupt_briberi' 'council_sustain' 'develop_sustain' 'electr_engin' 'electr_heat' 'electr_purchas' 'electr_retail' 'emiss_reduc' 'employe_includ' 'employe_turnov' 'energi_consum' 'energi_insur' 'energi_produc' 'energi_technolog' 'environment_challeng' 'european_commiss' 'gas_drill' 'gas_ghg' 'geograph_exposur' 'global_market' 'green_energi' 'impact_environ' 'impact_gain' 'indian_ocean' 'industri_develop' 'invest_sustain' 'key_personnel' 'kilomet_rpk' 'local_market' 'local_stakehold' 'manag_environment' 'methan_emiss' 'nutrit_health' 'oper_safeti' 'particul_matter' 'power_industri' 'power_plant' 'probabl_oil' 'product_sourc' 'product_stewardship' 'promot_healthi' 'provid_inform' 'recycl_packag' 'regulatori_complianc' 'regulatori_risk' 'requir_report' 'resourc_effici' 'review_committe' 'risk_intern' 'sell_electr' 'social_environment' 'spill_respons' 'stage_life' 'transport_chemic' 'wast_generat' 'water_qualiti' 'water_withdraw' 'wind_solar' 'account_metric' 'achiev_regulatori' 'activ_pharmaceut' 'adopt_practic' 'amount_exposur' 'anim_feed' 'antifoul_system' 'chemic_ingredi' 'child_labor' 'combin_heat' 'company_environment' 'complianc_polici' 'conduct_clinic' 'consumpt_capita' 'contribut_sustain' 'corpor_activ' 'custom_power' 'custom_sustain' 'data_collect' 'data_consolid' 'data_traffic' 'daytoday_oper' 'decommiss_project' 'drill_oil' 'econom_growth' 'econom_right' 'electr_deliveri' 'electr_grid' 'emiss_increas' 'emiss_limit' 'emiss_result' 'employe_accord' 'energi_cost' 'energi_includ' 'energi_renew' 'ensur_equal' 'environment_manag' 'environment_risk' 'eu_direct' 'fractur_fluid' 'gas_develop' 'global_food' 'global_standard' 'health_nutrit' 'heat_cool' 'human_health' 'impact_product' 'incid_report' 'includ_contribut' 'includ_custom' 'includ_disclosur' 'industri_foreign' 'inform_investor' 'intern_labour' 'intern_safeti' 'kilowatt_hour' 'legisl_regul' 'local_author' 'marin_casualti' 'maritim_organ' 'oecd_guidelin' 'packag_product' 'packag_recycl' 'patient_treat' 'photovolta_pv' 'physic_damag' 'physic_risk' 'power_produc' 'process_water' 'produc_reserv' 'product_human' 'product_impact' 'product_materi' 'public_health' 'reduc_fuel' 'reduc_water' 'refer_sustain' 'regul_includ' 'regul_relat' 'regul_sx' 'relev_employe' 'renew_materi' 'renew_power' 'requir_supplier' 'scrap_metal' 'social_licens' 'sustain_fish' 'sustain_perform' 'water_risk' 'abandon_ship' 'activ_marin' 'agreement_parti' 'agricultur_organ' 'air_water' 'aircraft_compon' 'altern_energi' 'chemic_safeti' 'chicken_pork' 'climat_disclosur' 'climat_impact' 'climaterel_financi' 'climaterel_risk' 'condit_sustain' 'consumpt_scope' 'contamin_pathogen' 'cool_steam' 'critic_infrastructur' 'custom_electr' 'custom_privaci' 'data_storag' 'depart_justic' 'determin_materi' 'diesel_engin' 'disclosur_tcfd' 'electron_wast' 'emiss_primari' 'emiss_wast' 'employe_activ' 'employe_report' 'endang_speci' 'energi_agenc' 'energi_manag' 'energi_mix' 'energi_total' 'environ_social' 'environment_measur' 'environment_respons' 'eu_emiss' 'farm_system' 'feed_sourc' 'femal_male' 'fertil_crop' 'financi_sustain' 'food_beverag' 'food_consum' 'food_label' 'govern_esg' 'govern_offici' 'govern_process' 'green_build' 'guidelin_standard' 'heat_power' 'heavi_fuel' 'imo_guidelin' 'impact_climat' 'increas_consumpt' 'increas_fuel' 'industri_gase' 'injuri_ill' 'integr_solar' 'intern_air' 'key_custom' 'manufactur_food' 'manufactur_solar' 'marin_environ' 'metric_unit' 'middl_manag' 'natur_capit' 'offer_employe' 'oper_accord' 'organ_imo' 'oxid_emiss' 'paper_product' 'pari_agreement' 'particip_clinic' 'peopl_communiti' 'percentag_employe' 'pet_food' 'pharmaceut_ingredi' 'power_deliv' 'power_market' 'prepar_sustain' 'privaci_data' 'privaci_user' 'produc_renew' 'product_food' 'product_packag' 'protect_privaci' 'pulp_paper' 'pv_modul' 'qualiti_safeti' 'rate_trir' 'recoveri_rate' 'recycl_industri' 'recycl_renew' 'reduc_energi' 'relat_legal' 'renew_electr' 'risk_climat' 'safeti_incid' 'safeti_initi' 'scenario_analysi' 'secur_risk' 'semiconductor_materi' 'slaughter_process' 'social_sustain' 'standard_requir' 'sulfur_acid' 'surfac_water' 'sustain_manag' 'sustain_palm' 'technolog_innov' 'technolog_solut' 'temporari_employe' 'train_employe' 'transpar_initi' 'transpar_intern' 'wast_recycl' 'wast_stream' 'water_manag' 'water_scarciti' 'water_shortag' 'water_stress' 'weather_event' 'wind_energi' 'wood_fibr' 'world_econom' 'access_energi' 'accid_fatal' 'accid_investig' 'accord_imo' 'acquir_dispos' 'activ_develop' 'activ_engag' 'address_children' 'affect_custom' 'agreement_note' 'agreement_ppas' 'aircraft_fuel' 'altern_sustain' 'amount_emiss' 'amount_materi' 'amount_recov' 'anim_health' 'anticorrupt_antibriberi' 'applic_law' 'approv_local' 'arctic_circl' 'assess_environment' 'awar_program' 'board_complianc' 'board_fasb' 'board_involv' 'board_sasb' 'briberi_act' 'build_materi' 'busi_environ' 'carbon_tax' 'care_profession' 'care_provid' 'catastroph_event' 'catastroph_reinsur' 'center_medicar' 'certifi_global' 'certifi_wood' 'chain_custodi' 'chain_sustain' 'chemic_petroleum' 'clean_air' 'clean_water' 'clinic_laboratori' 'closur_decommiss' 'co_e' 'communiti_action' 'communiti_develop' 'communiti_local' 'communiti_stakehold' 'compani_environment' 'complet_solar' 'complianc_intern' 'condens_natur' 'consum_electron' 'consumpt_energi' 'consumpt_wast' 'contract_employe' 'contribut_global' 'countri_risk' 'custom_inform' 'cycl_includ' 'defens_system' 'develop_hydrocarbon' 'develop_local' 'develop_wind' 'dioxid_nitrogen' 'direct_employe' 'direct_scope' 'discharg_limit' 'disclos_gender' 'disclosur_requir' 'diseas_develop' 'dispos_compani' 'divers_talent' 'divid_facil' 'domin_market' 'drill_apprais' 'drive_growth' 'drug_applic' 'drug_product' 'due_curtail' 'editori_content' 'educ_materi' 'effici_reduc' 'electr_aircraft' 'electr_bill' 'electr_consum' 'emerg_market' 'emiss_greenhous' 'emiss_target' 'emiss_transport' 'employ_compani' 'employe_classifi' 'employe_contract' 'employe_health' 'employe_subcontractor' 'encourag_report' 'energi_data' 'energi_energi' 'energi_green' 'energi_materi' 'energi_project' 'energi_report' 'energi_resourc' 'energi_revenu' 'energi_storag' 'energi_water' 'engag_employe' 'engag_local' 'ensur_health' 'ensur_privaci' 'ensur_safeti' 'environ_includ' 'environment_consider' 'environment_effect' 'environment_effici' 'environment_inform' 'environment_polici' 'environment_principl' 'environment_qualiti' 'environment_standard' 'environment_sustain' 'equival_coe' 'esg_environment' 'estat_chain' 'estat_compani' 'estat_invest' 'estat_project' 'ethic_complianc' 'ethic_conduct' 'european_aviat' 'exempt_requir' 'exposur_industri' 'exposur_limit' 'fair_trade' 'farm_requir' 'feder_trade' 'focus_grow' 'food_wast' 'forest_manag' 'fuel_expens' 'fuel_fuel' 'fuel_usag' 'gas_compress' 'gas_facil' 'generat_energi' 'geograph_region' 'ghg_report' 'global_energi' 'govern_commerci' 'govern_entiti' 'govern_market' 'green_power' 'green_technolog' 'grid_electr' 'guarante_origin' 'guidanc_note' 'guidelin_compani' 'guidelin_report' 'handl_hazard' 'hazard_materi' 'health_claim' 'health_risk' 'healthcar_product' 'human_consumpt' 'hydraul_fluid' 'hydrocarbon_resourc' 'identifi_mitig' 'ii_noxious' 'impact_account' 'impact_aris' 'impact_compani' 'impact_local' 'impact_scope' 'implement_ballast' 'improv_energi' 'incid_incid' 'includ_climat' 'includ_cloud' 'includ_emiss' 'includ_environment' 'includ_land' 'includ_power' 'includ_scope' 'increas_recycl' 'independ_ep' 'industri_associ' 'industri_classif' 'industri_custom' 'industri_protect' 'industri_report' 'inform_link' 'ingredi_compani' 'ingredi_product' 'initi_ensur' 'initi_gfsi' 'initi_research' 'insul_materi' 'intergovernment_panel' 'intern_report' 'invest_report' 'labor_child' 'land_sea' 'lca_analysi' 'limit_environment' 'liquid_substanc' 'local_regul' 'lower_emiss' 'lower_energi' 'maintain_level' 'manag_forest' 'marin_protect' 'market_renew' 'materi_includ' 'materi_inform' 'materi_materi' 'medic_equip' 'medicaid_servic' 'medicar_medicaid' 'mental_health' 'mitig_impact' 'municip_water' 'nation_global' 'negat_media' 'newsprint_paper' 'occup_rate' 'offshor_wind' 'oil_heavi' 'onlin_privaci' 'oper_impact' 'origin_equip' 'ownership_structur' 'paper_industri' 'partner_employe' 'patent_licens' 'pay_electr' 'permit_standard' 'person_protect' 'physic_activ' 'plastic_rubber' 'pork_beef' 'power_compani' 'power_deliveri' 'prevent_corrupt' 'price_volatil' 'privaci_secur' 'procedur_assess' 'proceed_briberi' 'process_chemic' 'process_control' 'process_emiss' 'process_technolog' 'process_wast' 'product_health' 'product_improv' 'product_ingredi' 'product_reduc' 'project_life' 'project_lifecycl' 'properti_damag' 'protect_agenc' 'protect_equip' 'protect_protect' 'protect_regul' 'protocol_corpor' 'provid_custom' 'purchas_power' 'purchas_renew' 'rainforest_allianc' 'rate_employe' 'recycl_technolog' 'red_list' 'reduc_raw' 'reduc_revenu' 'reduc_wast' 'regul_drive' 'regulatori_environ' 'relat_clinic' 'relat_compens' 'relat_data' 'relat_electr' 'relat_emiss' 'relat_govern' 'relat_natur' 'relev_busi' 'remuner_polici' 'renew_fuel' 'report_disclosur' 'report_ghg' 'report_legal' 'report_suspect' 'reput_risk' 'requir_approv' 'requir_estim' 'requir_relat' 'reserv_determin' 'reserv_polici' 'respons_care' 'respons_climat' 'review_board' 'right_issu' 'risk_briberi' 'safeguard_health' 'safeti_concern' 'safeti_data' 'safeti_issu' 'secur_manag' 'secur_provis' 'seek_reduc' 'segment_electr' 'serv_custom' 'servic_saa' 'set_sustain' 'share_data' 'social_risk' 'solar_hydro' 'solar_technolog' 'solar_wind' 'standard_qualiti' 'standard_sustain' 'strategi_sustain' 'sulfur_oxid' 'suppli_seafood' 'suspect_violat' 'sustain_account' 'sustain_disclosur' 'sustain_esg' 'sustain_sourc' 'system_materi' 'technic_skill' 'technolog_improv' 'term_plan' 'term_sustain' 'total_electr' 'total_ghg' 'transport_cost' 'transport_road' 'trial_includ' 'turnov_rate' 'uk_briberi' 'uniqu_posit' 'uniqu_user' 'utz_certifi' 'wage_salari' 'warm_potenti' 'wast_wast' 'water_contamin' 'water_exchang' 'water_suppli' 'water_water' 'welfar_standard' 'wind_farm' 'wind_power' 'wind_project' 'worker_right' 'worklif_balanc' 'world_health'

Appendix 2: SASB Materiality Map Example

Figure 11: SASB materiality map example

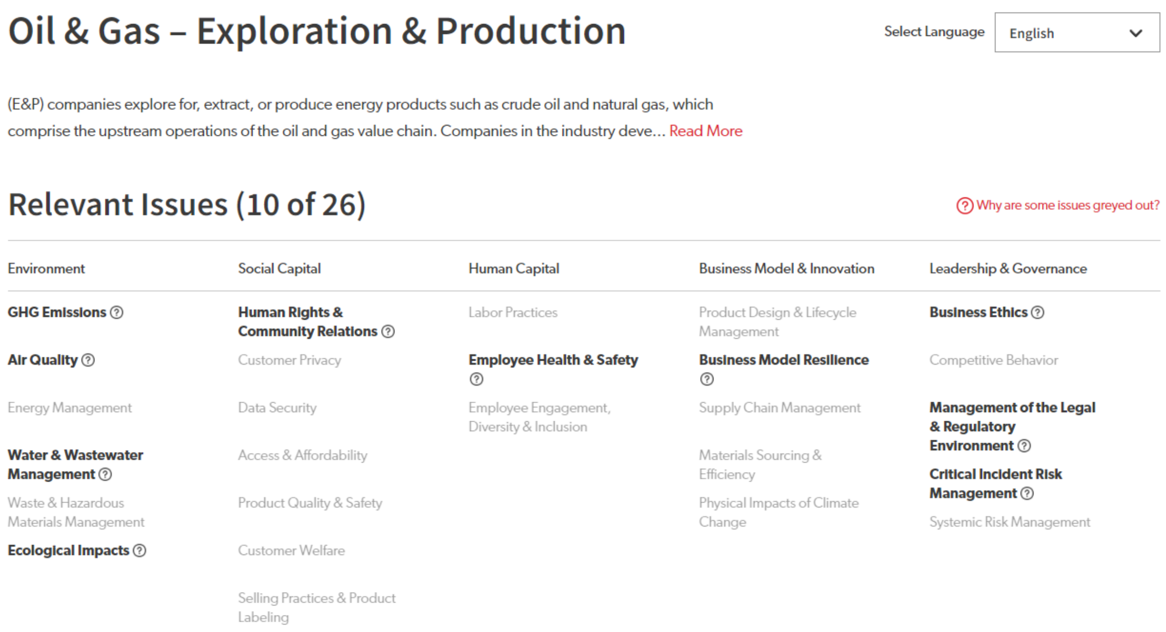


Figure 11 illustrates the pillars that SASB prioritize when considering materiality for a given industry. The black text indicate that these factors are material for the exploration and production industry.

Appendix 3: Norwegian Index Regressed on European Market

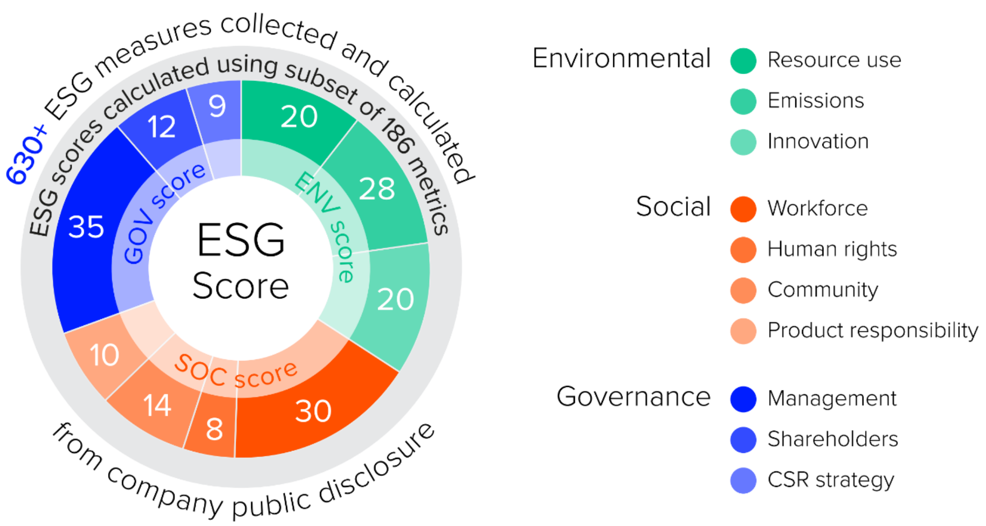
Table 12: Norwegian index regressed on European market

	Dependent variable: *Norwegian Market – R_f*
European market – R_f	1.300^{***}
	p = 0.000
SMB	-0.026
	p = 0.834
HML	0.847^{***}
	p = 0.000
RMW	0.750^{***}
	p = 0.003
CMA	-1.313^{***}
	p = 0.00003
WML	-0.135^{*}
	p = 0.065
α	0.001^{**}
	p = 0.043
Note:	$^{*}p<0.1$, $^{**}p<0.05$, $^{***}p<0.01$

The table show that the Norwegian index constructed of our sample has significant alpha vs the European index. This is the reason why the market factor in the Fama French five-factor model plus momentum's market factor had to be changed. Most portfolios constructed of Norwegian companies, regardless of the ESG performance would achieve abnormal returns. P-values use robust standard errors.

Appendix 4: Refinitiv Score

Figure 12: Refinitiv score



Appendix 5: Regression Output for Practical Portfolios (2008-2014)

Table 13: Regression output for practically portfolios (2008-2014)

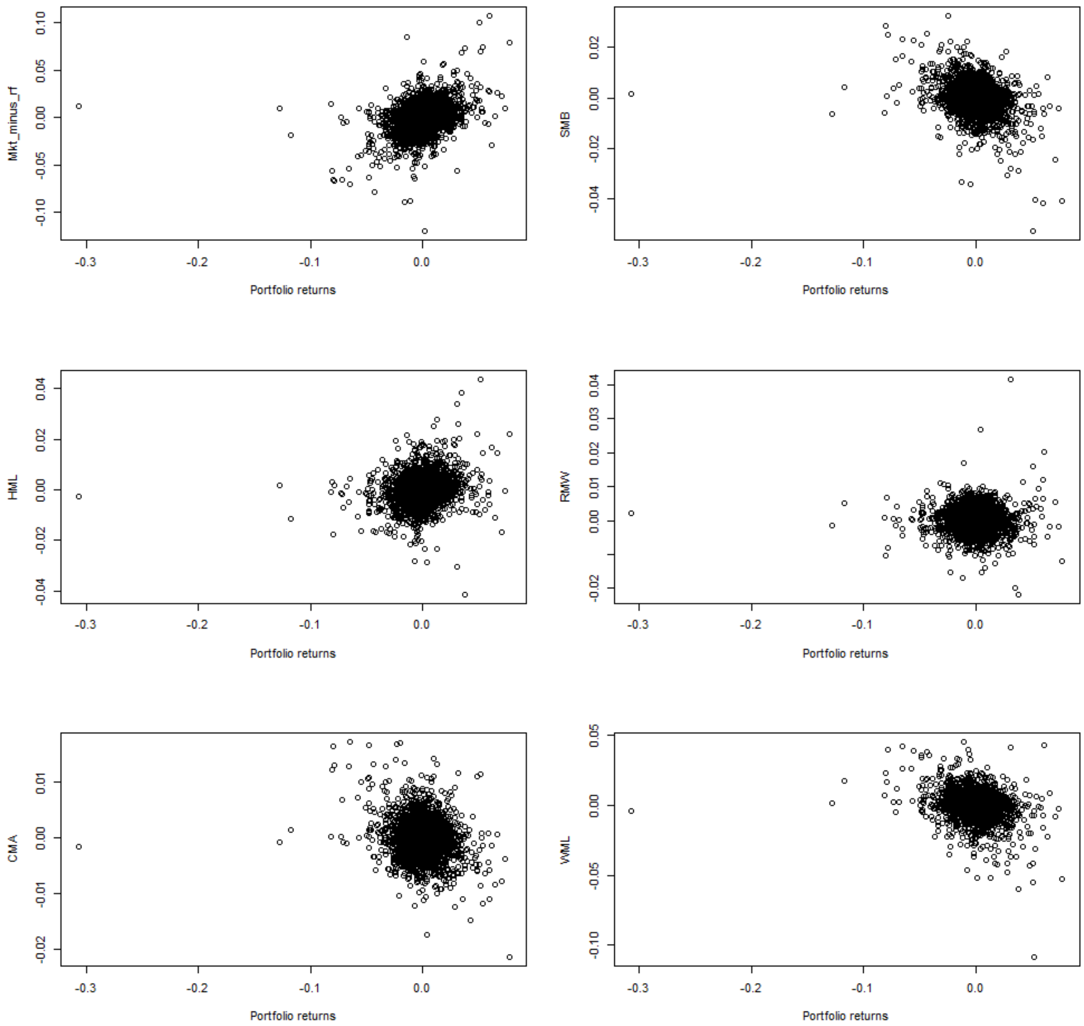
	Dependent variable: R_i-R_f		
	ZI	BIC	NS
R_m-R_f	0.190***	0.613***	0.510***
	p = 0.000	p = 0.000	p = 0.000
SMB	-0.266***	-0.235***	-0.022
	p = 0.005	p = 0.00001	p = 0.248
HML	0.145	0.113	0.017
	p = 0.350	p = 0.386	p = 0.497
RMW	0.061	0.164	-0.087**
	p = 0.754	p = 0.315	p = 0.028
CMA	-0.171	-0.143	-0.040
	p = 0.249	p = 0.135	p = 0.335
WML	-0.048	-0.073*	0.001
	p = 0.396	p = 0.059	p = 0.959
α	-0.0005	-0.0004*	0.00005
	p = 0.112	p = 0.062	p = 0.538
Observations	1,758	1,758	1,758
R2	0.257	0.807	0.966
Adjusted R2	0.255	0.806	0.966
Note:			*p<0.1, **p<0.05, ***p<0.01

This table present the results on practical portfolios (2008-2014) regressed on Fama French Five-Factor model with momentum. The table show risk factor loadings and abnormal returns (α) on three equal-weighted practical investing strategies constructed with ESG score. The regression has excess returns over the risk-free rate as dependent variable. R_m-R_f is the market factor, SMB is the size factor, HML is the value factor, RMW is the profitability factor, CMA is the investment factor and WML is the momentum factor. The α of the models represents any effects that ESG will have on the portfolio's abnormal returns. The p-values are based on robust standard errors. These portfolios are restricted to rely solely on ESG scores that are available at the time of investment. The portfolio's composition is a result of the scores the companies received 2 years prior to investment date. The regression thus shows the relationship between ESG score and future performance.

Table 13 show that the ZI (α = -0.0005, p = 0.112) and BIC (α = -0.0004, p<0.1) portfolios achieve negative abnormal returns.

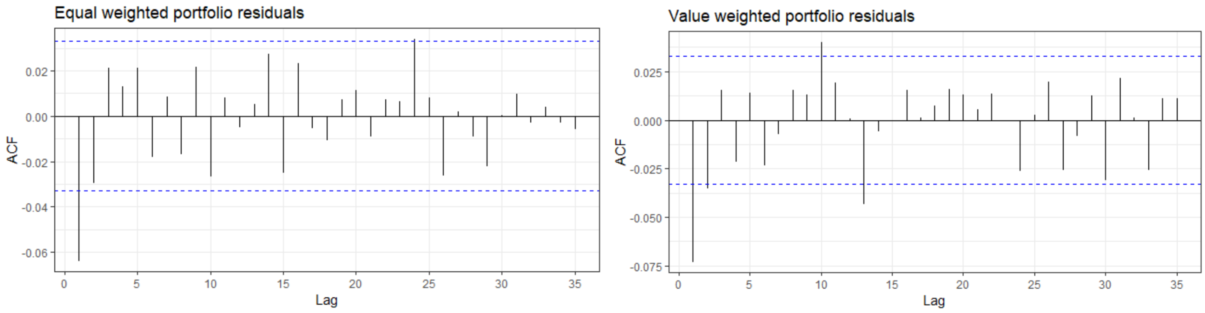
Appendix 6: Linearity check

Figure 13: Linearity check



Appendix 7: ACF Plots

Figure 14: ACF plots

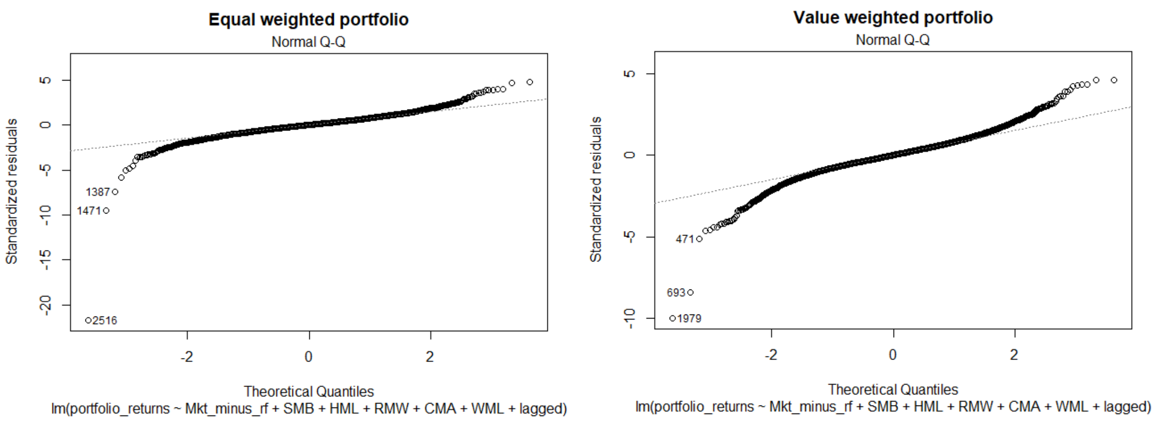


Appendix 10: If one of the vertical line crosses the blue dotted horizontal line, it means that we may have autocorrelation for the respective lag. We can see that the EWP model has a negative autocorrelation on lag = 1, while the VWP model has autocorrelation on lag = 1 and possibly on lag = 2. It may also look like there is autocorrelation on lag = 10 and lag = 13, but this is probably random. Basic intuition tells us that portfolio's returns today should not be dependent on returns exactly 10 and 13 days ago. Autocorrelation on lag = 1 means that yesterday's returns affect today's returns.

Appendix 8: Quantile plots

Quantile plots compare two probability distributions by plotting the respective quintiles against each other. The x-axis consists of theoretical quintiles, and the y-axis consists of the model's quintiles. If the points are in a straight line, the residuals of the model will completely coincide with theoretical quintiles.

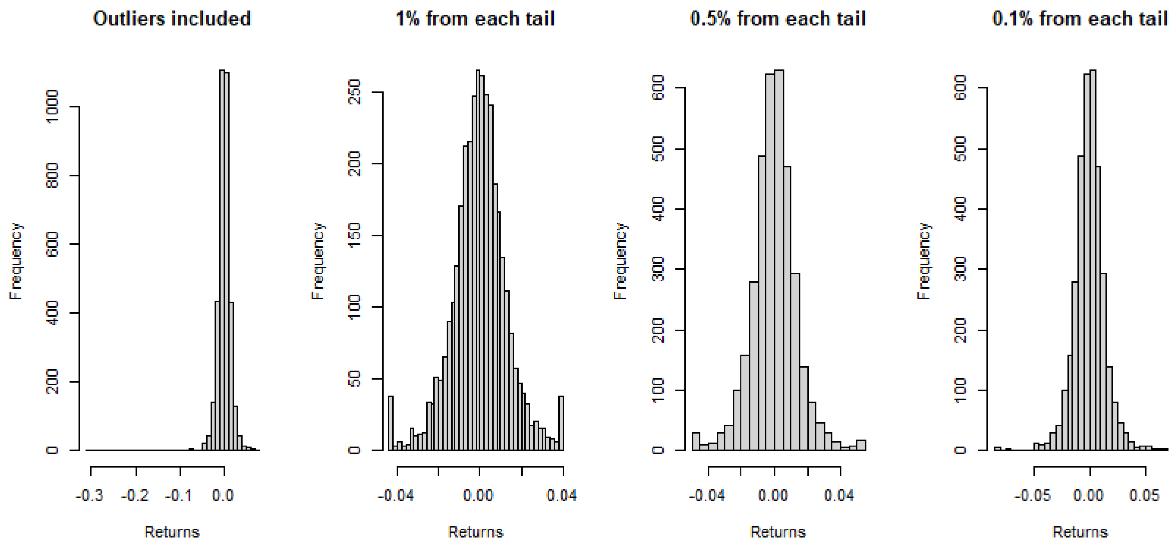
Figure 15: Quantile plots



We see that the points are mostly on the same line but with a downward float to the left and an upward float to the right of the graph. This pattern can indicate that we generally have normally distributed residuals, but with longer tails than what is theoretical.

Appendix 9: Outlier Removal

Figure 16: Outlier removal



Appendix 10: CAPM of Equal-Weighted Portfolios

Table 14: CAPM of equal-weighted portfolios

	Dependent variable: R_i-R_f		
	ZI	BIC	NS
R_m-R_f	0.294^{***}	0.681^{***}	0.524^{***}
	p = 0.000	p = 0.000	p = 0.000
Constant	-0.001^{**}	-0.0005^{**}	-0.0001
	p = 0.034	p = 0.020	p = 0.342
Observations	3,264	3,264	3,264
R^2	0.100	0.766	0.950
Adjusted R^2	0.100	0.765	0.950
Note:			*p<0.1, **p<0.05, ***p<0.01