

Neumann János Egyetem

MNB intézet: Fenntartható pénzügyek központ

Nemzetközi Gazdaság és Gazdálkodás mesterszak



OTDK 2025

Do firms take seriously ESG considerations? Application of the Structural Topic Modeling

Komolyan gondolják a cégek az ESG szempontokat? Strukturális Témamodellezés (STM) alkalmazása

Siket Bence

Konzulens: Granát Marcell Péter

Kézirat lezárásának időpontja: 2023.04.09.

Abstract

In this paper, I test the ESG-related considerations of the S&P500 firms using their annual CSR disclosures. A large part of the ESG related information comes from verbal communication, and thus it is vital to understand textual documents in order to get the most accurate picture of firms' ESG-related attitudes. I estimate whether the emphasis of certain topics extracted from the S&P500 firms' yearly CSR disclosures can significantly predict the deterioration or improvement of future ESG scores. By retrieving the ESG-related scores of Refinitiv from 2014 to 2022, I define three types of predictors to the Structural Topic Model. Based on the empirical results, I find that qualitative aspects play a significant role in the development of ESG scores. The results also support the argument that external rating agencies also pay significant attention to the topical content of disclosures; therefore the topical content of CSR disclosures has an impact on the ESG scores. The research also highlights that within the ESG framework there are clearly identifiable topics that play a very crucial role for both firms and external rating agencies. The topics related to food waste, supply chain, waste management and greenhouse gas emissions significantly contribute for the development of the ESG and its subcategories scores.

Table of contents

1. Introduction	1
2. Related literature	4
2.1 Overview of the ESG framework	4
2.2 Discussing sustainability related topics: ESG and CSR reports	6
3. Data	8
3.1 Textual data: CSR reports.....	8
3.2 Measuring CSR impact: ESG scores	10
3.3 The covariates estimated in the model	12
4. Methodology	14
4.1 STM overview	14
5. STM model.....	16
5.1 Running the base STM models.....	17
5.2 Finding optimal number of topics	18
6. Empirical results.....	21
6.1 Estimating the relationship between the proportion of topics and covariates	21
6.2 Identify and explore the statistically significant topics	26
7. Conclusion.....	30
Literature	31

List of tables

1. Table: Descriptive statistics about the webscraped ESG reports per year and the average number of pages related to the documents	9
2. Table: Summary statistics of yearly ESG scores grouped by the presence or absence of ESG report availability	11
3. Table: Description of covariates used for the estimation of topic proportions	12
4. Table: Description of the final inputs derived from the Eikon data to test the significance of topic proportions across groups.....	13
5. Table: Exploring document topic proportions and topic-term distributions in STM models with 5 and 10 topics	20
6. Table: Table: Results for the statistically significant topics in case of the ESG and CSR Strategy scores.....	26
7. Table: Results for the statistically significant topics in case of the Emission Score	27
8. Table: Results for the statistically significant topics in case of the Resource Usage Score.	28
9. Table: Results for the statistically significant topics in case of the Community Score	29

List of figures

1. Figure: A possible graphical representation of the LDA model	15
2. Figure: A possible graphical representation of the STM model	16
3. Figure: Model diagnostics with different criterias by number of topics	18
4. Figure: The p values from the regressions estimating the topic proportions for different groups in case of CSR Strategy and ESG scores	22
5. Figure: The p values from the regressions estimating the topic proportions for different groups in case of Environmental pillar related variables	23
6. Figure: The p values from the regressions estimating the topic proportions for different groups in case of Community ('S' pillar) and Management ('G' pillar) scores	25

1. Introduction

Topics related to climate change and sustainability has become a natural part of today's political and economic debates. These debates are becoming more prominent and intense since a result of climate change and global warming, extreme weather conditions and events are becoming more frequent and larger (Stott, 2016; Huber and Gullede, 2011).

Sustainability efforts have inevitably reached companies as well. Recently firms place much greater emphasis on understanding sustainability related risks and mitigating the risks arising from the climate change. In fact, in recent years, in addition to profit seeking and efficient operation, taking into account the environmental sustainability and social aspects has become almost equally important objective of the firms. So that companies can be compared in some framework, there was a significant demand for the quantification of environmental and social performance of firms from the regulatory and investment sides.

ESG scores capture the performance of companies along three major pillars – Environmental (“E”), Social (“S”) and Governance (“G”) – and serve as a basis for investors who would consciously incorporate these aspects into their investments during portfolio formation. Although the financial markets faced many challenges in 2022 - and accordingly ESG funds too – based on the Morningstar data¹, 65 percent of ETF inflows in 2022 were concentrated in funds that invest in stocks or bonds aimed at reducing the risks arising from climate change or pays significant attention to its social or management aspects. Despite there are many criticisms of the ESG framework – mainly due to the high degree of disagreement between the ESG rating agencies –, Talan and Sharma (2019) found ESG scores to be one of the most effective and widely used indicators for quantifying the environmental, social and governance performance of firms.

The external rating agencies usually calculate ESG scores as a combination of qualitative (aspirations, programs, policies) and quantitative information (CO2 emission, water usage efficiency, inequality in incomes) focusing on ESG aspects (Christensen et al, 2022). The annual CSR report disclosures published by the firms serve serves as one of the most important input information for their rating. Since the ESG scores are partly derived from hard data that can be calculated relatively easy, and partly from soft information – which include aspects such as emphasising certain topics on a larger extent or introducing energy efficiency program –, a precise understanding of the ESG scores requires significant work.

¹ <https://www.ft.com/content/a3e9d87f-fa6f-4e5e-be6e-e95b42af2fec>

The explanatory power of text information content therefore creates an opportunity for researchers to apply text analysis methods to discover the ESG framework.

In this research I focus on exploring the qualitative aspect of the ESG framework. In particular the research aims to answer whether the greater proportion and emphasis of certain topics extracted from the S&P500 firms' yearly CSR disclosures can significantly predict the deterioration or improvement of future ESG scores. The question can also be approached in the way that, apart from the numerical indicators, is it possible to extract additional information from the 60-70 page reports, which can have explanatory power in the future worse or better ESG performance of the firms?

To assess the predictive power of ESG related topics on the ESG scores (and on the sub dimensions of the ESG), I first retrieve the yearly ESG disclosures of S&P500 firms and convert them into a format suitable for topic modelling. Focus on the S&P500 companies is beneficial in several ways. First, aspiration towards reporting and transparency is higher. Second, investors and external rating agencies pay more attention to the ESG performance of these companies. Then in order to distinguish between good and bad performing companies in a year-on-year (YoY) comparison, I use Refinitiv's data to construct explanatory variables between 2014 and 2021. These variables then serve as predictors in the STM models, where I estimate whether the proportion of certain topics is significantly different between the groups. The stronger emphasis of some topics and in parallel the improvement in the received hard data may indicate that the firms are really serious about the given ESG topic and would take large-scale steps to mitigate risks. This result would suggest that the disclosure practice of CSR reports is effective, meaning that their topical content has an impact on the ESG scores.

Based on the empirical results, I can draw the following conclusions regarding the CSR disclosures of the S&P500 firms. First, in line with the definition of the CSR Strategy and ESG scores, the former is more sensitive to the qualitative aspects, therefore the topic proportions alone can significantly explain the better and worse performance of firms. Second, although the Refinitiv takes into account the sector when assessing the ESG performance of firms, the performance compared to the sector explains more significantly the topic proportions compared to the firms' own previous year's performance. Third, although the ESG framework covers a lot of topics, the topic discussing food waste and supply chain is particularly important: the smaller proportion of food waste and supply chain terms compared to other ESG-related information may result in a worse score compared to the sector after the disclosure of the CSR report. Fourth, regarding the development of two significant indicators

in the environmental pillar – the Emission and Resource Usage scores – fossil fuels, waste management, food waste and supply chain topics play significant roles in explaining the scores received by the Refinitiv. The higher proportion of fossil fuels related topic, while the lower share of other topics compared to other ESG-related topics contribute to the deteriorating performance. As for the social pillar of the ESG framework, the Community Score is sensitive to many topics even for many non-social related topics. Nevertheless, the topic dealing with management and business aspects by a larger extent corresponds to a worse performance in community score compared to the sector. Considering the governance pillar of the ESG, no topic was identified, that would separate the firms in terms of their future performance.

The research can provide several implications. First, apart from the mandatory publication, the disclosure of CSR reports is indeed important as the text content of these reports complements the quantitative indicators. This confirms the fact that qualitative aspects play a significant role in the development of ESG scores. Based on the empirical results, it seems that external rating agencies also pay significant attention to the topical content of disclosures. Accordingly, this result indicates that the disclosure practice of CSR reports is effective, meaning that their topical content has an impact on the ESG scores. Second, within the ESG framework there are clearly identifiable topics that play a very crucial role for both firms and external rating agencies. The topics related to food waste, supply chain, waste management significantly contributes for the development of the ESG and its subcategories scores. Last, in addition to the conventional regression methods, supplementing the analysis toolbox with completely different approaches – for instance with topic models – can be beneficial for the more accurate understanding of the very diverse field of ESG.

The rest of the paper continues as follows. In the second section, I describe the related literature discussing the ESG framework and the annual Corporate Social Responsibility report disclosures of firms. The third section presents the data including the yearly CSR disclosures of firms and the description of covariates used for the estimation of topic proportions. In the fourth section, I give an overview of the applied Structural Topic modelling methodology. Section fifth presents the estimation and comparison of STM models. In the sixth section, I present the main empirical findings of this research. The final section concludes the research.

2. Related literature

In this chapter, I divide the literature overview into two distinct parts to present the quantitative and qualitative dimensions of the growing corporate social responsibility framework. In the first part, I cover a controversial topic: the ESG scores used by external rating agencies and the criticism surrounding them. In the second part of the related literature, the definition and development of corporate social responsibility reports (henceforth CSR reports) and related studies are discussed.

2.1 Overview of the ESG framework

The concept of the ESG framework dates back to 2004 when the UN Global Compact Initiative published the “Who cares Wins” (UN, 2004). With the ESG framework and indicators related to the term, three distinct non-financial pillars are defined: environmental (“E”), social (“S”) and governance (“G”). The environmental pillar captures issues, such as climate change, water pollution, and biodiversity loss. The social pillar focuses on aspects that consider gender policies, labour standards, public health, and income distribution. The governance pillar is related to the strength of shareholders’ rights, control procedures, respect for the law, or the CSR strategy of the given firm. From the above examples, it is clear that the pillars of the ESG framework not only differ significantly in terms of what they capture, but also in terms of how an external rating agency can quantify the extent to which a company is exposed to certain ESG components. For example one of the largest ESG rating providers, Refinitiv calculates the environmental and social category scores compared to the industry group as a benchmark, while in the case of governance categories, the country of incorporation is used as a benchmark.²

Given the growing demand from investors – in parallel with the spike in sustainable investing – and government institutions for the most accurate possible quantification of firms’ ESG performance, external rating agencies have allocated significant resources to these areas in recent years (Billio et al., 2021). In fact, some larger financial data provider companies have recognised the potential of the ESG rating market so much that they have acquired many early ESG rating providers. For instance, MSCI acquired the first ESG rating provider, KLD, in 2010, while another significant participant Morningstar bought 40% of the shares of Sustainalytics (Berg et al, 2022). These rating agencies are similar to credit rating agencies in the sense that they allow investors to screen companies’ performance along some kind of

² <https://www.refinitiv.com/en/sustainable-finance/esg-scores#methodology>

division, in this particular case, in the field of addressing environmental, social, and governance issues.

The daily consideration of ESG aspects, in addition to other factors such as market cap, price-to-earnings ratio, or book-to-market ratio, indicated a remarkable increase in assets related to environmental, social, and governance dimensions. Accordingly, the inflow to ESG investments has continued to skyrocket in recent years: ESG-focused funds set a record in 2021, after 2019 and 2020, and according to Refinitiv data, ESG funds now account for approximately 10 percent of the world's funds.³ Although managing the risk of “greenwashing” from the regulatory side will continue to be a great challenge in the future, significant attention is expected to be focused on ESG-based and responsible investments in the future as well. Therefore, ignoring relevant topics, such as climate change, water pollution, or biodiversity loss, may indicate a competitive disadvantage for larger companies listed on the stock exchange. Several studies have investigated the beneficial effects of favourable ESG performance through the channel of strong ESG performers to gain higher stock returns (Albuquerque et al., 2019; Lins et al., 2017; Engelhardt et al., 2021). In addition, some other channels were also recognised: a firm's commitment to ESG aspects reduces uncertainty and risk and publicly increases positive perception among investors (Bassen et al., 2006), and firms with higher ESG scores face a lower cost of capital (Kotsantonis and Serafeim, 2019).

However, many other studies find that addressing ESG issues are not necessarily accompanied by significant benefits for firms. Stambaugh and Taylor (2021) point out that firms with high ESG scores generate lower expected returns than firms with low ESG scores. The pursuit of sustainability issues during the operation of companies is partially opposite to the well-known concept raised by Friedman (2007), according to which the main responsibility of a company is the maximisation of shareholders' returns. The positive relationship between increasing awareness of ESG-related issues and financial performance has been questioned; however, the opposite has been declared. The growing operational costs stemming from addressing ESG issues worsen the profitability of companies.

Overall, empirical evidence is not conclusive, in which the disagreement of ESG rating agencies also partially plays an important role (Abhayawansa and Tyagi, 2021; Baker et al., 2022). Numerous studies agree that the large difference between ESG ratings for the same company stems from the immense number of aspects that fit into the ESG framework,

^{3 3} <https://www.reuters.com/markets/us/how-2021-became-year-esg-investing-2021-12-23/>

accurate measurement of ESG aspects, and the judgement of introduced tools addressing environmental, social, and governance issues. Abhayawansa and Tyagi (2021) highlight the disagreement in ESG ratings is mainly due to the differences in the definitions of ESG constructs (theorization problem) and the methodological differences (commensurability problem). Berg et al (2022) also argue that the low pairwise Pearson correlations across the ESG rating is mostly explained by differences in measurement, indicating the imperfect measurement of ESG attributes. Since ESG factors include a wide range of topics, and the assessed ESG factor significantly depends on the company being analysed, the ESG scoring framework does not have a clearly defined list of the factors that external rating agencies observe. All of these rating agencies have defined their own proprietary definitions of materiality (Billio et al., 2020). As a consequence of disagreement in several aspects, the correlation across ESG ratings is quite weak overall. According to Billio et al (2021) the average correlation between the ESG ratings of the largest participants is on average 0.58 and ranges from 0.43 (between MSCI and Refinitiv) to 0.69 (Refinitiv and RobbcoSAM).

Although there have been many criticisms of the ESG framework, according to Talan and Sharma (2019), ESG scores are one of the most effective and widely used indicators for quantifying the environmental, social, and management aspects of companies. Accordingly, in this study, I also use the ESG score framework in the STM model to capture the relationship between ESG score performance and topics in CSR reports. In fact, due to the lack of standardization and universally accepted approach in ESG activities measurement, extracting information from annual reports may add relevant explanatory power to the separation of favorable and unfavorable ESG performances.

2.2 Discussing sustainability related topics: ESG and CSR reports

The corporate social responsibility framework captures a firm's commitment to operating economically, socially, and environmentally sustainably (Dao et al, 2011). The concept of Corporate Social Responsibility for capturing firms' activities addressing environmental issues, employment discrimination, and unethical practices has become a critical aspect of operations since the 1980s. This was also the point when companies allocated special resources to these areas and accordingly began to report these activities and initiatives to the customers and investors in a transparent manner.

ESG reports are usually annual text disclosures including both qualitative (aspirations, programs, policies) and quantitative information (CO2 emission and water usage efficiency)

focusing on ESG aspects (Kiriū and Nozaki, 2020). Since sustainability and socially responsible investing (SRI) are increasingly important for investors, sustainability reporting concerning ESG activities is becoming more crucial for large firms. The SRI is a simple process identifying investments which put a great emphasis on environmental, social and governance (ESG) values (Renneboog et al., 2008).

According to McKinsey, nowadays more than 90% of S&P500 companies publish ESG reports.⁴ In line with this increasing disclosure tendency, KPMG also highlighted that based on early revenues almost 80% of the 100 largest companies in the world publish some kind of sustainability report⁵. The reporting rate increased by almost 10 percentage points within a decade. The significant emergence of the ESG framework is mostly driven by the environmental component. However, the social dimension has also caught great attention in recent years. China also recognized the upward tendency in CSR and ESG reporting: two large stock markets, Shenzhen Stock Exchange (SZSE) and Shanghai Stock Exchange (SSE) have published guidelines to incentive listed companies to disclose textual reports focusing on ESG concerns. The remarkable growth in the disclosure of CSR reports is also enhanced by the Purpose Driven Disclosure (PDD) Project. The World Business Council for Sustainable Development (WBCSD) initiated the project with the aim of by 2050 all companies will measure and report their truthful value and profit including natural and social capital besides the financial capital.

Unlike structured financial reports, CSR reports vary to a great extent not just in terms of appearance but also in the way ESG topics are conveyed. As standardization is still ongoing in the field of ESG reporting these reports may include a large part of narrative information (Ho et. al, 2022). In parallel with this, Baier et al. (2020) argue there is no standardized definition of ESG criteria. Focusing on the United States, Rouen et al (2022) highlight the lack of standardization, mandate and audit and regulation contributes to a high degree of noise in the field of ESG reporting. ESG criteria are often referred to as corporate social responsibility, and according to SRI researchers, the measurement is quite challenging since CSR activity is mostly based on qualitative information. To solve this, regulators gradually introduce reporting mandates to provide a more transparent framework for ESG disclosures (IFRS, 2021). The growing need for transparency and non-financial disclosures from investors and stakeholders strengthens voluntary ESG disclosures (Beattie, 2014).

⁴ <https://www.mckinsey.com/capabilities/sustainability/our-insights/does-esg-really-matter-and-why>

⁵ <https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2022/10/ssr-small-steps-big-shifts.pdf>

The purely quantitative-based approach for exploring CSR reports has received considerable criticism over the years as it has not considered enough aspects besides being too mechanistic (Tregdida et al. 2007). Given the large length and wide variety of content of CSR reports researchers started to exploit the advantages and potentials of text mining approaches. Extracting information from textual data has received a great deal of attention in recent years. In addition to the growing number of studies in the field of consumer satisfaction extracted from soft data, analysing textual information has also received considerable emphasis in the field of economics. Exploring the content and underlying factors of non-financial disclosures (such as ESG or CSR reports) is not only interesting from a descriptive point of view, but also by providing insights to the environmental and social goals of companies (Jaworka and Nanda, 2018).

3. Data

In this chapter I present both the qualitative and quantitative data used for the estimation of the structural topic model. First, I describe the selected ESG scoring agency and the approach the rating agency follows to quantify the ESG attributes. In the second section, the method and source of retrieving the firms' CSR reports are discussed. Third covariates estimated in the model are presented.

3.1 Textual data: CSR reports

Most of the published research within the field of textual analysis uses a wide range of sources to obtain the CSR and ESG reports of firms. In previous years, the global reporting initiative (GRI) – a non-government organization providing CSR disclosure guidelines and publish a numerous number of CSR reports in a structured database – became a standard source for academics to capture the ESG aspirations of companies in a textual form (Ning et al, 2021; Goloshchapova et al., 2019). Unfortunately for some reason, the GRI finished collecting the ESG reports and it is no longer possible to access Sustainability Disclosure Database. Due to the fact that there is no centralized database of ESG reports, Rouen et al (2022) use a large-scale hand-collection approach. Rouen et al (2022) also point out that the empirical analysis of ESG reports is hindered by the fact there exists no official data source which contains a large number of corporate social responsibility reports.

Given the limited number of free sources and that, as far as I know, there is no officially accepted benchmark ESG report database I use the easily accessible

ResponsibilityReports.com website.⁶ According to the company's description the site allows to search for 13,011 responsibility reports (including sustainability reports, corporate responsibility reports, corporate social responsibility reports, and ESG reports) from 2,468 global companies. Since in the research I focus on the listed S&P 500 companies the New York Stock Exchange (NYSE) and Nasdaq Technology Exchange (NASDAQ) markets were chosen. Retrieving annual reports are advantageous since it contributes to the comparability of firms' CSR activities and annual reports convey textual information with a high degree of credibility (Chan et al, 2014).

1. Table: Descriptive statistics about the webscraped ESG reports per year and the average number of pages related to the documents

Sector	Number of CSR reports	Average number of pages	Median number of pages
Communication Services	40	68.1	63.5
Consumer Discretionary	205	58.9	52.0
Consumer Staples	164	75.1	73.5
Energy	111	72.6	65.0
Financials	275	58.9	49.0
Health Care	221	61.9	52.0
Industrials	266	61.7	53.0
Information Technology	229	63.8	54.0
Materials	139	62.0	52.0
Real Estate	129	47.4	45.0
Utilities	138	79.2	66.5
Total	1918	63.5	54

The table shows descriptive statistics about retrieved CSR reports from ResponsibilityReports for the S&P500 companies between 2010 and 2021. The statistics are based on sector grouped values. The second column of the table presents the number of CSR reports in a given sector, the third column refers to the average number of pages in the reports calculated in a specific sector, while the last column shows the median values for the number of pages. Source: ResponsibilityReports.com

According to Table 1, I retrieved overall 1918 CSR reports from the ResponsibilityReports site between 2010 and 2021.⁷ On average the number of pages is 63.5 and the median number of pages is only 54, indicating that within the CSR reports the number of pages draw out a left skewed shape. Looking the table in more detail, the most reports are available for the financial sector (275) followed by the industrials sector (266). The least number of CSR reports were scraped in case of the communication services sector (40). As for the average number pages across the sectors, quite significant differences can be observed in terms of the length in the reports. The number of pages is the highest in the utilities and consumer staples sectors, while the lowest in the real estate sector. Using textual analysis, Kimbrough et al (2022) find that

⁶ <https://www.responsibilityreports.com/>

⁷ Most of the results can be found on the related GitHub storage: https://github.com/kukszo8/TDK_CSR-reports-and-ESG-scores

longer reports are associated with reduced disagreement among ESG raters. Although the focus of the research is not the examination of length and complexity of the CSR reports, an important research question can rise from how these aspects related to the performance in ESG score.

3.2 Measuring CSR impact: ESG scores

In addition to differences in the definitions of ESG and measurement divergences, the ESG rating industry – unlike the credit risk rating industry where only three major firms can be identified (S&P, Fitch, Moody’s) – is a quite fragmented industry with a lot of participants. Without claiming to be complete, the following are the biggest players in the industry, of which there are quite a few came from a different business area: MSCI, ISS ESG, Sustainalytics, Refinitiv, Moody’s ESG, S&P Global (previously RobecoSAM).⁸

In this research as a covariate input to the structural topic model (STM) I use the ESG data provided by the Refinitiv Eikon Datastream (TRESG score). Several studies obtain the firms’ ESG data from the Refinitiv dataset (De Lucia et al, 2020; Eliwa et al, 2019). Berg et al (2022) emphasize the complexity of Refinitiv’s TRESG, as it captures the most ESG-related topics with 282 indicators followed by Sustainalytics with 163 indicators. The ESG score measures the company’s ESG performance based on verifiable reported data in the public domain. Therefore, a firm’s ESG performance is based on the content of annual reports, CSR reports and information found on company and NGO websites. Most of companies disclose their ESG or CSR report once a year, which gives the most comprehensive picture of the company's performance and aspirations. Accordingly, in most cases, the ESG data is also updated once a year.

Refinitiv assesses the firms’ relative ESG performance, commitment and effectiveness across 10 main themes based on publicly available and auditable data.⁹ The three main pillar scores – environmental, social and governance – are derived from the 10 category scores. The final ESG pillar score is calculated as the relative sum of category weights which vary in the case of ‘Environmental’ and ‘Social’ categories across industries, while the weight of ‘Governance’ remains unchanged. Refinitiv treats underlying data points as either boolean or numeric. The type of boolean data points is typically answered with yes or no and assesses mostly soft information related to ESG topics. For instance, a boolean question can be the

⁸ <https://www.spglobal.com/en/research-insights/articles/sp-global-to-acquire-the-esg-ratings-business-from-robecosam-shell>

⁹ <https://www.refinitiv.com/en/sustainable-finance/esg-scores>

following: “Does the company have a water efficiency policy?” It is clear from the example question that these boolean types of data points capture the goals and strategies of the companies and perhaps these types of evaluations can be better explored using textual analysis methods such as introducing the Structural Topic Model (STM). Table 2 shows the yearly summary statistics of Refinitiv ESG scores for firms with available CSR reports and companies without CSR reports.

2. Table: Summary statistics of yearly ESG scores grouped by the presence or absence of ESG report availability

	Q1 ESG values		Q2 ESG values		Q3 ESG values	
	Has report	Has no report	Has report	Has no report	Has report	Has no report
2014	56.3 (91)	37.3 (198)	67.6 (91)	49.5 (198)	74.9 (91)	62 (198)
2015	55.3 (105)	41.1 (192)	67.4 (105)	52.5 (192)	77.3 (105)	64.6 (192)
2016	58.1 (130)	41.9 (178)	68.2 (130)	55.1 (178)	76.8 (130)	66.2 (178)
2017	61.7 (159)	42.8 (150)	69.0 (159)	55.0 (150)	77.7 (159)	63.7 (150)
2018	61.9 (187)	43.3 (125)	69.7 (187)	56.2 (125)	77.7 (187)	67.9 (125)
2019	61.6 (209)	46.7 (104)	68.7 (209)	61.3 (104)	76.7 (209)	72.1 (104)
2020	63.0 (192)	53.4 (92)	71.2 (192)	68.1 (92)	77.8 (192)	76.3 (92)
2021	79.33 (4)	56.3 (18)	83 (4)	70.5 (18)	85.4 (4)	72.5 (18)

The tables present the distribution of ESG scores retrieved from the Refinitiv Eikon for the S&P500 companies between 2014 and 2021. The distributions are captured by the Q1, Q2 and Q3 quarterlies. The rows represent the values grouped by the years, while the column separates further the ESG values by the presence or absence of ESG report availability. The number of observations per groups can be found in parenthesis. Source: Refinitiv Eikon

Based on the summary statistics of yearly ESG scores for S&P500 companies several conclusions can be made. First, in general, there is an increasing tendency in the ESG scores both for companies with ESG reports available and both for companies where I could not obtain the annual text report. Second, the share of firms providing ESG related information in a form of non-financial reports represents an increasingly large proportion compared to firms where ESG reports were not available. Third, considering the sample with the ESG score retrieved from the Refinitiv Eikon and the CSR reports scraped from Responsibilityreports.com, the companies with yearly ESG disclosure has a lower ESG score compared to companies where the ESG report was available. Although the difference between the two groups has decreased somewhat since 2014, companies with ESG reports still perform better in environmental, social and governance aspects. This highlights an aspect according to which the presence of the ESG disclosure report in itself leads can lead to an improvement in ESG score due to the increase in transparency and credibility. However, examination of this issue is not the subject of this research.

3.3 The covariates estimated in the model

Understanding the relationship between the ESG scores and the textual content of CSR reports may be quite challenging for researchers in the field of sustainability. As described previously in detail, ESG scores are given as a combination of soft (qualitative) and hard (quantitative) indicators. In order to identify which of the ESG indicators are related to the topics conveyed from the STM models, I take into account a number of indicators. In the focus of this research I quantify whether the emphasis of the topics (topic distributions) differs significantly in the groups constructed along some dimension from the ESG related covariates. In particular, I am interested in whether it is possible to predict the deterioration or improvement of firms' future ESG performance by emphasizing certain topics more prominently?

Accurate consideration of the time dimension of the ESG scores and CSR report disclosures is also critical, as reporting on ESG activities is usually done retrospectively. For instance, a report summarizing ESG activities for the year 2020 is usually published by the company in 2021. So that the topics derived from the ESG reports and the different covariates are consistent in terms of time dimension, I take the lead value of covariates into account when interpreting the relationships. In table 3, I present the covariates and their description constructed from the ESG related scores from Refinitiv.

3. Table: Description of covariates used for the estimation of topic proportions

Pillar	Name	Code	Range	Description
ESG	ESG score	TRESGS	0-100	ESG Score is an overall company score based on the self-reported information in the environmental, social and corporate governance pillars.
E	Emission	TRESGENERS	0-100	Emissions Score measures a company's commitment to and effectiveness in reducing environmental emission in the production and operational processes.
E	Resource usage	TRESGENRRS	0-100	Resource Use Score reflects a company's performance and capacity to reduce the use of materials, energy or water, and to find more eco-efficient solutions by improving supply chain management.
S	Community	TRESGSOCOS	0-100	Human Rights Score measures a company's effectiveness in respecting the fundamental human rights conventions.

G	Management	TRESGCGBDS	0-100	Management Score measures a company's commitment to and effectiveness in following best practice corporate governance principles.
None	CSR Strategy	TRESGCGVSS	0-100	CSR Strategy Score reflects a company's practices to communicate that it incorporates the economic (financial), social and environmental dimensions in its day-to-day decision-making processes.

The table shows description of covariates used for the estimation of topic proportions. It is important that these are not the specific inputs to the model, but this is the basis from which I form the various categorical variables. The first column refers to the pillar where the certain indicator belongs to. The 'none' term at the CSR strategy means that this variable is not part of the composite ESG score but rather an indicator developed by Refinitiv to capture mostly the qualitative side of the ESG performance of a given company. The second and third column presents the name and code under which the certain indicator can be found in the Thomson Reuters Datastream. Fourth column refers to the score range on which the indicator can take the value. The last column provides detailed information with respect to the selected covariates. Further information can be found the following site. <https://www.refinitiv.com/en/sustainable-finance/esg-scores>.

It is important that the abovementioned raw scores are not the specific inputs to the model, but they form the basis from which I construct the various categorical variables to separate the good and bad performing companies along some dimension. Therefore, as a next step from the raw scores I conduct three input variables for each variable. In each year, for each company, the created variables reflect the given firm's ESG performance compared to itself, compared to the performance of the sector and compared to the average performance of the S&P 500 companies. With using these variables as independent variable and the average topic proportion as dependent variable in the STM model, I can estimate along which dimension the emphasis (topic proportion) of each topic differs significantly. This estimation method allows the possibility to draw conclusions from the predictability of ESG scores from the textual content of the yearly CSR reports. Table 4 shows the final covariates used as input to explain the topic proportions.

4. Table: Description of the final inputs derived from the Eikon data to test the significance of topic proportions across groups

Abbreviation	Data type	Description
Change_comp_firm	Character	Measuring whether the firm improved its score in a given dimension compared to the previous year (YoY change).
Change_comp_sector	Character	Measuring whether the firm improved its score in a given dimension to a greater extent compared to the sectorial average change (YoY change).

Comp_avg	Character	Capturing whether the firm performs better in a given dimension compared to the average score in the S&P500.
----------	-----------	--

The table presents the final inputs derived from the raw scores obtained from Eikon. For each of the indicator described at Table 3 (ESG score, Emission score, Resource usage score, Community score, Management score, CSR score) these three types of covariates were calculated. The first column shows the abbreviation of the given covariates which will contribute to the understanding of the estimated models in the next chapter. Second column of the table shows the type of data. The data type is character in all cases and calculated from with an if-else statement based on the description statement. The third column gives an overview how the categorical covariates were calculated. The first two covariates is based on a YoY performance, while the third covariate captures the performance in a given indicator compared to the average performance of the sample for the S&P500 companies.

4. Methodology

In this section, I describe the estimation method used to examine whether there is a relationship between the topics of CSR reports and ESG scores. In the section, in addition to the theoretical background, I also emphasize what kind of benefits arising from choosing the Structural Topic Model (STM) technique over the widely used Latent Dirichlet Allocation (LDA) topic model.

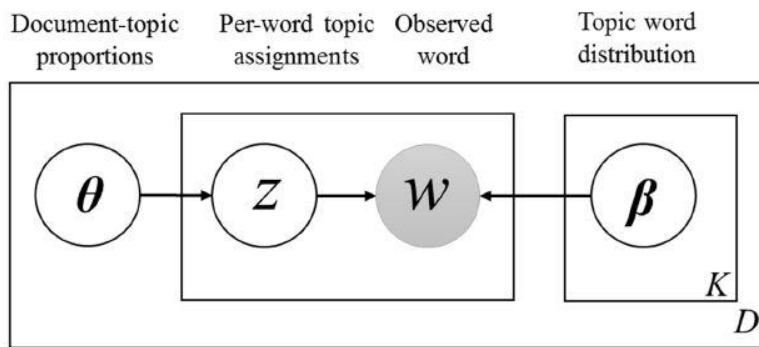
4.1 STM overview

The *Structural Topic Modeling* (STM) belongs to the topic modelling family namely the Bayesian generation which assumes that each text document is a mixture of topics and each topic is derived from distribution of words. These types of models allow uncover clusters (hidden topics) by summarizing very large amount of unstructured text documents. In other words, statistical topic models explore rich latent topics automatically derived from articles or documents (Roberts et al, 2014). Topic models are usually referred as unsupervised methods in the sense they infer rather than assume the content of topics. Typically when choosing supervised topic modelling techniques the user of the model determines the documents ex ante into one of the predetermined topics (Laver, Benoit, and Garry 2003). In topic modelling we focus on two types of information from the textual data: document-topic distribution and topic-term distribution. In my research the document-topic distribution provides information how can we break down the ESG reports in terms of topics. Meanwhile topic-term distribution conveys a list of words representing the topics in the documents. For instance, in ESG reports, a topic related to environmental protection (“E” pillar of the ESG framework) can dominate significant share of document topics, which is then represented by such words as water (40%), emission (20%), energy (20%), resources (10%), and reduction (10%). In

topic models the sum of topic proportions across all topics is equal to one and the sum of word probabilities for a given topic is also one (Roberts et al, 2019).

To understand the consideration behind topic models first I present the theoretical background of the Latent Dirichlet Allocation (LDA) model which is the most discussed model type in the field of topic modelling literature (He et al, 2020). By examining the LDA model, we can understand what additional and other desirable properties the STM model offers. In the LDA model we can typically observe the documents or texts and the words composing the document (w). However, document-topic proportions for a given document (θ), the topic word distribution (β) within a topic and the per-word topic distributions (z) are referred as latent (or hidden) layers. To explain as accurate as possible the observed word (w) we can run a generative process to discover the topic structure (θ, β, z) (Chae and Park, 2018).

1. Figure: A possible graphical representation of the LDA model



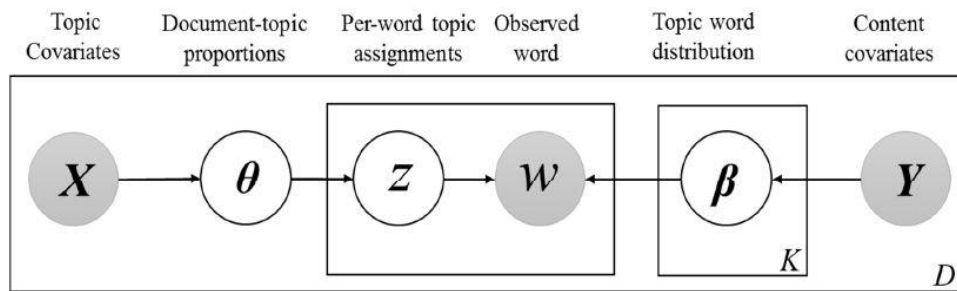
Source: Chae and Park (2018).

In general, choosing LDA for topic modelling we are able to discover large amounts of textual data in a structured manner by grouping the text into topics. However, in the LDA model both topic-word distribution and document-topic proportions are determined by the Dirichlet distribution of the data. The preset parameters in LDA do not allow incorporating any kind of information that may affect the document-topic and topic-word distributions of the text.

Since information related to the documents and texts has continuously expanded over time and the demand for additional information that can be extracted from documents has also increased by decision makers and researchers, STM has gained significant attention in recent years. For instance, Chae and Park (2018) highlights that by analysing corporate social responsibility reports we can include metadata which includes the year of ESG disclosure, company name, industry related information or any financial data. We can further extend the scope of possible variables with ESG related information such as the yearly change of ESG score or the ESG performance of the company compared to peer companies.

The key advantage and reason behind its recent popularity of the STM model is that permits its users to include covariates or additional document related information into the topic model (Roberts et al, 2014; Chae and Park, 2018; He et al, 2020). Under the application of the STM model, we apply a logistic-normal generalized linear model ($\theta \sim \text{LogisticNormal}(X)$), instead of the previously used Dirichlet distribution ($\theta \sim \text{Dirichlet}(\alpha)$) at the LDA model. Roberts et al (2014) identify three crucial differences in the STM model compared to the LDA model: (1) we allow topics to be correlated; (2) instead of sharing a global mean, over the selected number of topics, each document or text has its own topic distribution defined by covariate X ; (3) term usage within topics can vary with content covariate Y .

2. Figure: A possible graphical representation of the STM model



Source: Chae and Park (2018).

By adding relevant topical information or numerical variables to the existing framework, we can influence the document-topic proportions in the topic model. In fact, with the STM approach, we dissolve the assumption of LDA that there is no heterogeneity in the texts and emotions of authors or information related to the texts do not influence document-topic proportion or topic-word distribution. Choosing the STM model, researchers can explore the topics in the documents and in parallel estimate the relationship between document related covariates. Accordingly, the model allows us to conduct preliminary hypotheses about relationships and test them. For instance in the field of CSR reporting we can test whether the proportion of topic (topical prevalence) related to environmental aspects has significantly differ across groups with high ESG score and low ESG score. In addition, we can make conclusions for instance whether social and health wellbeing of employees are emphasized stronger in the recent years compared to the early of 2010's.

5. STM model

In this section, first the preliminary steps and filters are described for constructing the database for the STM modelling. Second, I run the STM model with a prevalence variable for

models with topics from 2 to 60. Third, I run different kind of selection criteria which serve as a basis of finding the optimal number of topics for the STM model. Finally, I consider whether it is worth choosing the final models based on the criteria or its rather worth prioritizing interpretability of topics over the maximization of the model fit.

5.1 Running the base STM models

Before estimating the STM models several data cleaning and filtering steps were made to mitigate the biasing effect due to terms that are irrelevant and meaningless in the ESG framework. Given that the STM model learn on these irrelevant words - because the co-occurrence of these words in specific documents – it ultimately results independently estimated topics containing only these words in the model. These so-called randomized topics are not in the focus of this research, according to which are the critical ESG topics from the improvement or deterioration of the firms ESG score can be predicted. In addition, these distortions are mainly errors from importing the PDF files.

After taking the regular step within text analysis methods – breaking the documents to words (tokenization step) – I carried out the following data cleaning steps. First, words containing any number or just having only one character has been removed from the analysis. Second, non-English alphabetic characters were filtered as in order not to form distinct topics for these non-comparable words. Third, company names, abbreviations and symbols alone do not convey useful information, so I did not take these words into account in the STM model either. Last, I removed the stop words – prepositions, pronouns, conjunctions – from the database that used for the topic model.

One of the key advantages of the STM model is that we can incorporate document or firm-level metadata into the topic analysis. These metadata can be used in either affecting topical prevalence or topical content. In case of metadata variables for topical prevalence the distributions of topics within the documents can vary with respect to the given variable. Using topical content covariates we allow the model to affect the words used within the topics.

In this research I use the year of the ESG reports as a covariate variable affecting the topical prevalence of the STM model. Nevertheless it is worth to note that the year of the ESG report may not necessarily correspond to the year of the ESG report disclosure. In fact, in general, companies always retrospectively evaluate the previous year, not only in financial reports but also in nonfinancial reports. This will be particularly important in the research later on when I estimate the relationship of variables derived from the ESG scores and the topics of the CSR

reports. The STM models with topic number from 2 to 60 were estimated by the following equation:

$$Prevalance = g(Time) \quad (1)$$

, where covariate Time can take a value of the year between 2014 and 2021 and provides information about the CSR activities in that given year.

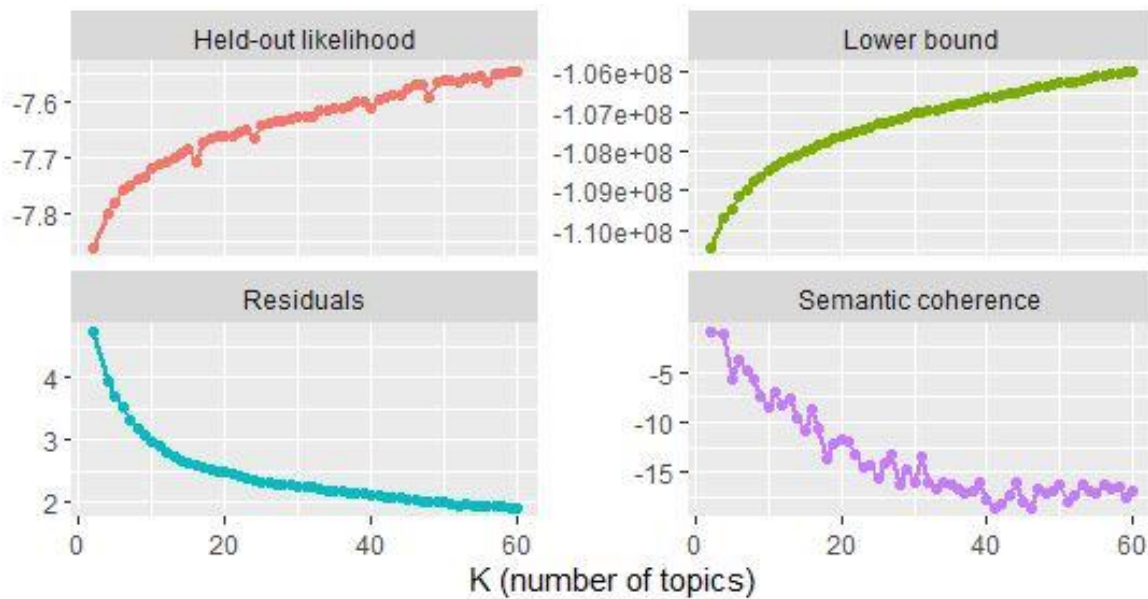
5.2 Finding optimal number of topics

STM assumes each document includes K number of topics with a vocabulary size of V. Similarly to clustering methods, preliminarily in case of the STM model it is uncertain what is the optimal number of topics with which the model user can describe the topical content of a given text. The number of topics, K parameter is one of the most important input parameter which must be defined by the user to run the model. In recent years several model selection criteria and indicators were developed that can serve as a compass to which topic number it is worth to proceed with STM modeling. However, Grimmer and Steward (2013) emphasize that there is not a „right” answer for what is the appropriate number of topics for a document. In practical applications, the intuitive interpretation of results may be preferred over the maximization of the model fit (Bi et al, 2019; Hu et al, 2019). Thus, in my research I also use the „hybrid” approach, that is, in addition taking into account the selection criteria I also place great emphasis on the interpretability of the topics.

Mimno et al (2011) introduce the concept of semantic coherence which is similar to pointwise mutual information (Newman et al, 2010): its value is maximized when in a given topic the most relevant words frequently co-occur together. It is a metric that has a high correlation with human judgement. Roberts et al (2014) present that reaching high semantic coherence is relatively easy when a few topics are dominated by very common terms. Thus, topic quality should be measured as a combination of semantic coherence and exclusivity of words to topics (Roberts et al, 2019). Evaluating topic models can be also conducted by the held-out likelihood (Wallach et al, 2009) calculation and the residual analysis (Taddy, 2012). In case of the held-out likelihood the higher the criteria the model fits better, meanwhile lower residuals in the STM model also indicate a more robust topic model.

After the pre-processing steps on the corpus dataset explained in the previous chapter, I calculated the held-out likelihood, lower bound, residuals and semantic coherence for topics from 2 to 60 to get an overview of model fits. The results of the criteria are presented in Figure 3.

3. Figure: Model diagnostics with different criterias by number of topics



The figure shows the model diagnostics with the calculation of different model criteria. The figure presents four subfigures namely the held-out likelihood, lower bound, residuals and the semantic coherence. In case of all the subfigures the x axis means the k parameter (the number of topics), meanwhile the y axis stands for the different value at the given criteria. Source: Refinitiv Eikon

Held-out likelihood constantly increasing with higher number of topics, meanwhile residuals also show a similar conclusion: with increasing number of topics the residuals are gradually decreasing. Nevertheless, the slope of the curves in both cases points to a lower improving robustness in terms of marginality. Based on these two criterias, choosing a topic with the highest possible number of topics would be the optimal choice.

As a next step I examine the trade-off between the semantic coherence and exclusivity of words to topics. Roberts et al (2014) argue that a semantically interpretable topic disposes two qualities. First, it is cohesive in the sense that major terms in the topic usually co-occur together. Second, it is exclusive in the sense that these high-probability words are unique in a given text, meaning they do not appear as top words in other topics.

Even though the majority of the model selection criteria indicate the higher the topic number we choose the more robust model we end up with, other practical aspects have to be considered during the topic modelling. This is particularly important in those cases when the final goal is to interpret and estimate the effect of certain covariates on the topical distributions. Under this investigation the aim is that how we can still clearly identify certain topics without the most important words in the topic do not appear in other topics in the model (exclusivity of words to topics). To discover what is the number of topics when the models still define well distinguished topics, I present the topic distribution and topic-term distributions in models with 5 and 10 topics.

5. Table: Exploring document topic proportions and topic-term distributions in STM models with 5 and 10 topics

Model with 5 topics		
Topic number	Topic share	Top words
2	23.1%	business, health, employees, program, people, corporate, work
3	22.1%	sustainability, products, water, waste, safety, global, food
1	17.6%	report, management, business, gri, employees, risk, governance
5	17.3%	safety, water, management, emissions, employees, sustainability, report
4	15.4%	energy, customers, company, gas, power, electric, pg
Model with 10 topics		
Topic number	Topic share	Top words
4	20.0%	business, employees, corporate, financial, report, risk, communities
1	17.0%	report, management, employees, gri, business, sustainability, energy
6	13.7%	global, sustainability, business, safety, energy, employees, emissions
8	10.4%	energy, customers, gas, employees, safety, pg, power
3	8.7%	water, sustainability, waste, products, global, sustainable, people
5	8.1	water, safety, emissions, management, gas, operations, performance
9	7.1	food, u.s, products, safety, percent, global, report
10	6.4	energy, water, sustainability, report, environmental, management, waste
2	6.2	health, care, program, patients, healthcare, employees, quality
7	2.4	financial, company, net, million, value, mellon, may

The table demonstrates two different STM estimations. The first part of the table shows the STM model with 5 topics, meanwhile the second part is the output from an estimation with 10 topics. In both cases the prevalence variable for the STM model was the disclosure year of the CSR report. Each row represents different topic number within the STM model. The first column defines the topics number, the second column presents topic proportion within the given topic model. The last column in each model shows the top words within the topics. Source: Refinitiv Eikon

Examining the top words in the topics in case of the first STM model, clearly identifiable topics emerge which we can often see within the ESG framework. Topic number 2 deals with questions related to workforce and community and can be classified to the social (“S”) pillar. Topic 3 and 4 addressing environmental (E”) issues and focus on the most frequently discussed topics in the field of ESG: resource usage and emission related considerations. Exploring the top words within topic 1, reporting processes (‘gri’ – Global Reporting Initiative, ‘report’) and governance (‘management’, ‘governance’) receive significant emphasis within the topic. In terms of topic proportions, we cannot clearly see that certain topics dominate a large part of the documents. In other words the topics distributions are

roughly equal with topic 2 has the largest share with 23.1%, while topic 2 makes up roughly 15 percent of the documents.

Exploring the STM models with 10 topics we obtain a quite different picture of both the distinguishableness of the topics and their proportionality. Although clear topics still can be recognized at the model with 10 topics – for instance the topic number 8 and number 2 which deal with environmental and social related considerations – significant overlap appears among the topics. The term of ‘employees’ and ‘water’ exist in five and three of the topics respectively. The absence of exclusivity may stem from the overall homogeneity of the ESG related topics. As for the topic distributions in case of the second STM model, a more unbalanced topic proportions can be seen as three topics dominate more than half of the documents, while the sum of topic number 10, 2 and 7 explains 15% of the CSR reports.

Given the limited heterogeneity in the topics and the significant overlap in top words already at the STM model with 10 topics at I will only estimate the relationship of the covariates and the topics only up to the 12-topic STM model.

6. Empirical results

In this section I present the main results of the research by estimating the relationship between the metadata and the topics. First, I show all the estimated effects across different covariates with topic models from 3 to 12. Second, I focus on the statistically significant topics by exploring the content of these topics.

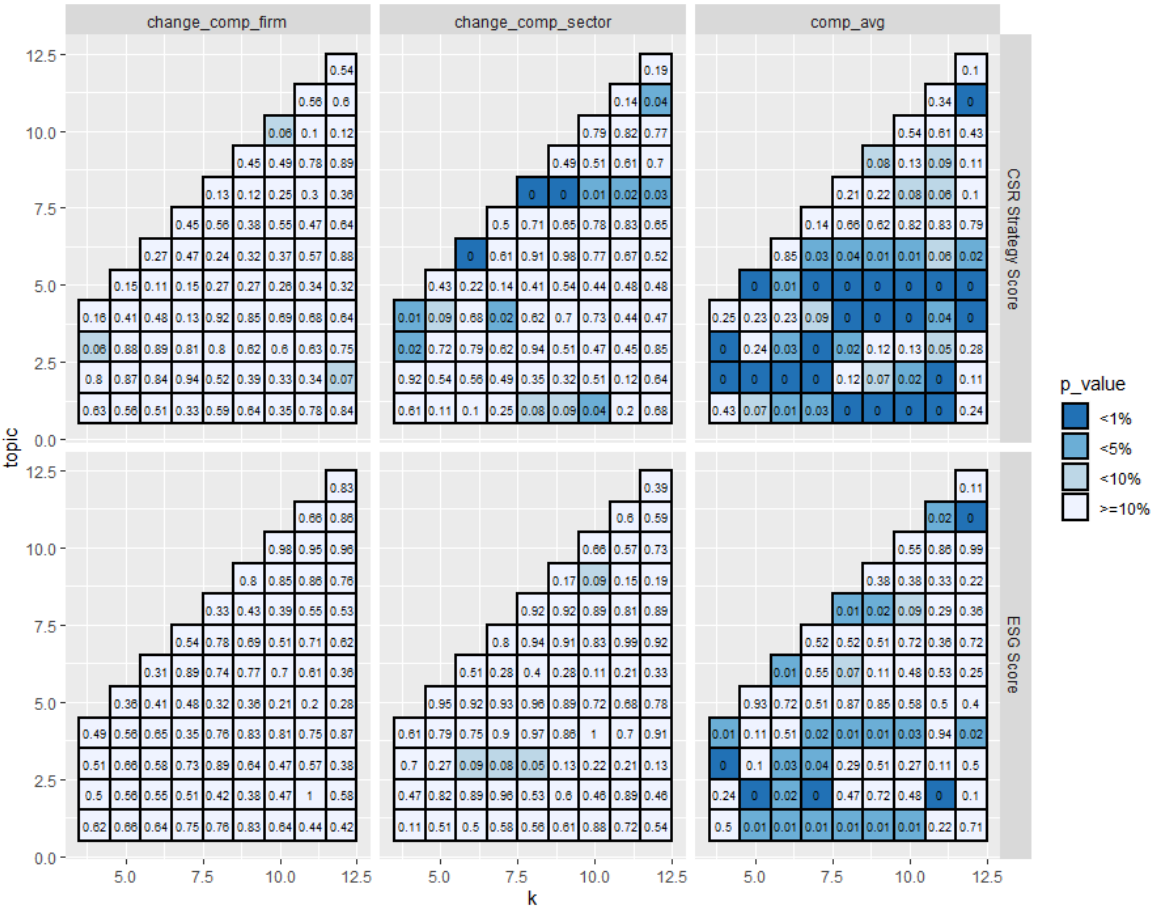
6.1 Estimating the relationship between the proportion of topics and covariates

After defining the categorical covariates from the numerical ESG related data obtained from the Refinitiv, I estimate the impact of covariates on the topic proportions. To explore the relationship between metadata and topics model users can validate the usefulness of the topic models (Grimmer, 2010; Grimmer and Stewart, 2013). In general, in the STM model we estimate regressions for the proportion of each document – in my case the yearly CSR reports – about topic with predefined metadata as the independent variables. Accordingly, we test whether the proportion of a given topic significantly differs across the defined groups. In addition to the previously presented ESG score related covariates, I also incorporate the year of the report as an additional independent variable to the regression. The variable time also carry relevant information with respect to the communication practices of companies. For instance, Chae and Park (2018) find that the proportion of topics related to employee

engagement and community charity have grown, while topics discussing health, tobacco and philanthropy have declined.

In the figures the p values from the regressions are presented. In the horizontal dimension the ESG related scores, while on the vertical axis predictors calculated from these variables can be seen. For the sake of interpretability and comprehensibility, I divided the 6 types of indicators into 3 parts. Starting from the broader indicators, I proceed towards the sub-categories and examine where the proportion of the topic differs between the groups. First I show the results from the regressions for the CSR Strategy and composite ESG scores for STM models from 4 topics up to 12 topics.

4. Figure: The p values from the regressions estimating the topic proportions for different groups in case of CSR Strategy and ESG scores



The figure presents the p values from the regressions estimating the topic proportions as dependent variable at different topic numbers with two types of covariates: CSR Strategy Score and ESG score. In the horizontal dimension the covariates, while on the vertical axis, predictors calculated from these variables can be seen. The numbers in the squares show the p values estimated from the particular regression. The darker colour the given square, the smaller the p-value is. Accordingly, the darker colour also indicates that on a smaller significance level the given topic proportion differs between the groups calculated from the given covariate. Source: Refinitiv Eikon

As previously explained in detail, both the CSR Strategy Score and the composite ESG score capture indicators related to the entire ESG framework, but there may be still significant

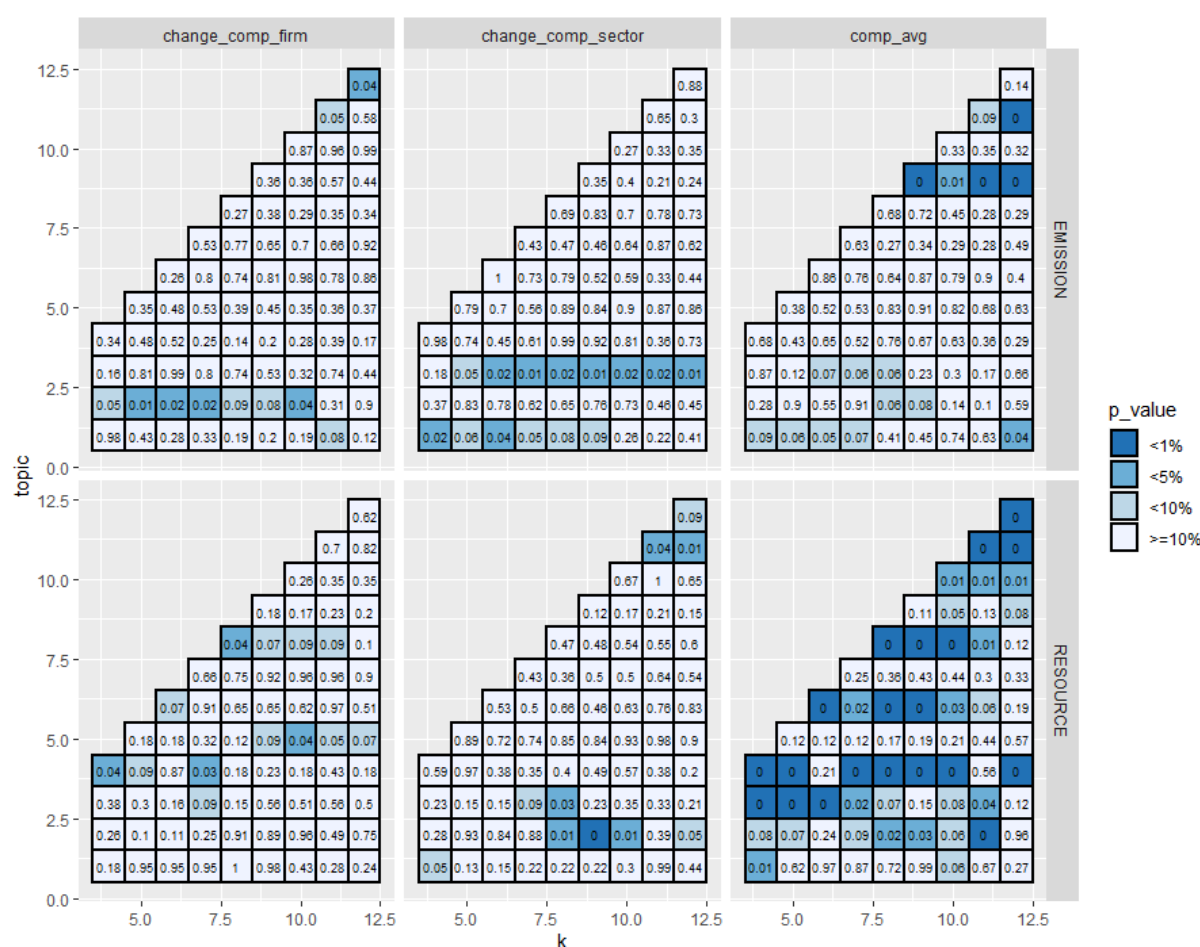
differences stemming from the methodology. First, the CSR Strategy score focus on the qualitative - communication of ESG initiatives – aspects, while the ESG score is a mixture of qualitative and quantitative aspects. Second, there are also differences in the source of the information from which they are evaluating the overall score: the ESG score is calculated from the self-reported information, while in case of the CSR strategy score the Refinitiv does not explicitly highlight the source of information.

Looking at the p values for the `change_comp_firm` variable– which measures whether the firm improved its score compared to the previous year – significant topics have not been estimated on a 5% significance level. Based on these results, it can be concluded that the CSR Strategy score and ESG score indicators do not necessarily result in a significant improvement or deterioration compared to the firm’s previous year's value in a given score, depending on the emphasis of certain topics in the CSR documents.

With respect to the indicator compared the performance to the sector (`change_comp_sector`) in a given year, the proportion of topics are significantly differ in the CSR Strategy Score across the groups. For instance, the share of topic 8 in the documents for the topics models with 8, 9, 10, 11, and 12 topics significantly different for those firms whose annual CSR Strategy Score change is better or worse compared to the average change in the sector.

Analysing the last column – where the regression results for the covariate firms with better than average or worse than average scores for a given year are presented – several significant topics emerge from the estimated models. Within this variable the S&P500 firms usually do not emigrate from one group to another, therefore this indicator is not necessarily capable of predicting improvement deterioration from one year to the next year. Nevertheless, it is able to show which topics are significantly more emphasized by the generally well-performing companies (above average value) in a given indicator. In general, the proportion of the first topic in the STM models from 4 topics to 8 topics significantly differs across firms with having above and below average CSR Strategy scores. The share of topics differs more significantly with respect to the CSR Strategy Score, which is compatible with prior intuition as this indicator includes more qualitative information compared to the ESG score.

5. Figure: The p values from the regressions estimating the topic proportions for different groups in case of Environmental pillar related variables

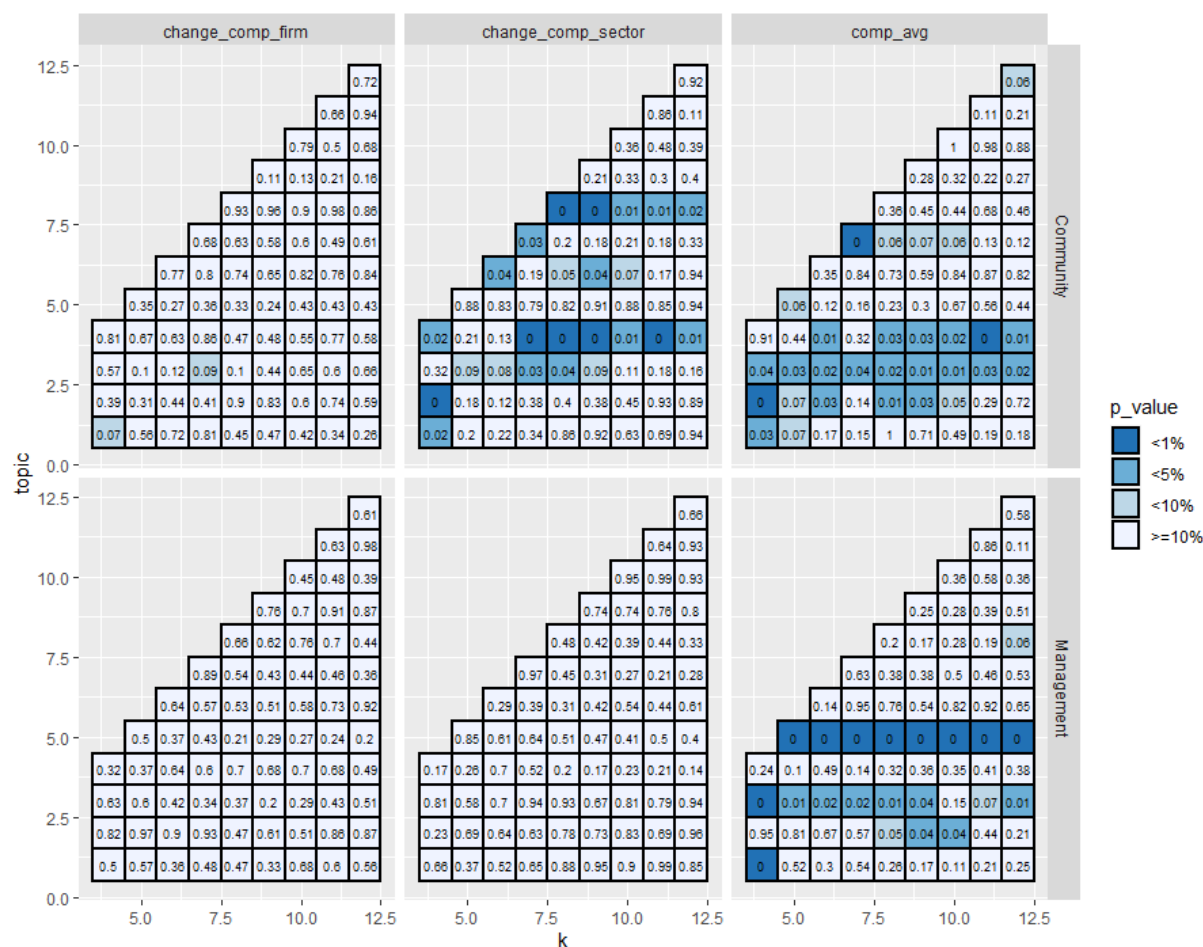


The figure presents the p values from the regressions estimating the topic proportions as dependent variable at different topic numbers with two types of covariates: Emission Score and Resource Score. In the horizontal dimension the covariates, while on the vertical axis, predictors calculated from these variables can be seen. The numbers in the squares show the p values estimated from the particular regression. The darker colour the given square, the smaller the p-value is. Accordingly, the darker colour also indicates that on a smaller significance level the given topic proportion differs between the groups calculated from the given covariate. Source: Refinitiv Eikon

Focusing only on the environmental pillar of the Eikon ESG framework, more significant topics appear in the YoY predictors compared to the composite CSR and ESG scores. Based on the estimations, varying degrees of emphasis topic number 2 in the yearly CSR reports on average may indicate a notable difference in the Emission Score value compared to the previous year's value. Remaining at the evaluation of the emission scores, the different proportion of the third topic can cause a greater difference in the score compared to the average change in the sector the firm operates in.

Highlighting several topics at a greater extent may contribute predicting more accurately the improvement or deterioration in the Resource Usage Score compared to the previous year. For instance, testing the equality of topic proportions in the documents at topic 2 in case of models with 8, 9 and 10 topics, statistically significant differences were estimated across firms separated by the performance with respect to the sector average.

6. Figure: The p values from the regressions estimating the topic proportions for different groups in case of Community ('S' pillar) and Management ('G' pillar) scores



The figure presents the p values from the regressions estimating the topic proportions as dependent variable at different topic numbers with two types of covariates: Community Score and Management score. In the horizontal dimension the covariates, while on the vertical axis, predictors calculated from these variables can be seen. The numbers in the squares show the p values estimated from the particular regression. The darker colour the given square, the smaller the p-value is. Accordingly, the darker colour also indicates that on a smaller significance level the given topic proportion differs between the groups calculated from the given covariate. Source: Refinitiv Eikon

Exploring the regressions for Community and Management Scores related to the social and governance pillars respectively, the share of topics mostly diverge in predictors derived from former indicator. In fact, based on the estimates, it seems that the predictor created based on the performance compared to the sector average explains effectively the difference in the emphasis of the topics. The proportion of several topics is statistically different between the two groups. For instance, by exploring topic number 4 and topic number 8 it is possible to understand the reason behind the better or worse performance in Community Score.

6.2 Identify and explore the statistically significant topics

In the previous chapter the p values for STM models with ESG-related predictors were explained. Out of the three types of covariates, two covariates – Change_comp_firm and change_comp_sector – can be considered as predictive variables for predicting the future performance of firms in ESG variables. The third covariate – Comp_avg, which aligns the firms based on a score higher or lower than the yearly average in the sample – is strongly related to the development of certain topics, and since there is no strong migration of companies jumping from a below-average group to an above-average group from one year to the next, prediction for the future performance in ESG related indicators from this variable alone is not possible. Thus in the following, I focus on estimating the significant topics for the first two covariates. I evaluate the significant topics along two dimensions: the sign of the effect (in which group the given topic is more emphasized compared to other topics) and the identified ESG-related topic.

6. Table: Results for the statistically significant topics in case of the ESG and CSR Strategy scores

Indicator		CSR Strategy Score	
Model		Model 1	Model 2
STM topics	Dependent variable	Worsened YoY	Worse than sector
12	Topic 11: Management	0.007 (-0.13)	0.027** (2.06)
8-12	Topic 8: Fossil fuels	-0.026 (-1.52)	-0.046*** (-2.86)
6	Topic 6: Greenhouse gas emissions	-0.020 (-1.11)	-0.050*** (-2.87)
4	Topic 3: Management & Reporting	0.039* (1.86)	0.048** (2.43)
4	Topic 4: Greenhouse gas emissions	-0.024 (-1.39)	-0.039** (-2.47)
Indicator		ESG Score	
Model		Model 1	Model 2
STM topics	Dependent variable	Worsened YoY	Worse than sector
8	Topic 3: Food waste & Supply chain	-0.002 (-0.13)	-0.030** (-1.99)

The table 6 shows the regression results for the statistically significant topics with predictors derived from the ESG Score. The dependent variable is the average topic proportion in the given group. Thus, the dependent variable captures to which extent firms discuss a topic in their yearly CSR report compared to other ESG-related information. Model 1 is estimated with the predictor measuring whether the firm improved its score in a given dimension compared to the previous year (YoY change). The independent variable for model 2 is assessing whether the firm improved its score in a given category to a greater extent compared to the sectorial average change (YoY change). When the dependent variable refers to multiple STM topics, the beta, significance level and the t statistic values correspond to the STM model with smallest number of topics. The t statistics are presented in the parenthesis. Significance levels are determined as follows: * p<.10, ** p<.05, *** p<.01. Source: Refinitiv Eikon

Analysing the regression results of table 6, I can draw the following conclusions. First, in the case of CSR Strategy scores, compared to the ESG scores, the proportion of more topics differs significantly between the groups. All of this corresponds to preliminary intuition, since former indicator takes into account purely qualitative (textual) aspects, while the latter includes both hard and soft information. Second, the predictors derived from the performance compared to the sector explain more strongly the difference in the distribution of topics both in the ESG and CSR Strategy indicators. Third, 5 percentage point lower share of the greenhouse gas emissions topic relative to other ESG-related topics in the CSR disclosures results worse performance in the CSR Strategy Score in the following year. The lower proportion of topics discussing fossil fuels, while the higher proportion of topics dealing with management and reporting practices also indicate an unfavourable CSR Strategy score compared to the sector.

Looking at the results for the regression with predictors derived from the composite ESG score, the proportion of topic 3 significantly differs between the firms separated by the performance relative to the sector. Putting the emphasis on topics assessing food waste and supply chain issues on a 3 percentage point lesser extent compared to other sustainability related topics results a worse performance compared to the sector after the disclosure of the CSR report. This leads to the conclusion that food waste is a critical topic within ESG framework, to which companies must allocate significant resources.

7. Table: Results for the statistically significant topics in case of the Emission Score

Indicator		Emission Score	
Model		Model 1	Model 2
STM topics	Dependent variable	Worsened YoY	Worse than sector
5-10	Topic 2: Health & Safety	-0.044** (-2.51)	-0.004 (-0.21)
6-12	Topic 3: Waste management	-0.002 (-0.01)	-0.043 ** (-2.43)
6-7	Topic 1: Governance	0.017 (1.08)	0.032** (2.06)

The table 7 shows the regression results for the statistically significant topics with predictors derived from the Emission Score. The dependent variable is the average topic proportion in the given group. Thus, the dependent variable captures to which extent firms discuss a topic in their yearly CSR report compared to other ESG-related information. Model 1 is estimated with the predictor measuring whether the firm improved its score in a given dimension compared to the previous year (YoY change). The independent variable for model 2 is assessing whether the firm improved its score in a given category to a greater extent compared to the sectorial average change (YoY change). When the dependent variable refers to multiple STM topics, the beta, significance level and the t statistic values correspond to the STM model with smallest number of topics. The t statistics are presented in the parenthesis. Significance levels are determined as follows: * p<.10, ** p<.05, *** p<.01. Source: Refinitiv Eikon

In terms of the Model 1 for the Emission Scores, the proportion of topic 2 – discussing health and safety related issues – is 4.4 percentage point lower in the CSR documents for firms

whose Emission Score worsened compared to the previous year's Emission Score value. As for the model 2, where the given score change is compared to average change in the sector, more intuitive results can be obtained. Firms performing worse in Emission Score compared to the sector average in relation to the previous year, focus to a lesser extent (4.3% percentage point lower share of topic) on waste management topics in the yearly CSR reports. This allows us to conclude that the emphasis on waste management topic (qualitative aspect) in the CSR reports will actually affect the next year's Emission Score value, which, by the way, makes up a significant part of the environmental pillar of the ESG framework defined by the Refinitiv. The share of governance related topic is significantly higher in case of the group who worsened the Emission Score in relation to the average of the sector. This indicates the fact that there can be a correlation between topics with different pillars.

8. Table: Results for the statistically significant topics in case of the Resource Usage Score

Indicator		Resource Usage Score	
Model		Model 1	Model 2
STM topics	Dependent variable	Worsened YoY	Worse than sector
8-11	Topic 8: Fossil fuels	0.031** (2.01)	-0.014 (-0.72)
10-12	Topic 5: Greenhouse gas emission	0.029** (2.07)	-0.001 (-0.09)
12	Topic 11: Governance	-0.003 (-0.22)	0.031** (2.47)
8-10	Topic 2: Health & Safety	0.001 (0.11)	0.032*** (2.65)
8	Topic 3: Food waste & Supply chain	-0.022 (-1.44)	-0.035** (-2.24)

The table 8 shows the regression results for the statistically significant topics with predictors derived from the Resource Usage Score. The dependent variable is the average topic proportion in the given group. Thus, the dependent variable captures to which extent firms discuss a topic in their yearly CSR report compared to other ESG-related information. Model 1 is estimated with the predictor measuring whether the firm improved its score in a given dimension compared to the previous year (YoY change). The independent variable for model 2 is assessing whether the firm improved its score in a given category to a greater extent compared to the sectorial average change (YoY change). When the dependent variable refers to multiple STM topics, the beta, significance level and the t statistic values correspond to the STM model with smallest number of topics. The t statistics are presented in the parenthesis. Significance levels are determined as follows: * $p < .10$, ** $p < .05$, *** $p < .01$. Source: Refinitiv Eikon

Looking at the results of the Model 1, the proportions of topic 8 and topic 5 – which discussing fossil fuels and greenhouse gas emission related topics – are 3.1 and 2.9 percentage points larger respectively at those companies whose Resource Usage Score decreased compared to the previous year's value. Firms discussing non-sustainable and fossil energy sources in a higher proportion perform worse in Resource Usage Score in an annual comparison.

The higher proportion of governance and health & safety related topics contribute to a worse performance in Resource Usage Score compared to the sector, while firms who put less attention to food waste and supply chain issues also face a lower Environmental Pillar score by deteriorating the resource-related category score.

9. Table: Results for the statistically significant topics in case of the Community Score

Indicator		Community Score	
Model		Model 1	Model 2
STM topics	Dependent variable	Worsened YoY	Worse than sector
4	Topic 1: Food Waste & Food safety	-0.032* (-1.79)	-0.042** (-2.27)
4	Topic 2: Management & Business	0.017 (0.86)	0.058*** (2.84)
4	Topic 4: Greenhouse gas emission	0.004 (0.24)	-0.036** (-2.28)
7-8	Topic 3: Food waste & Supply chain	-0.27* (1.70)	-0.036** (-2.20)
7-12	Topic 4: Financial risk & Community	0.003 (0.86)	-0.044*** (-2.84)
8-12	Topic 8: Fossil fuels	-0.001 (0.93)	-0.049*** (-2.99)

The table 9 shows the regression results for the statistically significant topics with predictors derived from the Community Score. The dependent variable is the average topic proportion in the given group. Thus, the dependent variable captures to which extent firms discuss a topic in their yearly CSR report compared to other ESG-related information. Model 1 is estimated with the predictor measuring whether the firm improved its score in a given dimension compared to the previous year (YoY change). The independent variable for model 2 is assessing whether the firm improved its score in a given category to a greater extent compared to the sectorial average change (YoY change). When the dependent variable refers to multiple STM topics, the beta, significance level and the t statistic values correspond to the STM model with smallest number of topics. The t statistics are presented in the parenthesis. Significance levels are determined as follows: * p<.10, ** p<.05, *** p<.01. Source: Refinitiv Eikon

Exploring the regressions results for the Community Score, several topics emerge which can explain the score after the disclosure of the CSR reports. Within the models, the topic proportions differ significantly in case of the Model 2. Topic dealing with management and business by a larger extent (5.8 percentage points) corresponds to a worse performance compared to the sector. This indicates that prioritizing governance aspects over social aspects can worsen the evaluation of the latter. Meanwhile, the lower proportion of food waste, greenhouse gas emission, supply chain, fossil fuels, financial risk and community related topics can contribute to the unfavourable performance compared to the sector.

7. Conclusion

Climate change is an increasingly growing issue today, so there is increasing focus on firms to incorporate these aspects into their daily operations. Both the regulatory and investor sides pay considerable attention to the development of companies' ESG score, which is a framework created for the comparability of environmental, social and governance performance.

In this research I estimate whether it is possible to predict the improvement or deterioration of ESG scores from the proportion of ESG related topics in the annual CSR report disclosures of S&P500 firms. By estimating the predictive power of the textual documents of firms, I explore whether the firms strongly emphasizing a given sustainability topic are very serious about dealing with the topic or rather they thought it was serious only in the sense of communication channels and did not necessarily take real steps to solve the issue. Since the ESG framework takes into account qualitative aspects in addition to quantitative indicators, their understanding can be therefore crucial in capturing as accurate as possible the ESG scores.

Regarding the explanatory variables in the STM model, I construct three types of binary predictors, which separate the outperforming or underperforming firms in certain ESG-related indicators. Out of three indicators I examine two in detail – own performance compared to the previous year and performance in relation to the sector average compared to the previous year – which can have predictive capabilities, as they try to explain whether there is a difference in the given ESG score after the disclosure of the annual text document. As for the CSR documents, with web scraping technique I retrieve all the possible S&P500 firms' disclosures from 2010-2021.

My research suggests that the CSR report disclosure practice of firms is effective, meaning that their topical content has an impact on the ESG scores. The greater proportion of certain topics is indeed reflected in the ESG scores, suggesting that external rating companies also evaluate the efforts positively or negatively depending on the topic content. Also a significant contribution of this research to the ESG field is that very crucial topics emerge even within the framework. Assessing topics related to food waste, supply chain, waste management and greenhouse gas emissions significantly contribute for the development of the ESG and its subcategories scores. Indeed, these topics prove to be so important that they can affect the scores of other categories in the ESG framework as well.

Literature

- Abhayawansa, S., & Tyagi, S. (2021). Sustainable investing: The black box of environmental, social, and governance (ESG) ratings. *The Journal of Wealth Management*, 24(1), 49-54.
- Albuquerque, R., Koskinen, Y., & Zhang, C. (2019). Corporate social responsibility and firm risk: Theory and empirical evidence. *Management Science*, 65(10), 4451-4469.
- Baier, P., Berninger, M., & Kiesel, F. (2020). Environmental, social and governance reporting in annual reports: A textual analysis. *Financial Markets, Institutions & Instruments*, 29(3), 93-118.
- Bassen, A., Meyer, K., & Schlange, J. (2006). The influence of corporate responsibility on the cost of capital. Available at SSRN 984406.
- Beattie, V. (2014). Accounting narratives and the narrative turn in accounting research: Issues, theory, methodology, methods and a research framework. *The British Accounting Review*, 46(2), 111-134.
- Berg, F., Koelbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6), 1315-1344.
- Bi, J. W., Liu, Y., Fan, Z. P., & Zhang, J. (2019). Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews. *Tourism Management*, 70, 460-478.
- Billio, M., Costola, M., Hristova, I., Latino, C., & Pelizzon, L. (2021). Inside the ESG ratings:(Dis) agreement and performance. *Corporate Social Responsibility and Environmental Management*, 28(5), 1426-1445.
- Chan, M. C., Watson, J., & Woodliff, D. (2014). Corporate governance quality and CSR disclosures. *Journal of business ethics*, 125, 59-73.
- Chae, B., & Park, E. (2018). Corporate social responsibility (CSR): A survey of topics and trends using Twitter data and topic modeling. *Sustainability*, 10(7), 2231.
- De Lucia, C., P. Paziienza, and M. Bartlett. 2020. "Does Good ESG Lead to Better Financial Performances by Firms? Machine Learning and Logistic Regression Models of Public Enterprises in Europe." *Sustainability* 12 (13).
- Engelhardt, N., Krause, M., Neukirchen, D., & Posch, P. N. (2021). Trust and stock market volatility during the COVID-19 crisis. *Finance Research Letters*, 38, 101873.
- Eliwa, Y., A. Aboud, and A. Saleh. 2019. "ESG Practices and the Cost of Debt: Evidence from EU Countries." *Critical Perspectives on Accounting* 79: 102097.
- Friedman, M. (2007). *The social responsibility of business is to increase its profits* (pp. 173-178). springer berlin heidelberg.
- Grimmer, J. (2010). A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in Senate press releases. *Political Analysis*, 18(1), 1-35.
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3), 267-297.

Goloshchapova, I., Poon, S. H., Pritchard, M., & Reed, P. (2019). Corporate social responsibility reports: topic analysis and big data approach. *The European Journal of Finance*, 25(17), 1637-1654.

Huber, D. G., & Gulledge, J. (2011). *Extreme weather and climate change: Understanding the link, managing the risk*. Arlington: Pew Center on Global Climate Change.

Ho, J. C., Chen, T. H., & Wu, J. J. (2022). Are corporate social responsibility reports informative? Evidence from textual analysis of banks in China. *China Finance Review International*, 12(1), 101-120.

Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? Text analysis using structural topic model. *Tourism Management*, 72, 417-426.

<https://www.ifrs.org/news-and-events/news/2021/11/ifrs-foundation-announces-issb-consolidation-with-cdsb-vrf-publication-of-prototypes/>

Jaworska, S., & Nanda, A. (2018). Doing well by talking good: A topic modelling-assisted discourse study of corporate social responsibility. *Applied Linguistics*, 39(3), 373-399.

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

Larcker, D. F., Pomorski, L., Tayan, B., & Watts, E. M. (2022). ESG ratings: A compass without direction. Rock Center for Corporate Governance at Stanford University Working Paper Forthcoming.

Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. *American political science review*, 97(2), 311-331.

Lins, K. V., Servaes, H., & Tamayo, A. (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *the Journal of Finance*, 72(4), 1785-1824.

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011, July). Optimizing semantic coherence in topic models. In *Proceedings of the 2011 conference on empirical methods in natural language processing* (pp. 262-272).

Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010, June). Automatic evaluation of topic coherence. In *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics* (pp. 100-108).

Ning, X., Yim, D., & Khuntia, J. (2021). Online sustainability reporting and firm performance: Lessons learned from text mining. *Sustainability*, 13(3), 1069.

Pástor, L., Stambaugh, R. F., & Taylor, L. A. (2022). Dissecting green returns. *Journal of Financial Economics*, 146(2), 403-424.

Renneboog, L., Ter Horst, J., & Zhang, C. (2008). Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of banking & finance*, 32(9), 1723-1742.

- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., ... & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American journal of political science*, 58(4), 1064-1082.
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software*, 91, 1-40.
- Rouen, E., Sachdeva, K., & Yoon, A. (2022). The Evolution of ESG Reports and the Role of Voluntary Standards. Available at SSRN 4227934.
- Stott, P. (2016). How climate change affects extreme weather events. *Science*, 352(6293), 1517-1518.
- Talan, G., & Sharma, G. D. (2019). Doing well by doing good: A systematic review and research agenda for sustainable investment. *Sustainability*, 11(2), 353.
- Tregidga, H., Milne, M. J., & Kearins, K. (2007). The role of discourse in bridging the text and context of corporate social and environmental reporting. Department of Accountancy, Finance and Information Systems, University of Canterbury.
- UN. (2004). Who cares wins: Connecting financial markets to a changing world. Tech. rep., United Nations Global Compact.
- Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009, June). Evaluation methods for topic models. In *Proceedings of the 26th annual international conference on machine learning* (pp. 1105-1112).