Norwegian School of Economics

Bergen, Spring 2021

Explicitness of 10-K reporting on CSR-related topics and board composition.

Aleksandr Narutto and Mostafa Raziani

Supervisor: Maximilian Rohrer

Master thesis, Economics and Business Administration Major:

Finance / Business Analytics

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgements

This thesis is written as part of our Master of Science in Economics and Business Administration at NHH, with majors in Business Analytics and Financial Economics.

We would like to thank our supervisor, Maximilian Rohrer for his support, for valuable feedback and guidance throughout the semester. We are grateful that his door have always been open for our questions, and for his constructive feedback throughout our project.

We would also like to thank the Norwegian School of Economics for providing an extensive and interesting Master of Science program.

Keywords – NHH, textual analysis, CSR, board characteristics, corporate governance

Executive summary

This paper examines the relationship between board composition characteristics, particularly board gender diversity, size, and independence level on one side, and the level of explicitness in reporting related to corporate social responsibility (CSR) on the other. For a sample of 2398 10-K forms of US firms, we conduct a textual analysis on annual reports published between April 1st, 2020, and March 31st, 2021. We identify important CSR-related topics and establish dictionaries of most indicative words for them. We then process texts with the reports to construct the measure that defines the level of explicitness. Our results confirm that companies with more women on the board tend to disclose more explicitly on CSR-related topics in general annual reports. However, board size is not significant in predicting the same measure. When controlling for additional factors, we confirm that independence level of the board has no correlation with explicitness level. Finally, we suggest some important avenues for future research in this field.

Contents

List of Figures

List of Tables

Introduction

Change in societies and culture of accountability for company's activities in various fields such as environmental and philanthropy-related issues led to the conceptualization of corporate social responsibility (CSR). Prior research focuses on the relationships between firms' board composition characteristics, such as board independence and board gender diversity on the one side, and CSR performance on the other, but there is a lack of research on how board composition affects CSR disclosure (i.e., explicitness of disclosure). Aguilera et al. (2006) mention that it is the responsibility of a board of directors to create an internal drive towards CSR activities and subsequent reporting, which lies in line with the implementation of basic mechanism of corporate governance. Thus, they argue that the firms' board of directors is the body responsible for firms' CSR performance and disclosure. A documented underlying mechanism (Rao & Tilt, 2016) points out that that the board structure has the potential to influence corporate decisions around development of socially and environmentally responsible policies and involvement CSR practices. Such influence can give motives to companies to reflect these activities in their disclosure. Thus, we expect that the board structure is related to the way companies report about their activities.

Apart from examining the effect of board independence on the level of explicitness, we follow the lead of prior literature by investigating whether other prominent board characteristics have a significant effect on the dependent variable that is central to this study. We additionally explore the relationship between size of the board and gender diversity level and the explicitness of CSR-related disclosure, using data for 2398 listed companies in the United States.

Since all U.S. publicly traded firms have to provide annual reports of the company's business and financial status to the Securities and Exchange Commission (SEC) on Form 10-K, it makes the reports an important tool for implementation of studies where companies stand in comparable frames.

According to Matten & Moon (2008), the explicitness of CSR disclosure measures the proportion of CSR disclosure that conveys specific claims to readers—more explicit CSR disclosures often include more useful information than less explicit ones within the same number of words. The explicitness of CSR disclosure can be considered as a measure of CSR

report's quality, meaning that CSR reports with higher explicitness level can provide more useful information for a broad range of stakeholders.

In this paper, we follow the methodology developed by Hummel et al. (2019), which is one of the most recent investigations on the explicitness level of CSR disclosure. In their study, researchers measure explicitness of CSR disclosure by the level of similarity between CSR reports and a pre-defined set of topic-related texts that contain the most relevant CSR terms. The higher the similarity level of CSR reports to the topic text, the more explicit it is.

According to Hummel et al. (2019), the explicitness of CSR can be a measure of CSR report's textual quality, hence we are interested in whether there are relationships between this textual characteristic and the board composition.

This paper deepens the research on characteristics of CSR-related disclosure. We analyze potential drivers that affect the explicitness level of reporting. The results of the study are mixed in relation to our hypotheses. When it comes to the explicitness level in corporate reporting on CSR-related topics, we find a positive significant relationship with "gender diversity", while considering some controlling variables like market capitalization, pretax ROA, and GRI reporting. The setup includes 9 models in 7 of which the measure is significantly positively connected with explicitness with parameter estimates between 0,2 and 0,43 depending on the model. Therefore, our result lets us admit the second hypothesis.

At the same time, we find evidence for non-correlation of explicitness level with size and level of independence of the board. In the same setup, we additionally control for board committees and some evidence of non-linear relationship when examining the connection of dependent variable with the level of independence of the board. We adapt categorization of the board that generalizes level of independence into wider definitions and assign companies in one of 4 categories, based on the methodology that is proposed by Bhagat & Black (2001) in a parallel research line. Initial results persist.

Literature Review

In this section, we start with the anticipated contribution and continue providing background on Corporate Social Responsibility systems and 3 elements of board composition: size, gender diversity, and independence. We further show in each respective subsection evidence for the hypotheses that form the focus of this study.

Background

Our paper relates to a research area that overviews the potential effect of board composition on CSR reporting. We try to detect if any of the characteristics of board composition can be considered a predictor for a more specific and explicit reporting on CSR activities.

Over the prior two decades, the rising awareness of environmental and social issues has pushed companies to engage in environmentally and socially responsible practices (Agnolucci & Arvanitopoulos, 2019 ; Alam et al., 2019) and the image of the large companies often depends on their performance regarding corporate social responsibility. CSR is an organizations' broader responsibility towards society.

Furthermore, Responsible stakeholders substantially increase the pressure for business organizations to direct their attention to explicit CSR policies; stakeholders themselves are guided in their actions by corporate social disclosure (alongside, for instance, media coverage or NGO campaigns) (Höllerer, 2016).

Bénabou & Tirole (2010) show how environmental activists effectively enforce their pro-social preferences on companies. For example, a court in the Netherlands has ruled in a landmark case that the oil giant Shell must cut its CO2 emissions by 45% compared to 2019 levels (BBC, 2021). An environmental group brought the case to the court in 2019, alongside six other bodies and more than 17,000 Dutch citizens. Though the decision only applies in this country, it could have broader effects elsewhere.

Why are there different CSR systems? Matten & Moon (2008) argue that understanding, scope, and content of CSR significantly differ between governance and business systems that diverge, i.e., between liberal market economies (the United States) and coordinated market economies (various continental European countries). While LME cultural setting provides greater

opportunity and incentives for business to address responsibility through explicit CSR policies, rhetoric, and action, CME represents a system of wider organizational responsibility (embedded in broader norms and regulation), yielding comparatively narrow opportunities and incentives for business to take explicit responsibility (Höllerer, 2016). Matten & Moon (2008) differentiate CSR reporting styles into either explicit, where companies explicitly formulate corporate policies, or implicit, where companies assume institutional frameworks implicitly. They provide a set of characteristics for the both types. Figure 1 provides a comparative overview over the implicit and explicit elements of CSR. Furthermore, Figure 2 illustrates a continuum between explicit and implicit CSR.

According to the previous research, different variables influence CSR disclosure. One of these factors is Corporate Governance. Jamali et al. (2008) assert that corporate governance encourages managers and executives to set CSR mission and vision, and the board plays an important role in achieving and supporting these goals. Given that boards of directors are major players in corporate governance, board composition is likely to influence CSR (Rao & Tilt, 2016). We have not identified studies that take into consideration the explicitness level in corporate reporting on CSR-related topics and board characteristics. Therefore, our paper focuses on 3 elements of board composition (size, gender diversity, and independence) and the possible impact on the explicitness level in corporate reporting on CSR-related topics. The following paragraphs explain previous research on these topics and formulate our expected input by formulating hypotheses.

Board Independence and CSR Reporting

Board members, as organization's key decision-makers, are both accountable and responsible for CSR issues to a broad range of stakeholders. Rao & Tilt (2016) argue that the board structure has the potential to influence corporate decisions around development of socially and environmentally responsible policies and involvement in CSR practices. According to the agency theory proposed by Jensen & Meckling (1976), board independence and board diversity lead to better monitoring of corporate management which encourages companies to be more focused on and concerned about the needs and expectations of various stakeholders (Ibrahim Hanefah, 2016). According to Herda et al. (2012), board independence contributes to implementing CSR reporting in 500 largest US companies.

An increase in the number of independent directors on a board strengthens the need for voluntary disclosure to shareholders which leads to improved surveillance (Mulcahy & Donnelly, 2008). Rouf & Hossan (2020) find in the study that covers all listed banks on the Dhaka Stock Exchange that the proportion of independent directors has a significant relationship with the CSR disclosure in the annual report by the listed banks in Bangladesh. Ahmad et al. (2017) find that by disclosing more details about CSR activities in annual reports, companies achieve multiple effects: asymmetric information reduction, and the company's reputation/protection enhancement. So, an efficient board puts as a priority improvement and advancements in CSR activities. The board's decisions can adjust managers' interests with the long-term goals of both shareholders and non-shareholding stakeholders. Their study concludes that having more independent directors does not necessarily enhance board effectiveness, especially if they are unable to contribute to the board effectively. Therefore, we want to investigate the influence of the proportion of independent members on board on the degree of explicitness of CSR. This type of independence would be more specific and contribute to achievement of the aforementioned goals.

Thus, we hypothesize the following:

Hypothesis 1. The proportion of strictly independent directors in a firm is connected with the explicitness level in corporate reporting on CSR-related topics.

Board gender diversity and CSR Reporting

In the literature, there are debates about one particular characteristic of the board – gender diversity. Some studies focus attention on the critical role of gender diversity in board's decision-making (Rao & Tilt, 2016). For instance, Carter et al. (2007) rely on data from Fortune 100 index companies and find that board gender (and ethnic) diversity increases independence and triggers board to address questions that are less likely to be asked by board directors without such diversity. Board gender diversity also contributes to creativity, innovation, and high-quality decision-making (Erhardt et al., 2003).

Some research points to the impact of board gender diversity on non-financial performance including CSR reporting. Bear et al. (2010) investigate how board gender diversity and the number of women on boards affect firms' CSR ratings and how, in turn, CSR affects firms' reputation. They find a positive relationship between CSR and the number of women on the

board. The authors argue that the presence of women on board affects CSR rating through two major strength including increased sensitivity and participative decision-making style. Further, according to the study, a contribution that women bring to CSR reporting results in enhanced corporate reputation.

In a similar vein, Krüger (2009) concludes that the higher number of women on firms' boards results in higher rates of positive social responsibility activities. Firms with higher proportions of female directors tend to be more generous towards society and show more care to the welfare of a firm's natural stakeholders (e.g., communities, workers, or environment). This points to the fact that a stronger presence of board members with altruistic preferences does indeed translate into more pro-social corporate behaviour.

Notwithstanding the above literature, there is less (and no research to our knowledge) exploring the relationship between board's gender diversity and the degree of explicitness CSR reporting. While we see that studies support the idea that CSR reporting is a manifestation that seeks enhancement of a company's reputation and overall performance measures, it is value-adding to define whether gender diversity is one of the contributors to the explicitness of CSR reporting.

Thus, we hypothesize the following:

Hypothesis 2. Gender diversity on the board is connected with the explicitness level in corporate reporting on CSR-related topics.

Board size and CSR Reporting

Board size is another attribute of corporate governance which CSR disclosure studies frequently emphasize. Literature on board size can be classified into two categories. One is in favour of large boards, whereas the other advocates for smaller boards. The findings that support smaller boards claim that larger boards are inefficient as they are weak in management control and increase agency costs. However, this notion can be called into question by statement that larger boards may be less influenced by management. On the other hand, small boards are assumed to be efficient, but they may be influenced by managers (Naseem et al., 2017). This group believes that a board's effectiveness may decline as board size increases above a moderate number (typical suggestions are boards of seven to nine members) (Jensen, 1994).

An indirect way to assess whether boards may be too large is to examine the factors that affect board size. Firms with strong insider control and, therefore, greater incentive to choose optimal board size tend to have smaller boards. For example, Gertner and Kaplan report that firms that have undergone reverse leveraged buyouts (undergone a leveraged buyout and then gone public again) have smaller boards than public firms generally (Gertner & Kaplan, 1996).

Several studies find a positive correlation between the size of the board of directors and CSR disclosure (Zaid et al, 2019; Esa & Ghazali, 2012). According to Abeysekera (2010), a larger board size assists boards in overcoming skill deficits in allowing more substantial disclosures about future earnings. Chapple & Moon (2005) assert that a larger board size could contribute to a higher engagement and participation in CSR operations, and thus its disclosure in the annual report, through a more comprehensive exchange of ideas and experiences. For total CSR, environmental and social disclosure, (Dyduch & Krasodomska, 2017) have not found any association between disclosure level and board size, profitability, financial leverage, and women on the board.

Thus, we hypothesize the following:

Hypothesis 3. Board's size is connected with the explicitness level in corporate reporting on CSR-related topics.

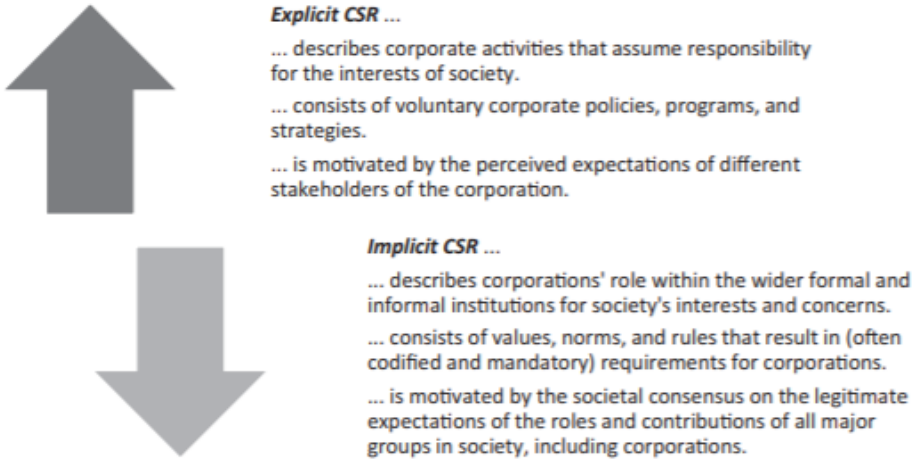

Figure 1: Explicit and implicit CSR (Source: adapted from Matten & Moon, 2008: 410)

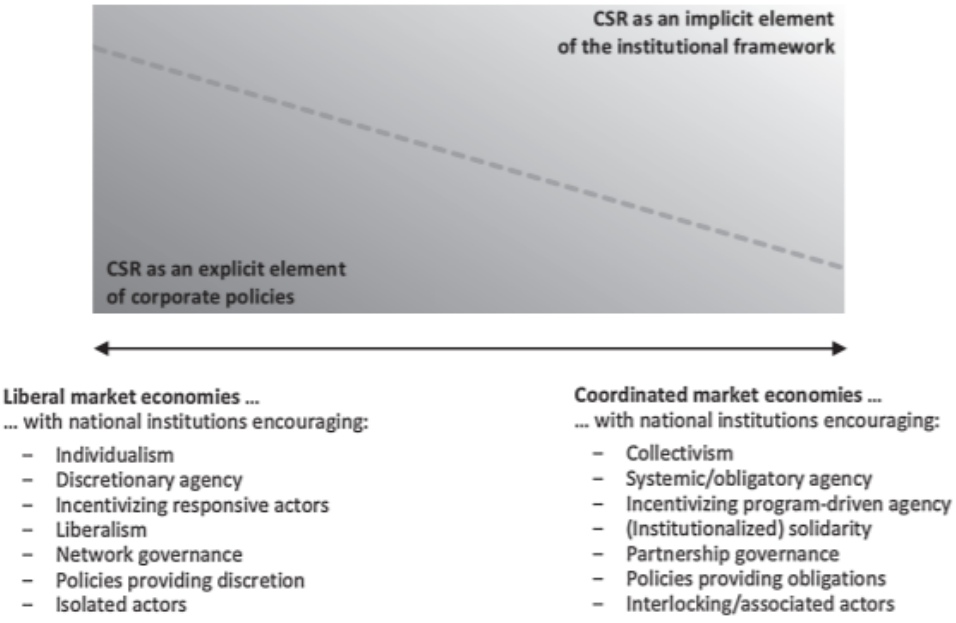

Figure 2: Continuum between explicit and implicit CSR (Source: adapted from Matten & Moon, 2008: 411)

Data

In this section, we start by presenting our research design. Then, we discuss our data collection and sampling process. It follows by the specification of dependent, independent, and baseline control variables. We subsequently provide descriptive statistics for the key variables in our sample.

Research Design

For our sample, we rely on 10-K reports from publicly traded U.S.-based companies that are accessible in the EDGAR database of the U.S. Securities and Exchange Commission (SEC). Data related to board characteristics is accessible for the 3 preceding years in Thomson Reuters Eikon, however, there is a substantial number of missing observations for some of the variables of interest which reduces the number of companies that we can overview in our work. We analyze annual reports published between April 1st, 2020 and March 31st, 2021, covering 1 full year of reporting.

We divide the data collection into two parts. Firstly, we identify the companies using Thomson Reuters Eikon and download data points of interest from the database. In the next phase, we download the reports of the companies from the EDGAR database. The reports are stored as 10-K forms that are submitted to the commission. When choosing the textual representation of the disclosure we are facing a choice between corporate social responsibility (CSR) report and more broad whole company-level reports that are available in EDGAR. One of the positive sides of analysis on 10-K reports is that the form has a rigid structure which allows for a more comparable analysis between companies.

In the following steps, we specify the dependent variables that relate to CSR topics. To identify the level of explicitness of the report based on the chosen topic, we pre-process reports and compare the number occurrences of the words from the topic training set with occurrences of such words in the report using cosine similarity.

Sample

Matten & Moon (2004) offer and justify the distinction between ‘explicit’ (American) and ‘implicit’ (embedded in broader norms and regulation) CSR and provide an argument that national CSR systems are a function of the intensity of the institutional framework and the

salience of social issues to individual corporations. Therefore, we consider all accessible U.S. companies.

In the first step, we choose 11727 U.S.-based companies that are publicly listed. To further align our research with comparable studies and increase the quality of the sample we identify criteria that ensure that we include more representative companies in the sample. Table 1 below provides a step-by-step process of sampling.

Companies that fall into our sample category should be fairly large and fairly liquid. Based on these criteria we include companies with a market capitalization of at least US$ 100 million that have a daily average traded volume in 2020 of at least US$ 100 000. Additionally, we consider all companies, except ETFs and investment funds that fall in the initial sample but need to be eliminated. This process narrows down the sample to 3393 companies. Some of the variables that are important for this research were missing in the dataset or look unreliable. For example, pretax ROA values of more than +100% and less than -100% and occurrences with board size values more than 17 are not in the final sample. It results in 945 remove observations. Finally, after we download and process reports from the EDGAR database, additional concerns appear from the second part of the data which deals with the length of the reports. We eliminate observations with a cleaned report length of fewer than 10 000 words. These steps reduce our final sample to a total of 2398 companies.

Table 1: Sample selection process

	Total
Initial population of US based firms	11727
Less: Companies with Market Capitalization less than US$ 100 m.	6044
	5683
Less: Companies with daily average traded volume less than US$ 100 000	1035
	4648
Less: investment funds and ETFs	1255
	3393
Less: observations with missing values	945
	2448
Less: oservations with dubious values	50
Total sample	2398

Table 2 shows that in total we have companies from 18 sectors. The manufacturing sector companies are an overwhelming part of the dataset, which reflects the nature of the economy.

Table 2: Sample distribution by industry group

NAICS sector	Number of companies
1 Accommodation and Food Services	48
2 Administrative and Support and Waste Management and Remediation Services	40
3 Agriculture, Forestry, Fishing and Hunting	5
4 Arts, Entertainment, and Recreation	10
5 Construction	40
6 Educational Services	12
7 Finance and Insurance	455
8 Health Care and Social Assistance	43
9 Information	192
10 Manufacturing	742
11 Mining, Quarrying, and Oil and Gas Extraction	78
12 Other Services (except Public Administration)	10
13 Professional, Scientific, and Technical Services	255
14 Real Estate and Rental and Leasing	178
15 Retail Trade	102
16 Transportation and Warehousing	67
17 Utilities	64
18 Wholesale Trade	57
Total Sample	2398

Analysis of the sample informs us about additional characteristics that are not included in the exploration part of this paper but are relevant as characteristics of the board of directors. When looking at the composition of board committees in the sample – all companies have a dedicated audit committee. The compensation committee is present in an overwhelming 98% of the firms in the sample. CEO is part of the board for 97,4 % of companies. Most companies (99,2%) have a policy on the experience of board members for qualification to be part of the board.

Information on CSR-related matters shows that 1,4% of companies are facing environmental controversies.

After we define the sample size, we use ticker information from the Eikon database and retrieve reports for the companies using the Central Index Key (CIK). For that, we use a transfer table between CIK and ticker that SEC provides on their site. To compile our unique dataset, for each of the firms we collect 10-K reports for the defined period.

We pre-process reports and clean them from HTML tags, stop words, numbers, and symbols. For cleaning meaningless words, we use a dataset containing a vector of Grady Ward's English words.

Dependent variables: constructing CSR-related topics

Previous research (Hummel et al., 2019 Fortuna et al., 2020; Matten & Moon, 2008) suggests that there are specific topics that can characterize CSR activities of the firm. Explicitness of disclosure on these topics can reflect the activities of the firms. 'Explicit CSR' refers to corporate policies that assume and articulate responsibility for some societal interests. They usually consist of voluntary programs and strategies by corporations that combine social and business values and address issues that are part of their social responsibility. As the sample for this study consists of companies that are located in liberal market economy (US), we suppose that, following deliverables of Matten & Moon (2008), most of the companies are following the tendency and are explicit when reporting on their CSR-related work. Thus, the only dimension that we should control for when defining the level of explicitness is measuring the alikeness of what the topic suggests and how the company is communicating on the chosen topic. Following the methodology refined by Hummel et al. (2019) and Fortuna et al (2020) and proposed by Campopiano & De Massis (2015), we choose 8 topics that are presented in Table 3:

Table 3: CSR-related topics identified by Campopiano & De Massis (2015) and Hummel et al. (2019)

Topic	Description
Customers	"This topic includes aspects such as satisfying customer expectations, customer loyalty, and involvement, fair prices"
Education	"This topic includes aspects that reflect firm's engagement in development of education and knowledge growth."
Employees	"This topic refers to good working conditions, to the involvement of employees in business strategies, to safe and non-discriminating working conditions"
Environment and green growth	"This topic relates to concerns for environmental conservation and includes issues such as responsible use of energy and material resources, reduction of pollution emissions, green research and innovation"
Philanthropy	"This topic refers to the respect for local community, engaging in projects for the quality of local community life and development"
Processes, products and services	"This topic is related to production efficiency, quality guarantees, and improvements, waste reduction, recycling materials, product safety"
Stakeholder management	"This topic includes all issues related to satisfying stakeholder claims, stakeholder dialogue, stakeholder involvement in decision-making process"
Values and general interests	"This topic refers to what is considered good, important, useful and desirable"

Building up on the methodology that is present, we process all the reports in our sample and create a dictionary with a following frequency list. In the first phase we include relevant terms that are representative of the topics and repeated at least 400 times in the data corpus. We

additionally create dictionaries and frequency tables for bigrams and trigrams in the corpus. In the next step we include some of the terms that contain unigram that we identified in the previous step. It assists in narrowing down the use cases of some terms that can have several connotations or areas of use. In such cases we prioritize relevant longer word combinations and eliminate shorter ones. The second column in Table 4 illustrates the resulting form of the topics that includes all chosen terms.

For every word or a word combination in the topic, we retrieve the words that appear before and after them. This approach helps to construct word windows that enable us to capture the topic more broadly than just the initial search terms. Each window has a length of 20 words. Depending on the length of the initial term, the word window takes 8 to 10 words before the term and 9 words after it. Due to a large number of terms, we want to avoid repetitions of parts of text in the new training corpus. To control for it, we define the position of the terms in the text and delete all word windows in which the distance from the previous word-window is less than 19 words. For each topic, we merge word windows and construct 8 training sets. Examples of the word windows presented in Appendix 2.

Before finalizing the new corpora of topics, we choose to narrow them down. The previous step adds from 17 to 19 new words that are initially not part of the topic that we select for the research. We notice that some of the more general terms reappear as a result of it. It is possible to identify popular general terms with higher reliability because of the larger number of terms that constitute each topic than previous research suggests.

We create a frequency table for all 8 new training sets and identify a list of general words that do not characterize CSR topics. We eliminate all of the word windows that include such terms. As it can be noted, the most popular word in the topic environment and green growth – ‘risk’, and in the topic values and general interests – ‘value’. Both of these words can be too general for their word windows to communicate information related to the topics and avoid nonessential connotations. We identify irrelevant bigrams and trigrams with these words and eliminate all word windows with occurrences that contain such terms. Table in Appendix 3 exemplifies the terms. These additional steps ensure a higher quality of the topic despite a larger number of terms included.

The third column in Table 4 depicts the number of topic words in the sample corpus. The next column summarizes the total number of words in the filtered training set for each of the topics and the final column presents the number of unique words in each topic. For comparison

reasons, we provide the number of unique words in the cleaned full sample corpus of reports – 38257.

Table 4: Description of the topics and variable assignment

CSR-related topic	Topic words	Number of topic words in the full corpus	Number of words in the topic corpus	Unique words in the topic corpus
Customers topic_1	client (14459), retention (9627), customary (8596), customer relationships (7407), satisfaction (6284), user (6024), purchasers (4703), customer base (4463), customer service (4333), loyalty (3345), buyers (3220), consumption (3129), purchaser (2561), customer experience (1825), relationships customers (1541), customer relationship (1540), consumer privacy (1469), service customers (1231), customer needs (1221), brand recognition (961), client relationships (823), technology customer relationships (530), good service customer (472)	89764	81589	6966
Education topic_2	research development (38500), studies (21004), study (16686), course (16358), training (12780), education (9088), university (8378), knowledge (7999), scientific (6460), learning (3647), science (3574), student (3246), school (2788), academic (2160), college (1541), tuition (1248), trained (928)	156385	316130	12193
Employees topic_3	professional (17707), labor (16672), staff (12843), workforce (11984), parent (11114), family (10358), employee stock (9163), health safety (8728), diversity (7660), talent (6906), salaries (6382), post retirement benefit (6117), salary (5750), workers compensation (5077), key personnel (4988), hire (4483), skill (1616), workplace (3648), hiring (3470), employee benefit (3397), wages (3292), qualified personnel (3074), recruit (3023), women (2770), families (2734), career (2667), employee compensation (2557), leave (2519), job (2319), illness (2223), base salary (2055), safety employees (1917), safety health (1556), work environment (1275), gender (1123), discrimination (949), health care benefits (915), employee stock options (908), employee health (785)	196724	449015	12529
Environment and Green Growth topic_4	risk (171958), global (64419), nature (24839), electric (15840), climate (8411), solar (7237), natural disasters (5905), globally (5836), renewable (5669), organic (5633), climate change (5457), contamination (4838), nuclear (4780), wind (3698), carbon (3459), footprint (3349), weather conditions (3135), biologic (2911), cleanup (2592), renewable energy (2526), greenhouse (2417), rate environment (2321), environmental health safety (2167), pollution (2135), toxic (1711), animal health (1551), solar energy systems (1260), greenhouse gas emissions (1215), local environmental (684)	367953	519108	14701
Philantropy topic_5	grant (42781), local (42556), care (34117), community (15108), social (14533), communities (12009), therapies (8153), sponsoring organizations (6636), therapeutic (6509), aid (4949), hospitality (3147), average grant (3031), initiative (2979), charitable (1413), sponsorship (967)	198888	410362	12955
Processes, products and services topic_6	quality (39479), reduction (34184), recoverable (10527), packaging (7393), quality control (2439), waste (7032), recycling (1767), recycled (995), product development programs (379)	104195	211031	10786

Stakeholder management topic_7	commitment (13901), shareholder (10225), investor (9886), stockholder (21379), service providers (10164), owner (7726), supplier (6840), engagement (6070)	86191	299496	10573
Values and general interests topic_8	value (413422), respect (54361), rules (38242), conduct (32179), recognition (31441), responsible (18826), responsibility (12964), participation (12238), reliability (8983), culture (6948), voluntary (5511), integrity (5185), transparency (3287), code ethics (2912), anticorruption (2257), justice (1960), voluntarily (1787), ethical (1359), welfare (1285), commonwealth (1238), transparent (590)	656975	377231	12678

The next step is to use the corpora with topics in comparison with the reports' texts. Measurement of similarity between frequency of the word in a topic and a report gives us a representation of the level of explicitness of the report. Topics contain fewer unique words than the dictionary of all topics. At the same time, in every report, the number of words is not consistent.

To measure how explicit a report addresses the pre-defined topics we compute the cosine similarity between the word vectors of each report and training set for every CSR topic. In the process the choices to be made about scaling of the measure and transformation of the number of occurrences of the words in a binary form. We wish to investigate the number of occurrences of words in a report, so we choose to preserve initial data and do not perform scaling as well. An important point when comparing a topic with a report is that we stress the importance of terms that exist in the topic. For every topic, only the words that are initially present in the topic are compared with the report. We include in the analysis words that are repeated at least 250 times in the reports to avoid excessive sparsity of the measure. Depending on the topic, we calculate the final measure on the window of unique 3458–5997 words. We further use an average of cosine similarity measures for all topics to define an average measure that takes into consideration all topics as an aggregative measure.

Another way of performing analysis on topics bases on a simple frequency of all terms of the topic and the creation of variables based on this measure. Cannon et al. (2020) employ this approach in a similar study that uses CSR-related topics with a substantial number of words in each group. We choose to use the cosine similarity measure because it captures the effects of similarity between every pair of words.

In the table of Appendix 3, we provide highlights of the 20-word windows for each topic and actual report text that corresponds to them.

Independent variables

Three independent variables that we include in this research are the level of independence of the board of directors, size of the board, and percentage of women on the board of directors. Variables for the size of the board and percentage of women on the board are part of the research as reported by companies.

For the level of independence, we choose to use the level of strict independence of the board as a more reliable measurement. The measure of strict independence provided by Thomson Reuters Eikon treats affiliated directors as dependent. Definition of a strictly independent board member includes the following:

1. not employed by the company;
2. not served on the board for more than ten years;
3. not a reference shareholder with more than 5% of holdings;
4. no cross-board membership;
5. no recent, immediate family ties to the corporation;
6. not accepting any compensation other than compensation for board service.

It could be that directors who are on the board for a long period, being nominally independent, create connections in the company and cannot be considered as such. Gilson & Kraakman (1991) mention that the concept of an independent director might not always follow the common wisdom. They argue that "corporate boards need directors who are not merely independent [of management], but who are accountable [to shareholders] as well." Some directors are connected to the company in ways that are not specified by existing definitions of "independence". Some examples of connections include paid consultancy or advisory roles for connected firms, or paid management roles in foundations or universities that are connected to the company. Bhagat & Black (1999) suggest that a way for these subtle relationships to be seen can be for the SEC to specify the need for additional disclosure of financial and personal ties between directors and the company, including CEO.

Control variables: determining baseline variables

We select control variables in alignment with prior research that examines CSR disclosure (Hummel et al., 2019; Fortuna et al., 2020; Cannon et al., 2020). The following measures are included in the analysis because they may exert influences on the outcome variable that overlap with or interfere with the influence of main independent variables. First, we rely on determinants of CSR disclosure and include the firm's market capitalization (control for size), industry, and pretax ROA (control for financial performance).

Market capitalization signifies a company's size (Wallace & Naser, 1995). Companies with large market capitalization are commonly exposed to political charges, such as demands by society for the performance of social responsibility or more comprehensive regulation, such as price controls and higher corporate tax (Watts & Zimmerman, 1990). More extensive disclosure reports like CSR can reduce these consequences. Conversely, companies with low market capitalization are more likely to feel that more open and specified disclosure would be detrimental to their competitiveness (Ahmad et al., 2017).

Rodriguez-Fernandez (2016) and Khan (2010) show that profitability is an essential factor in CSR disclosures by companies. Climent-Serrano & Pavía (2015) and Seo et al. (2015) believe that Return on Assets (ROA) is a trustworthy proxy for this term. Stakeholders expect both financial and non-financial disclosures. As disclosure reports need some costs, companies with poor profitability circumstances try to follow the least demanding way to disclose. When companies are more profitable, they have more tendency to spend money on reporting.

Second, we include additional variables that control for reporting patterns of the firm. Namely, we use a binary variable to define if a company uses the GRI framework in its reporting practices. The GRI is the most commonly used format worldwide for sustainability reporting. It provides standardization by requiring participants to report on economic indicators, environmental compliance, labor practices, human rights, society, and product responsibility. Reports are maintained in a publicly accessible database.

Third, we add a proxy to disclosure quality. Here we use the ESG disclosure performance score which tracks the company's success based on the reported information for specific data points in the environmental, social, and corporate governance pillars in the range between 0 and 100.

Descriptive statistics

Table 5 provides summary statistics for non-binary variables used in this study. Most of the measures for topics follow similar patterns. Interestingly, for topics 'education' and 'values and general interests' standard deviation is substantially higher than in the other topics, which suggests a higher disparity between reports in the sample. At the same time, max values for these topics are substantially larger, reaching 0,89 and 0,86 accordingly. For Independence, size, gender diversity, and market capitalization mean and median have similar values, which suggests symmetrical distribution. Roapretax has a negative mean but a positive median. Additionally, the standard deviation is 0.179 which suggests a substantial deviation from the mean. Market capitalization's values allocate between 1.67 and 27.4, as we take a natural logarithm of initial values to normalize them before including them in our analysis.

Table 5: Descriptive statistics

Variable	mean	median	sd	min	max	n
topic_1	0.1517	0.1392	0.0866	0.0206	0.6791	2398
topic_2	0.123	0.0331	0.2196	0.0046	0.8983	2398
topic_3	0.1621	0.1266	0.1086	0.0176	0.5088	2398
topic_4	0.1303	0.1033	0.0881	0.0123	0.6396	2398
topic_5	0.1636	0.1056	0.15	0.0165	0.6694	2398
topic_6	0.1248	0.0848	0.1055	0.0091	0.7431	2398
topic_7	0.1909	0.1810	0.0712	0.0262	0.4077	2398
topic_8	0.1626	0.0823	0.2161	0.0134	0.8689	2398
topics_all	0.1511	0.1161	0.1095	0.0177	0.4885	2398
strictindepbm	0.4969	0.4933	0.1719	0	0.923	2398
boardsize	9.2151	9	2.2381	5	17	2398
genderdiv	0.2108	0.2	0.1075	0	0.625	2398
markcap	21.78	21.70	1.67	18.43	27.4	2398
roapretax	-0.022	0.014	0.179	-0.989	0.827	2398
esgcombscore	37,95001	35,3759	17,1482	3,09	92,7544	2398

Correlation Results

Table 6 shows bivariate Pearson correlation coefficients. Statistics presents the full sample of 2398 firm-year observations for the regression variables. The coefficients indicate that all topic-based measures (except topic 1, which has close to zero correlation level with topics 2, 4, 6, 8) are positively correlated. At the same time, we cannot assume a high level of correlation of these variables with market cap, size, strict independence level, diversity roa pretax, and ESG

score. Among our main variables, we conclude that gender has a positive correlation with topics "Education", "Environment and Green Growth", "Processes, products and services", "Values and general interests". Therefore, based on the univariate statistics, we find support for our hypothesis only for gender diversity. In addition, the correlation statistics indicates that the ESG score has a moderately positive correlation with financial performance, market capitalization, level of gender diversity, and size of the board. Kutner et al., (2005) mention that the fact of correlation among independent variables "does not inhibit our ability to obtain a good fit nor does it tend to affect inferences about mean responses or predictions of new observations". Thus, we choose to preserve ESG score in the model.

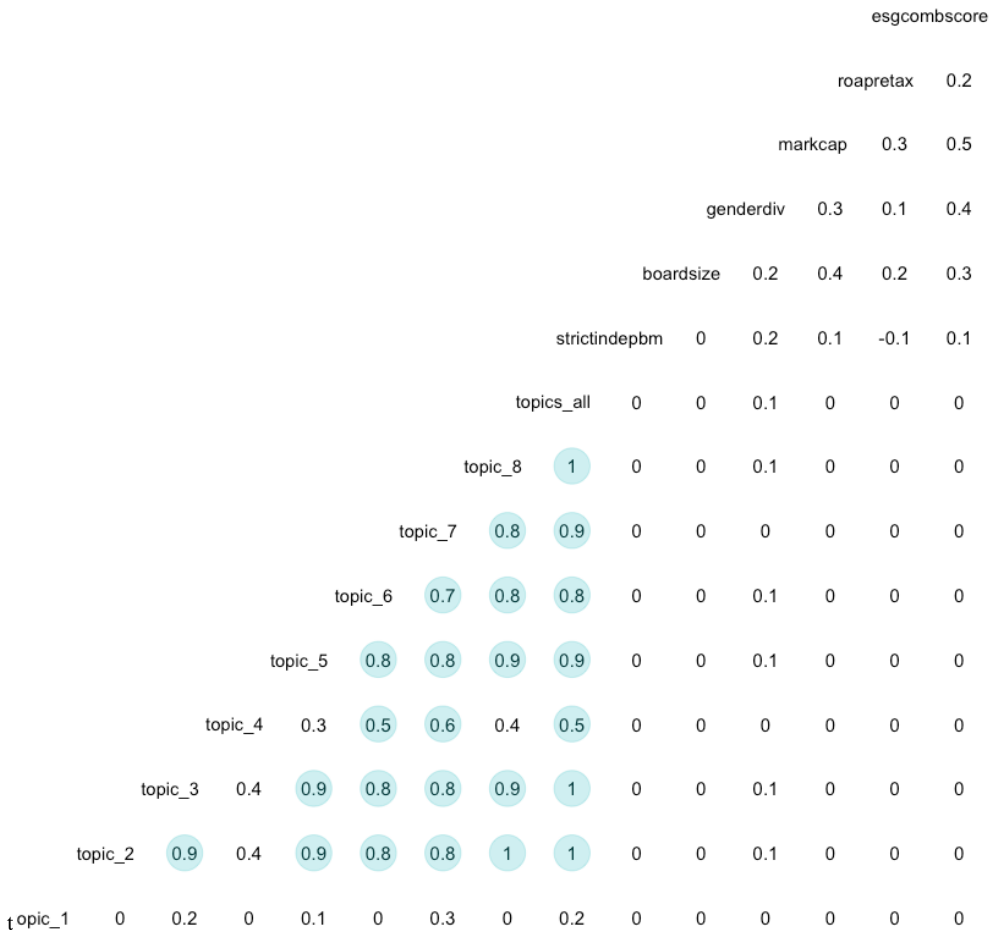

Figure 3: Correlation statistics with highlighted large values

Results

In this section, we specify the general form of the model. After that, we proceed with the analysis of the baseline regression models analyzing 8 topics and using the data from the previous steps. In the last phase, we include a set of modifications in our model to carry out a set of robustness checks to verify that our results hold in these modified states.

Empirical Model

We use the following empirical model to investigate whether there are differences in the level of explicitness of company reports and answer hypotheses about such difference depending on the level of independence of the board, the size of the board, and level of gender diversity:

$$Level\ of\ explicitness = \beta_0 + \beta_1\ size + \ \beta_2\ \ roapretax + \beta_3 stricindp$$

$$+\beta_4 boardsize + \beta_5 genderdiv + \beta_5\ esgcombscore + \beta_6\ grireport + \beta_7 naicssector$$

Level of explicitness proxies for our dependent topic include: topic_1– topic_8 and topics_all

In the model specification step, we use the Akaike information criterion to choose between different variations of the model and choose the most suitable variables for the representation of one or another characteristic of choice. The dependent variable in that phase is topics_all which reflects an average between all 8 specified topics. For the same case we also assess robustness of model. We test for multicollinearity by computing a variance inflation factor (or VIF).Values for all of the predictors do not raise suspicion for presence of multicollinearity: all GVIF are lower than reference boarders 3,5, and 10.

Table 6: Results for variance inflation factor of all predictors in the general model

	GVIF	Df	GVIF^(1/(2*Df))
markcap	1.762	1	1.328
roapretax	1.341	1	1.158
strictindepbm	1.076	1	1.037
boardsize	1.470	1	1.212
genderdiv	1.256	1	1.121
esgcombscore	1.989	1	1.410
grireport	1.700	1	1.304
naicssector	1.695	17	1.016

Baseline results from maximum likelihood beta regression

To establish a baseline relationship between the level of independence of the board of directors, size of the board, percentage of women on the board of directors, and the level of explicitness of topics, we estimate a set of beta regressions using maximum likelihood method:

Table 7: Results from regression analysis

	topic_1 (1)	topic_2 (2)	topic_3 (3)	topic_4 (4)	topic_5 (5)	topic_6 (6)	topic_7 (7)	topic_8 (8)	topics_all (9)
markcap	-0.001	-0.008	-0.004	0.002	-0.003	0.001	-0.001	-0.007	-0.003
	(0.009)	(0.017)	(0.011)	(0.010)	(0.013)	(0.012)	(0.007)	(0.016)	(0.011)
roapretax	-0.129*	0.061	0.004	-0.120	0.069	-0.035	-0.066	0.071	-0.007
	(0.072)	(0.135)	(0.086)	(0.083)	(0.104)	(0.098)	(0.058)	(0.127)	(0.087)
strictindepbm	0.022	-0.016	-0.069	0.001	-0.062	0.025	-0.003	-0.065	-0.017
	(0.067)	(0.126)	(0.080)	(0.079)	(0.097)	(0.092)	(0.055)	(0.118)	(0.081)
boardsize	-0.009	-0.004	-0.008	-0.003	-0.004	-0.008	-0.003	-0.005	-0.006
	(0.006)	(0.011)	(0.007)	(0.007)	(0.009)	(0.008)	(0.005)	(0.011)	(0.007)
genderdiv	0.060	0.373*	0.369***	0.183	0.432***	0.301*	0.203**	0.405**	0.341**
	(0.116)	(0.217)	(0.138)	(0.135)	(0.167)	(0.158)	(0.094)	(0.204)	(0.139)
esgcombscore	0.0001	-0.0001	-0.001	-0.0002	-0.001	0.0001	-0.0004	-0.0003	-0.0005
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
grireport	0.017	0.089	0.100**	0.007	0.091*	0.045	0.029	0.114*	0.079*
	(0.038)	(0.071)	(0.045)	(0.044)	(0.054)	(0.052)	(0.031)	(0.067)	(0.046)
naicssectorAdministra…	0.071	0.162	0.247*	-0.026	0.340**	0.153	0.103	0.181	0.194
	(0.118)	(0.218)	(0.139)	(0.138)	(0.167)	(0.160)	(0.096)	(0.206)	(0.142)
naicssectorAgriculture…	0.239	0.151	0.433	0.002	0.413	0.278	0.329*	0.193	0.290
	(0.247)	(0.480)	(0.289)	(0.303)	(0.355)	(0.339)	(0.199)	(0.451)	(0.301)
naicssectorEducational…	0.004	0.194	0.120	0.184	0.167	0.624***	0.159	0.184	0.254
	(0.181)	(0.329)	(0.212)	(0.200)	(0.254)	(0.220)	(0.142)	(0.310)	(0.209)
naicssectorOther…	0.146	0.234	0.260	-0.068	0.259	0.095	0.272*	0.273	0.239
	(0.188)	(0.354)	(0.222)	(0.227)	(0.270)	(0.260)	(0.149)	(0.332)	(0.225)
naicssectorRetail Trade	0.102	0.196	0.170	0.046	0.230*	0.155	0.124	0.230	0.195*
	(0.097)	(0.179)	(0.116)	(0.112)	(0.139)	(0.132)	(0.079)	(0.169)	(0.117)
naicssectorWholesale…	-0.007	0.235	0.212	0.055	0.274*	0.205	0.230***	0.285	0.230*
	(0.110)	(0.200)	(0.129)	(0.125)	(0.155)	(0.146)	(0.087)	(0.188)	(0.130)
Constant	-1.702***	-1.725***	-1.628***	-1.967***	-1.661***	-2.067***	-1.528***	-1.437***	-1.769***
	(0.194)	(0.361)	(0.232)	(0.226)	(0.279)	(0.264)	(0.158)	(0.340)	(0.234)
Observations	2,398	2,398	2,398	2,398	2,398	2,398	2,398	2,398	2,398
R^2	0.008	0.010	0.012	0.006	0.013	0.010	0.008	0.012	0.010
Log Likelihood	2,981.018	2,562.230	2,466.250	2,960.187	2,028.865	2,746.084	3,047.864	1,617.131	2,574.111

Note: This table reports maximum likelihood Beta-regression estiomates and P-values (in parentheses).We employ the same procedure on each CSR-related topic. We provide details on the dependent, independent and control variables in Appendix A1.

*p<0.1; **p<0.05; ***p<0.01

Cribari-Neto & Zeileis (2009) specify that beta regression is a standard maximum likelihood (ML) task for which there is no closed-form solution but numerical optimization is required. Practitioners commonly use a class of beta regression models, as introduced by Ferrari &

Cribari-Neto (2004), to model variables that assume values in the standard unit interval (0, 1). This approach naturally incorporates features such as heteroskedasticity or skewness commonly observed in data taking values in the standard unit interval, such as rates or proportions. The regression parameters are interpretable in terms of the mean of y.

As Brooks (2002) specifies: even though it could be possible to calculate the values of the standard goodness of fit measures such as RSS, R^2, or $\bar{R}^2$ for linear dependent variable models, these cease to have any real meaning. The objective of ML is to maximize the value of the likelihood function (LLF), not to minimize the residual sum of squares (RSS). Pseudo-R^2 is in use instead, where we target value 1 for the best possible model. However, we lose the simple interpretation of the standard R^2 that it measures the proportion of variation in the dependent variable that is explained by the model. Indeed, pseudo-R^2 does not have an intuitive interpretation. Pseudo-R^2 is highest for models that take topics Employees, Philanthropy and Values and general interests.

One of the control variables that refers to the GRI reporting shows significance in 4 of 9 models in which positive relation persists across all of these models. This variable shows the highest significance level in the model in which 'Employees' topic explicitness level is dependent.

The financial performance measure is interestingly not significant in the models, except having a 10% significance level in the model for topic Customers in which the direction is negative with. a coefficient value of -0,129. As specified in Sprinkle & Maines (2010), companies incur costs for CSR activities. They also suggest that NGOs often coerce companies to incur even more costs to report on their sustainability efforts, which sometimes does not have any significant positive shift in the paradigm but only creates additional costs.

Based on the Log-Likelihood measure, we can conclude that models in which topics 'environment and green growth' and 'stakeholder management' presented as dependent variables show better results. In the model with 'stakeholder management,' Log-Likelihood reaches 3047,9 in comparison to the lowest 1617,1 in model 'Values and general interests'.

For some models, industry sectors have shown a significant relationship with the topics. We present an overview of the findings in Table 8. For all presented pairs the connection is positive.

Table 8: Relationship between industry of the companies and dependent topics

Sector	Dependent topic(s)	direction of a relationship (+ or -)
Administrative and Support and Waste Management and Remediation Services	Employees / Philanthropy	+ / +
Agriculture, Forestry, Fishing and Hunting	Stakeholder management	+
Educational Services	Processes, products and services	+
Other Services (except Public Administration)	Stakeholder management	+
Retail Trade	Philanthropy / All topics	+ / +
Wholesale Trade	Philanthropy / Stakeholder management / All topics	+ / + / +

An important finding is that most of the models indicate the significance of gender diversity characteristic. Depending on the CSR-topic of the model, this ratio is significant at 10% level (models 2, 6), 5% (models 7, 8, 9) or 1% (models 3, 5) level. In models with explicitness level for topics 'Customers' and 'Environment and green growth' this variable is not significant.

In conclusion, we do not find a clear significant relation between two of the measures that are part of this work: level of independence and size of the board. However, there is a clear positive connection between the similarity of disclosure with the identified CSR-related topics and the proportion of women on the corporate boards. Log-Likelihood ratio test suggests that the model that provides better fit is model with topic 7 'Stakeholder management'.

Further analysis

We further conduct a battery of robustness checks to ensure that our instruments are valid and our system of equations is well-identified. The goodness of fit is assessed using different types of diagnostic displays following the methodology proposed for beta regression by Ferrari & Cribari-Neto (2004). Figure 3 depicts the results of the panels for the model 7 'Stakeholder management' that we take as an example for this section:

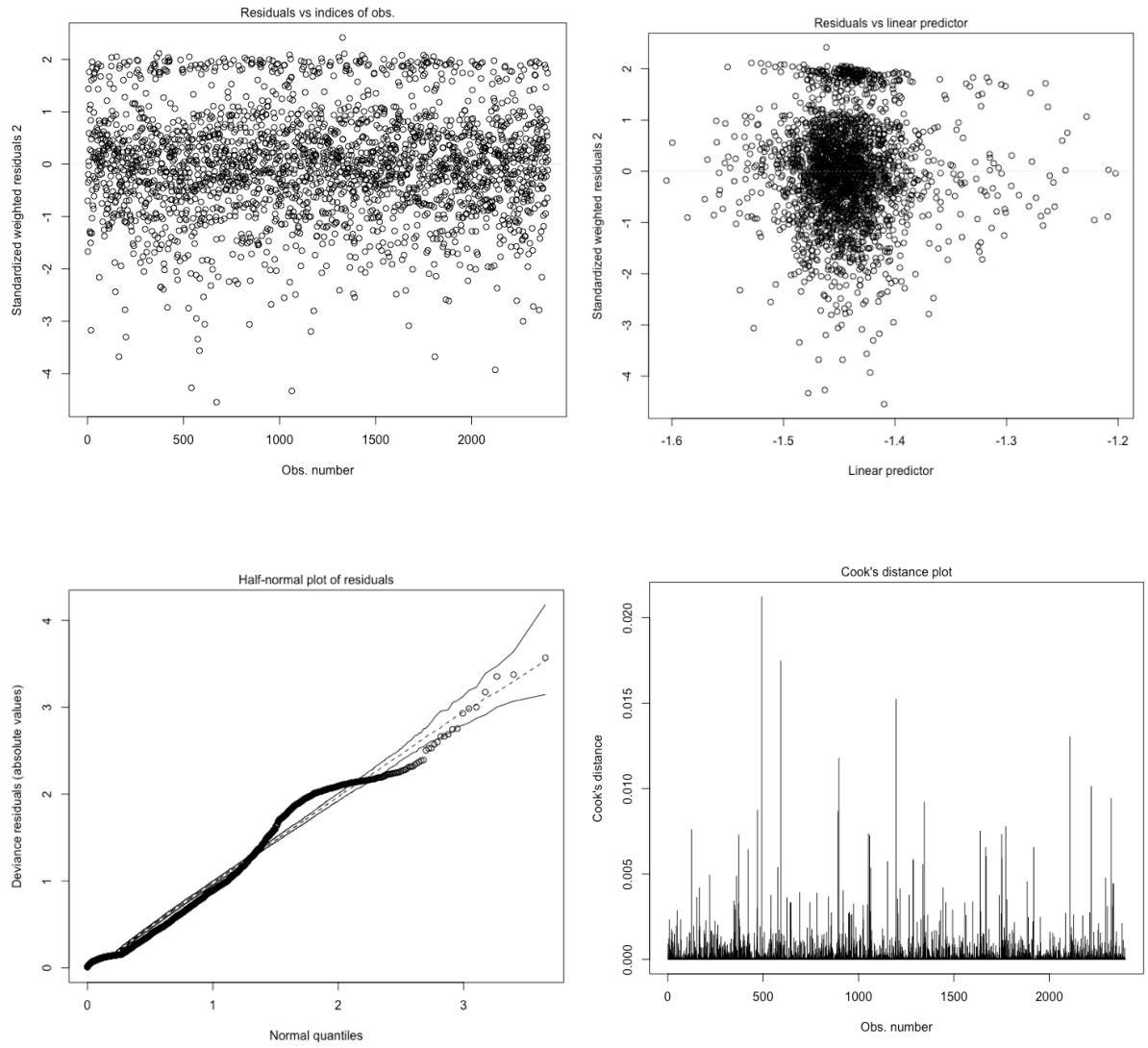

Figure 4: Diagnostics displays

The upper left panel plots the standardized weighted residuals against observation number. The upper right panel plots standardized residuals against linear predictor. The lower left panel displays the half-normal plot of absolute deviance residuals with a simulated envelope. The lower right panel presents a plot of Cook's distance measures versus observation number.

There are no substantive differences. In the case of beta regression, Espinheira et al. (2008) recommend using standardized weighted residuals. And we follow the proposed methodology. We assess residuals and note observations with values more than 2 (22 obs.) and less than -2 (66 obs.). We consider that any observations beyond these boundaries can be potential outliers. Additionally, from Cook's distance plot we see discrepant values. We use the reference number of 0,01 to identify 6 observations that fall out from the standard expected values.

The next point refers to the half-normal plot of absolute deviance residuals, diagnostics for beta regression, for now, is an area of active research. There are no universally accepted answers.

The outcome conditional on 2 parameters, as Ferrari & Cribari-Neto (2004) discuss in the reparameterization, has beta distribution. Consequently, we should not expect the raw residuals of a beta regression to be normally distributed; they should be beta distributed.

One concern that might arise is that 88 observations that have extreme weighted residuals and 6 potentially unusual but influential observations can substantially influence the setup of the model. In Table 8 we present a comparison of results between the actual model and the updated version that eliminates these values.

Table 9. Comparison between topic 7 models before and after cleaning data based on diagnostics displays

	topic_7	
	(1)	(2)
markcap	-0.001	-0.001
	(0.007)	(0.007)
roapretax	-0.066	-0.097*
	(0.058)	(0.054)
strictindepbm	-0.003	-0.029
	(0.055)	(0.051)
boardsize	-0.003	-0.006
	(0.005)	(0.005)
genderdiv	0.203**	0.297***
	(0.094)	(0.088)
esgcombscore	-0.0004	-0.00005
	(0.001)	(0.001)
grireport	0.029	0.042
	(0.031)	(0.029)
naicssectorAgriculture, Forestry, Fishing and Hunting	0.329*	0.302
	(0.199)	(0.231)
naicssectorFinance and Insurance	0.109	0.121*
	(0.070)	(0.066)
naicssectorManufacturing	0.100	0.129**
	(0.068)	(0.065)
naicssectorOther Services (except Public Administration)	0.272*	0.103
	(0.149)	(0.157)
naicssectorRetail Trade	0.124	0.146*
	(0.079)	(0.075)
naicssectorWholesale Trade	0.230***	0.221***
	(0.087)	(0.082)
Constant	-1.528***	-1.522***
	(0.158)	(0.147)
Observations	2,398	2,304
R^2	0.008	0.016
Log Likelihood	3,047.864	3,120.003
Note:	*p<0.1; **p<0.05; ***p<0.01	

We conclude that there are no substantial differences. We note that gender diversity level has a slightly higher power of significance in the new assessment. Financial control variable pretax ROA has a negative value close to -0.1 that is now significant on a 10% confidence level. There are some changes in industry control variables as well. However, we do not denote drastic positive changes in Log-likelihood and Pseudo-R^2 values.

Development of hypothesis about the level of independence

While we report in the literature review that we expect that the results on the connection of CSR and the board size might be mixed when looking back at the existing research that connects, there is supportive consideration for such a connection in boards with the high level of independence.

One of the reasons explaining the absence of connection might be that a company does not have as many as possible independent directors, but it is better to have some proportion of them.

Another concept that one might take into consideration when assessing the presence of inside directors may be a balance between independence and other factors that influence board choices. Inside directors are conflicted but well informed. Independent directors are not conflicted, but they might be less familiar with the company.

Having a reasonable proportion of inside directors on the board could be beneficial, which is why there is no clear linear connection between independence level and level of explicitness in reporting. Baysinger & Butler (1985) define an optimal board that consists of a mix of inside, independent, affiliated directors, who bring different skills and knowledge to the board.

In the literature review by Bhagat & Black (2001), we find some references to studies that promote the idea of the difference in behavior between firms with majority-independent boards and firms without such boards (Hermalin & Weisbach, 1998). However, none of the studies reviewed above investigate whether a supermajority-independent board, with only one or two inside directors, behaves differently than a majority-independent board. The theoretical case for such a high degree of independence affecting the board's monitoring ability is unclear.

We try testing this idea about a special level of independence that can have a meaningful implication. Relying on the methodology of Bhagat & Black (2001), we create a factor variable in which we assign the board of directors to one of four states:

Table 10: Explanation of a new factor variable

Proportion between: **Strictly Independent directors / (Insider directors + Affiliated directors)**	**Name of the category:**
[0/100 - 35/65]	Inside dominant
[35/65 - 50/50]	Inside prevalent
[50/50 - 65/35]	Independent prevalent
[65/35 – 100/0]	Independent dominant

The results for all 9 models in this new setup (Table 1 in Appendix 4) do not provide any significant contribution. We do not observe an interconnection of the level of independence of the board and more or less explicit reporting.

A final possibility is that independent directors can add value if they are part of an appropriate committee structure. In this way, independent directors might perform best within the monitoring function that they mostly represent, while inside and affiliated directors perform the informing and advising function to which they bring more firm-specific expertise. However, most large firms already have such committee structures and Klein (1998) finds little evidence that the outsider-dominated—audit, compensation, and nominating committees have no significant impact. Though, his study works with financial performance as a dependent variable. We try to analyze the relationship in our setup by observing two alternative committees. We add two binary variables that communicate the presence of corporate governance board committee and corporate and social responsibility or sustainable development board committee to the initial model. Table 2 in Appendix 4 presents that this direction of hypothesis development does not provide any significant results.

Finally, we conclude that changes in the level of independence on boards are unrelated after controlling for reporting performance, firm size, industry effects, board committees, and financial performance.

Conclusion

In this research, we explore the relationships between the characteristics of the board in firms and the explicitness of disclosures on CSR-related topics. We rely on measures derived from textual analysis following Hummel et al. (2019).

We test whether some board characteristics would improve the explicitness degree of reporting that companies submit in a 10-K form to the SEC. We do this by utilizing textual analysis on unique words that are categorized by topics. We do this using the forms of 2398 listed firms in the US capital markets that were submitted in one year period.

The findings in the previous chapter indicate that the characteristics of the board of directors can impact the quality of CSR disclosure. We consider size, independence, and gender diversity as three main elements of board characteristics. The result of the set of regressions shows that companies with more women on their boards tend to have more explicit CSR reporting. The results are consistent with the notion from previous research (Rao & Tilt (2016); Rouf & Hossan (2020); Cucari et al (2018)). At the same time, we conclude that both the number of members on the board and strict independence level has no significant influence on the level of explicitness. For the level of independence, we additionally control results for board committees and perform simplification of measures to control for non-linear results.

Our research has at least two main contributions. First, to the best of our knowledge, our study is the first to investigate the impact of organizations' board characteristics on the explicitness of CSR disclosures. The most recent study about the explicitness of CSR disclosures is by Hummel et al. (2019) which explores how economies that firms are located in impact the CSR disclosure level in terms of explicit and implicit CSR reporting. Our study goes beyond this and through a textual measure of CSR reports reveals how board composition affects the quality of CSR disclosure while executing analysis for one specific economy with a well-defined homogeneous pattern of reporting. Future research can extend this study by focusing on other characteristics of firms' board (e.g., director ownership, CEO duality, educational qualification of board members, and frequency of board meetings) and how they could affect the explicitness of CSR disclosures. Another extension is to analyze CSR reports, instead of general reports, which can provide new perspectives on the case.

Second, we contribute to the literature on gender-diverse boards and their impact on CSR reporting. Previous research has shown the importance of board gender diversity on boardroom

decisions and CSR outcomes (Rao, 2016). Our study contributes to this literature by illustrating a positive relationship between gender diversity, as one of the board characteristics, and the explicitness of CSR-related disclosure in annual reports. In other words, a more gender-diverse board contributes to higher quality and transparency of CSR report. While one could test further if explicitness is a result of actions that women directors direct to CSR initiatives.

Our study also has some limitations: first, we use cross-sectional data to test the hypotheses. Using panel data may produce more accurate results on the links between board composition and the explicitness of CSR reporting through analysis of larger data. Second, we analyze firms' reports texts based on 10-K disclosure. As the formats of CSR reports vary by organization, it is challenging to have a setup for comparative analysis. At the same time, some companies who already do CSR reporting might transfer some parts of it into general reporting frameworks. Meanwhile, execution of textual analysis on the text of 10-K forms brings additional noise in the analysis. Companies in some of the industries might discuss specific topics more explicitly, while the reporting form does not have a clear setup that would incorporate only CSR-related matters. When applied to 10-K filings, CSR-related topics do not always objectively represent information and reflect attempts of the companies to make positive changes. Whereas, they also can display a bias to certain industries. Third, our sample is constrained by data from one country. It might be interesting to explore the relationship between board composition and the explicitness of CSR across regions and in coordinated market economies.

Viewed collectively, the findings of this thesis indicate that there is an emerging need for a sustainability reporting framework, and discussions from SEC on the topic go in line with this. In a recent comment paper, Fleming &Ledbetter (2020) discuss policy proposals on Sustainability discussion and analysis that they evaluate as an initiative of critical importance.

In addition, In March of 2021, the U.S. Securities and Exchange Commission (SEC, 2021) addressed the public asking for inputs on Climate Change disclosure which is a clear positive step towards transparent reporting.

References

Abeysekera, I. (2010). The influence of board size on intellectual capital disclosure by Kenyan listed firms. *Journal of Intellectual Capital*, *11*(4), 504–518.

Agnolucci, P., & Arvanitopoulos, T. (2019). Industrial characteristics and air emissions: Long-term determinants in the UK manufacturing sector. *Energy Economics*, *78*, 546-566.

Aguilera, R. V., Williams, C. A., Conley, J. M., & Rupp, D. E. (2006). Corporate governance and social responsibility: A comparative analysis of the UK and the US. *Corporate Governance: an international review*, *14*(3), 147-158.

Ahmad, N. B. J., Rashid, A., & Gow, J. (2017). Board independence and corporate social responsibility (CSR) reporting in Malaysia. *Australasian Accounting, Business and Finance Journal*, *11*(2), 61–85. https://doi.org/10.14453/aabfj.v11i2.5

Alam, M. S., Atif, M., Chien-Chi, C., & Soytaş, U. (2019). Does corporate R&D investment affect firm environmental performance? Evidence from G-6 countries. *Energy Economics*, *78*, 401–411.

Baysinger, B. D., & Butler, H. N. (1985). Corporate governance and the board of directors: Performance effects of changes in board composition. *Journal of Law, Economics, & Organization*, *1*(1), 101-124.

Bear, S., Rahman, N., & Post, C. (2010). The impact of board diversity and gender composition on corporate social responsi- bility and firm reputation. *Business Ethics*, *97*, 207–221.

Bénabou, R., & Tirole, J. (2010). Individual and corporate social responsibility. *Economica*, *77*, 1–19.

Bhagat, S., & Black, B. (1999). The uncertain relationship between board composition and firm performance. *The Business Lawyer*, 921-963.

Bhagat, S., & Black, B. (2001). The non-correlation between board independence and long-term firm performance. *J. CorP. l.*, *27*, 231.

Brooks, C. (2002). Introductory econometrics for finance. Cambridge: Cambridge University Press.

Campopiano, G., & De Massis, A. (2015). Corporate social responsibility reporting: A content analysis in family and non-family firms. *Journal of Business Ethics*, *129*(3), 511-534.

Cannon, J. N., Ling, Z., Wang, Q., & Watanabe, O. V. (2020). 10-K disclosure of corporate social responsibility and firms' competitive advantages. *European Accounting Review*, *29*(1), 85-113.

Carter, D., D'Souza, F. P., Simkins, B. J., & Simpson, W. G. (2007). The diversity of corporate board committees and firm financial performance. *Available at SSRN 972763*.

Chapple, W., & Moon, J. (2005). Corporate Social Responsibility (CSR) in Asia: A Seven-Country Study of CSR Web Site Reporting. *Bus. Soc*, *44*, 415–441.

Climent-Serrano, S., & Pavía, J. M. (2015). Determinants of profitability in Spanish financial institutions. Comparing aided and non-aided entities. *Journal of Business Economics and Management*, *16*(6), 1170-1184.

Cribari-Neto, F., & Zeileis, A. (2009). Beta regression in R.

Cucari, N., Esposito De Falco, S., & Orlando, B. (2018). Diversity of Board of Directors and Environmental Social Governance: Evidence from Italian Listed Companies. *Corporate Social Responsibility and Environmental Management*, *25*(3), 250–266. https://doi.org/10.1002/csr.1452

Dyduch, J., & Krasodomska, J. (2017). Determinants of corporate social responsibility disclosure: An empirical study of Polish listed companies. *Sustainability (Switzerland)*, *9*(11). https://doi.org/10.3390/su9111934

Erhardt, N. L., Werbel, J. D., & Shrader, C. B. (2003). Board of director diversity and firm financial performance. *Corporate Governance: An International Review*, *11*, 102–111.

Esa, E., & Ghazali, N. A. M. (2012). Corporate social responsibility and corporate governance in Malaysian government-linked companies. *Corporate Governance: The international journal of business in society*.

Espinheira, P. L., Ferrari, S. L., & Cribari-Neto, F. (2008). On beta regression residuals. *Journal of Applied Statistics*, *35*(4), 407-419.

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. Journal of applied statistics, 31(7), 799-815.

Fleming, Rick A. & Ledbetter, Alexandra M. (2020) Making Mandatory Sustainability Disclosure a Reality, SEC Office of the Investor Advocate White Paper

Fortuna, F., Ciaburri, M., Testarmata, S., & Tiscini, R. (2020). CSR reporting and ownership structure: Evidence from Italian listed companies. *Corporate Ownership and Control*, *17*(3), 146–157. https://doi.org/10.22495/cocv17i3art11

Gertner, R., & Kaplan, S. N. (1996). THE VALUE-MAXIMIZING BOARD. *Working Paper*.

Gilson, R. J., & Kraakman, R. (1991). Reinventing the outside director: An agenda for institutional investors. *Stanford Law Review*, 863-906.

Herda, D. N., Taylor, M. E., & Winterbotham, G. (2012). The effect of board independence on the sustainability reporting practices of large US firms. Issues in Social & Environmental Accounting, 6.

Hermalin, B. E., & Weisbach, M. S. (1998). Endogenously chosen boards of directors and their monitoring of the CEO. *American Economic Review*, 96-118.

Höllerer, M. A. (2016). 2 Corporate social responsibility (CSR). In *Between Creed, Rhetoric Façade, and Disregard*. https://doi.org/10.3726/978-3-653-01464-8/6

Hummel, K., Mittelbach-Hörmanseder, S., Cho, C. H., & Matten, D. (2019). *Explicit corporate social responsibility disclosure: A textual analysis. December*, 1–54.

Ibrahim, A.H. and Hanefah, M. M. (2016). Board Diversity and Corporate Social Responsibility in Jordan. *Financial Reporting and Accounting*, *14*, 279–298.

Jamali, D., Safieddine, A. M., & Rabbath, M. (2008). Corporate governance and corporate social responsibility synergies and interrelationships. *Corporate Governance: An International Review*, *16*(5), 443–459.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of financial economics*, *3*(4), 305-360.

Jensen, M. C. (1994). the Modern Industrial Revolution, Exit, and the Failure of Internal Control Systems. *Journal of Applied Corporate Finance*, *6*(4), 4–23.

Khan, H. U. Z. (2010). The effect of corporate governance elements on corporate social responsibility (CSR) reporting: Empirical evidence from private commercial banks of Bangladesh. *International Journal of Law and Management*, *52*(2), 82-109.

Klein, A. (1998). Firm performance and board committee structure. *The Journal of Law and Economics*, *41*(1), 275-304.

Krüger, P. (2009). Corporate social responsibility and the board of directors. *Job Market Paper. Toulouse School of Economics, France*.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models*. McGraw-Hill.

Matten, D, & Moon, J. (2004). *Implicit and Explicit Corporate Social Responsibility: A conceptual Framework for Understanding CSR in Europe*. *April 2015*.

Matten, Dirk, & Moon, J. (2008). "Implicit" and "Explicit" CSR: A conceptual framework for a comparative understanding of corporate social responsibility. *Corporate Social Responsibility*, *33*(2), 257–277.

Mulcahy, R., & Donnelly, M. (2008). Board Structure, Ownership, and Voluntary Disclosure in Ireland. *Accounting in Emerging Economies*, *16*(5), 416–429.

Naseem, M. A., Riaz, S., Rehman, R. U., Ikram, A., & Malik, F. (2017). Impact of Board Characteristics on CSR Disclosure. *The Journal of Applied Business Research*, *33*(4), 801–810.

Rao, K., & Tilt, C. (2016). Board Composition and Corporate Social Responsibility: The Role of Diversity, Gender, Strategy and Decision Making. *Journal of Business Ethics*, *138*(2), 327–347.

Rodriguez-Fernandez, M. (2016). Social responsibility and financial performance: The role of good corporate governance. *BRQ Business Research Quarterly*, *19*(2), 137-151.

Rouf, M. A., & Hossan, M. A. (2020). The effects of board size and board composition on CSR disclosure: a study of banking sectors in Bangladesh. *International Journal of Ethics and Systems*, *37*(1), 105–121.

SEC. (2021, March 15). Public Input Welcomed on Climate Change Disclosures. https://www.sec.gov/news/public-statement/lee-climate-change-disclosures.

Seo, K., Moon, J., & Lee, S. (2015). Synergy of corporate social responsibility and service quality for airlines: The moderating role of carrier type. *Journal of Air Transport Management*, *47*, 126-134.

Sprinkle, G. B., & Maines, L. A. (2010). The benefits and costs of corporate social responsibility. *Business Horizons*, *53*(5), 445-453.

Wallace, R. O., & Naser, K. (1995). Firm-specific determinants of the comprehensiveness of mandatory disclosure in the corporate annual reports of firms listed on the stock exchange of Hong Kong. *Journal of Accounting and Public policy*, *14*(4), 311-368.

Watts, R. L., & Zimmerman, J. L. (1990). Positive accounting theory: a ten year perspective. *Accounting review*, 131-156.

Zaid, M. A., Wang, M., & Abuhijleh, S. T. (2019). The effect of corporate governance practices on corporate social responsibility disclosure. *Journal of Global Responsibility*.

Appendix

A1 Variable definitions

Table A1.1: Definitions of dependent and independent variables

Variable	Description	Source of data
Dependent Variable set		
Customers (topic_1)	Similarity between 10-K disclosure and constructed topic model that reflects disclosure on customer relationship	Textual analysis
Education (topic_2)	Similarity between 10-K disclosure and constructed topic model that reflects disclosure on education	
Employees (topic_3)	Similarity between 10-K disclosure and constructed topic model that reflects disclosure on employee-related subjects	
Environment and green growth (topic_4)	Similarity between 10-K disclosure and constructed topic model that reflects disclosure on subjects about environment and green growth	
Philanthropy (topic_5)	Similarity between 10-K disclosure and constructed topic model that reflects disclosure on philantropy	
Processes, products and services (topic_6)	Similarity between 10-K disclosure and constructed topic model that reflects disclosure on subjects about processes, products and services	
Stakeholder management (topic_7)	Similarity between 10-K disclosure and constructed topic model that reflects disclosure on how company manages its stakeholders	
Values and general interests (topic_8)	Similarity between 10-K disclosure and constructed topic model that reflects disclosure on company's value alignment and principles	
All topics (topics_all)	Similarity between 10-K disclosure and mean value for 8 defined CSR-related topics	
Main variables of interest		
Independence ratio (strictindepbm)	Level of independence of the board of directors	Thomson Reuters Eikon
Size of the board (boardsize)	The size of the board (between 5 and 17)	
Gender diversity ratio (genderdiv)	Percentage of women on the board of directors	
Control Variables		
Sector (naicssector)	Primary North American Industry Classification System (NAICS) Sector Description	Thomson Reuters Eikon

Market capitalization (markcap)	Natural logarithm of market capitalization in US
ROA (roa)	Return on assets before taxes.
Corporate governance board committee (cgboardcom)	Binary variable that shows whether company has a corporate governance board committee
CSR board committee (csrsustboardcom)	Binary variable that shows whether company has a nomination board committee
GRI repoort (grireport)	Binary variable that shows whether company's CSR report published in accordance with the GRI guidelines
ESG combined score (esgcombscore)	Overall company score (ESG Score) based on the reported information in the environmental, social and corporate governance pillars.

A2 Examples of twenty-word windows

Table A2.1: Examples of twenty-word windows

Twenty-word window	Corresponding CSR disclosure
Customers (topic_1)	Colgate-Palmolive
['also', 'face', 'strong', 'local', 'competitors', 'may', 'agile', 'better', 'local', 'consumer', 'insights', 'private', 'label', 'brands', 'sold', 'retailers', 'also', 'source', 'competition', 'certain']	"…In certain geographies, we also face strong local competitors, who may be more agile and have better local consumer insights than we do. Private label brands sold by retailers are also a source of competition for certain…"
Education (topic_2)	PepsiCo
['continued', 'growth', 'development', 'associates', 'supports', 'develops', 'associates', 'variety', 'global', 'training', 'development', 'programs', 'build', 'strengthen', 'employees', 'leadership', 'professional', 'skills', 'including', 'career']	"…the continued growth and development of our associates. PepsiCo supports and develops its associates through a variety of global training and development programs that build and strengthen employees' leadership and professional skills, including career…"
Employees (topic_3)	General Mills
['minor', 'temporary', 'workforce', 'disruptions', 'supply', 'chain', 'result', 'pandemic', 'implemented', 'employee', 'safety', 'measures', 'exceed', 'guidance', 'centers', 'disease', 'control', 'prevention', 'world']	"…minor temporary workforce disruptions in our supply chain as a result of the COVID-19 pandemic. We have implemented employee safety measures, based on guidance from the Centers for Disease Control and Prevention and World…"
Environment and green growth (topic_4)	Autodesk
['making', 'software', 'choice', 'poised', 'become', 'next', 'generation', 'professional', 'users', 'climate', 'change', 'addressing', 'global', 'challenges', 'posed', 'climate', 'change', 'make', 'possible']	"…making software of choice for those poised to become the next generation of professional users. Climate Change In addressing the global challenges posed by climate change, we make it possible…"
Philanthropy (topic_5)	3M
['strategy', 'brands', 'becoming', 'planet', 'positive', 'strengthening', 'roots', 'communities', 'advancing',	"…strategy and brands by becoming planet positive, strengthening our roots in our communities, and advancing

'social', 'justice', 'includes', 'supporting', 'practices', 'technologies', 'improve', 'farmer', 'livelihoods', 'agricultural', 'resiliency']	social justice. This includes supporting practices and technologies that improve farmer livelihoods and agricultural resiliency…"
Processes, products and services (topic_6)	Kellogg Company
['across', 'value', 'chain', 'meeting', 'targets', 'responsibly', 'ingredients', 'reducing', 'organic', 'waste', 'providing', 'packaging', 'living', 'founder', 'values', 'engaging', 'billion', 'people', 'issue', 'global']	"…across our value c hain by meeting our science-based targets, responsibly sourcing our ingredients, reducing organic waste and providing sustainable packaging. â€¢Living our founderâ€™s values by engaging 1.5 billion people in the issue of global…"
Stakeholder management (topic_7)	BestBuy
['procure', 'seek', 'mitigate', 'risk', 'enhance', 'partnership', 'suppliers', 'create', 'value', 'stakeholders', 'responsible', 'supply', 'chain', 'program', 'active', 'members', 'responsible', 'business', 'alliance', 'allows']	"…procure, we seek to mitigate risk, enhance the partnership with our suppliers and create value for all stakeholders through our Responsible Supply Chain Program. We are active members of the Responsible Business Alliance, which allows..."
Values and general interests (topic_8)	Packaging Corp of America
['areas', 'along', 'efforts', 'continue', 'develop', 'promote', 'maintain', 'diverse', 'workforce', 'culture', 'environment', 'respect', 'inclusion', 'principles', 'designed', 'develop', 'promote', 'strong', 'increasing', 'engagement']	"…areas along with efforts to continue to develop, promote and maintain a diverse workforce with a culture and an environment of respect and inclusion. These principles are designed to develop and promote strong and increasing engagement…"

A3 Terms for additional cleaning of topics

Table A3.1: Terms to perform additional cleaning on topics

Subgroup	Words and word combinations
'General' terms for cleaning 20-word windows for each topic	account, acquire, acquisition, agreements, america, approved, april, asset, asterisks, balance sheet, bank, business, california, capital, cash, collateral, compliance, cost, credit, debt, december, deposit, discount, dividend, drug, earning, end, equity, expense, february, federal, filed, financial, first, fiscal, fluctuation, government, impairment, inflation, invest, january, july, june, laws, lease, legal, liability, litigation, loan, march, market, may, million, mr, obligation, october, operating results, policy, portfolio, president, price, pricing, pursuant, rates, receivable, regulation, regulator, return, revenue, securities, september, settlement, stock, tax, transaction, valuation, volatility, words, years, york
Combinations with 'value' in the topic values and general interests	fair value (237312), carrying value (32642), fair values (21629), market value (16094), estimated fair value (15740), par value (14198), present value (13606), value reporting (12390), value measurements (11116), fair value measurements (11105), fair value reporting (10751), value assets (10674), value per (10164), stock par value (10027), value hierarchy (9476), fair value hierarchy (9440), value reporting unit (9213), changes fair value (8988), value per share (8765), date fair value (8456), measured fair value (7534), value measurement (7534),fair value measurement (7449), value company (7044), value shares (6270), book value (6116), fair market value (6015), change fair value (5793), value common (5524), par value per (5461), intrinsic value (5363), fair value assets (5254), level fair value (5210), value stock (5112)…
Combinations with 'risk' In the topic environment and green growth	risk factors (22977), credit risk (17978), risks related (14608), risk management (13385), risks uncertainties (11990), market risk (11426), item risk (10920), item risk factors (10790), rate risk (10243), interest rate risk (9131), risks associated (8889), disclosures market risk (5361), subject risk (4887), subject risks (3898), risk controls (3310), periods subject risk (3289), risks material (3265), subject risk controls (3249), risk controls may (3246), risk associated (3195), risk factors item (2952), risk loss (2937), assess risks (2825), risks relating (2811), risks procedures (2712), assess risks material (2708), respond risks procedures (2708), procedures respond risks (2707), procedures assess risks (2705), risk characteristics (2538), risk material (2526), risk item (2354), market risk item (2264), assessing risk (2202), business item risk (2187), risk item financial (2170), assessed risk (2164), based assessed risk (2140), assessing risk material (2139), reporting assessing risk (2139), risk material weakness (2139), market risks (2076), price risk (2065)…

A4 Modified models in relation to level of board independence and board committees

Table A4.1: Regression models with factor variable for independence level

	topic_1 (1)	topic_2 (2)	topic_3 (3)	topic_4 (4)	topic_5 (5)	topic_6 (6)	topic_7 (7)	topic_8 (8)	topics_all (9)
...									
strictindepbm1independent prevalent	0.017	0.016	0.024	-0.029	0.037	-0.024	0.009	0.014	0.011
	(0.033)	(0.062)	(0.040)	(0.038)	(0.048)	(0.045)	(0.027)	(0.058)	(0.040)
strictindepbm1inside dominant	-0.026	0.008	0.028	-0.024	0.033	-0.011	-0.002	0.030	0.003
	(0.037)	(0.069)	(0.044)	(0.043)	(0.053)	(0.050)	(0.030)	(0.065)	(0.045)
strictindepbm1inside prevalent	0.025	-0.006	0.018	-0.020	0.022	-0.042	-0.007	-0.002	-0.002
	(0.032)	(0.060)	(0.039)	(0.038)	(0.047)	(0.044)	(0.026)	(0.057)	(0.039)
genderdiv	0.043	0.371*	0.362***	0.179	0.425**	0.313**	0.201**	0.405**	0.336**
	(0.116)	(0.217)	(0.138)	(0.135)	(0.167)	(0.158)	(0.094)	(0.204)	(0.140)
grireport	0.013	0.089	0.100**	0.007	0.090*	0.048	0.030	0.115*	0.080*
	(0.038)	(0.071)	(0.045)	(0.045)	(0.055)	(0.052)	(0.031)	(0.067)	(0.046)
Constant	-1.691***	-1.729***	-1.670***	-1.948***	-1.704***	-2.034***	-1.523***	-1.469***	-1.773***
	(0.196)	(0.364)	(0.234)	(0.228)	(0.281)	(0.266)	(0.159)	(0.343)	(0.236)
Observations	2,398	2,398	2,398	2,398	2,398	2,398	2,398	2,398	2,398
R^2	0.009	0.010	0.011	0.007	0.013	0.011	0.008	0.012	0.010
Log Likelihood	2,982.312	2,562.313	2,466.135	2,960.480	2,028.993	2,746.589	3,048.083	1,617.165	2,574.170

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

Table A4.2: Regression models with binary variables for board committees

	topic_1 (1)	topic_2 (2)	topic_3 (3)	topic_4 (4)	topic_5 (5)	topic_6 (6)	topic_7 (7)	topic_8 (8)	topics_all (9)
…									
cgboardcom	0.033	0.008	0.043	0.051	0.024	0.074	0.030	0.007	0.028
	(0.046)	(0.085)	(0.055)	(0.054)	(0.066)	(0.063)	(0.037)	(0.080)	(0.055)
genderdiv	0.059	0.372*	0.364***	0.178	0.428**	0.289*	0.200**	0.405**	0.338**
	(0.116)	(0.217)	(0.138)	(0.135)	(0.167)	(0.158)	(0.094)	(0.204)	(0.140)
grireport	0.031	0.100	0.109**	0.019	0.092	0.056	0.036	0.126*	0.089*
	(0.040)	(0.074)	(0.047)	(0.047)	(0.057)	(0.054)	(0.032)	(0.070)	(0.048)
csrsustboardcom	-0.036	-0.029	-0.016	-0.025	-0.0002	-0.016	-0.015	-0.031	-0.022
	(0.034)	(0.063)	(0.040)	(0.040)	(0.049)	(0.046)	(0.028)	(0.060)	(0.041)
Constant	-1.755***	-1.758***	-1.670***	-2.025***	-1.676***	-2.134***	-1.562***	-1.471***	-1.808***
	(0.199)	(0.370)	(0.237)	(0.231)	(0.286)	(0.271)	(0.161)	(0.348)	(0.240)
Observations	2,398	2,398	2,398	2,398	2,398	2,398	2,398	2,398	2,398
R^2	0.009	0.010	0.012	0.007	0.013	0.011	0.009	0.012	0.010
Log Likelihood	2,981.852	2,562.340	2,466.641	2,960.847	2,028.933	2,746.856	3,048.360	1,617.276	2,574.397

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