The thematic content of ESG reports
A topic modeling approach

ABSTRACT

This paper aims to show the impact of the thematic content of ESG reports on information asymmetry between managers and investors. Using a sentence-based topic modeling algorithm, we determine and empirically quantify 30 topics in a large collection of 1,667 ESG reports between 2007 and 2020. We find that the algorithm produces a semantically meaningful set of topics and that there exist clear patterns in the topics discussed in ESG reports, along with significant temporal shifts in managers' focus on environmental, social or governance issues. We also show that the content of such reports helps explain investors' behavior and reduces information asymmetry around and after the release date of the report. Furthermore, we evidence that there are substantial differences in the value relevance of the topics discussed in the reports and that investors react more to ESG reports that discuss environmental matters, while they tend to find the social and governance pillars less informative. Finally, there exists a concave relationship between the number of topics discussed in ESG reports and their informativeness. Overall, this paper theoretically contributes to prior literature by showing that investors are not only sensitive to whether ESG reports are disclosed, but also to what information is provided.

Highlights

(1) This paper contributes to prior literature by providing evidence that the thematic content of ESG reports decreases information asymmetry.
(2) This paper unlocks the black box of ESG reports, examines its thematic content and distinguishes 30 topics contained in ESG reports.
(3) After categorizing topics across the three ESG pillars, we find that the three pillars show significant temporal shifts in managers' focus on environmental, social and governance issues.
(4) Although the thematic content is informative to investors over the short and long term, a limited number of topics leads to a significant stock price reaction.
(5) Investors react more to ESG reports that discuss environmental matters and less to social and governance topics.
(6) The value relevance of ESG reports has a concave relationship with the diversity of topics discussed in the report.

***JEL* Classification:** G15, M13, L26, D80.

Keywords: ESG reports; Topic modeling; Information value.

1. Introduction

Investors have been a key driving force in bringing Environmental, Social and Governance (ESG) reports to their current prominence by placing an ever greater focus on ESG criteria in their capital allocation process (Eccles et al., 2011).[1] According to the Government Accountability Institute (Governance Accountability Institute, Inc., 2021), while less than 20% of firms of the S&P500 provided an ESG report in 2011, this percentage increased to 92% by 2020. Although ESG reports are today an essential part of corporate disclosure, we still know very little on how investors use ESG information and, in particular, *what* information investors find of value relevance in ESG reports. This question is an important one as investors base their decisions on information that remains unaudited, voluntary and unregulated (Allen and Ramanna, 2013). By relying on unsupervised machine learning, this study opens the black box that constitute ESG reports and aims at investigating the informativess of ESG reports' thematic content in explaining investors' behavior around and after the report's release. As such, while many of the current studies in the literature are interested in the value relevance of ESG rating disagreements (Christensen et al., 2021; Dimson et al., 2020; Brandon et al., 2021), our paper extends on earlier conflicts between the shareholder and stakeholder theory to show that an ESG reports' informational content is a function of *what* is discussed in the document and provides evidence on whether managers can effectively mitigate information asymmetry through the thematic content of ESG reports.

The objective of this paper fits with the current efforts by authorities to define a framework regarding informative ESG reporting. While the European Union (EU) tightened its "Non-Financial Reporting Directive in 2021, the SEC has yet to follow in its footsteps.[2] One reason for the SEC's

[1] As the ESG construct is generally viewed as a representation of a firm's contribution to sustainability (Kleine and Von Hauff, 2009; Carrasco-Monteagudo and Buendía-Martínez, 2013), reports containing this type of information are published under a variety of labels (Davis and Searcy, 2010; Threlfall et al., 2020), including sustainability reports, sustainable development reports, CSR reports and triple bottom line reports. In this paper, we take it that these names are equivalent and use "ESG reports" as an umbrella label.

[2] The Sustainable Finance Disclosure Regulation 2019/2088/EU addresses the lack of consistency and clarity regarding the consideration of ESG factors in investment decision-making processes. This Regulation follows the Directive 2014/95/EU on non-financial reporting, which already required large firms to provide meaningful sustainability information. The new regulation aims at encouraging the investment sector to meet the goals of The Paris Agreement by reducing greenwashing and driving capital

hesitation is that it needs to identify "which, if any, sustainability disclosures are important to an understanding of a registrant's business and financial condition" (Securities and Exchange Commission, 2016; Grewal et al., 2021). Although the SEC's Investor Advisory Committee recently called for Regulation S-K to be modified in order to subject issued ESG reports to the same rules as other types of information (e.g. GAAP principles in financial reporting), the progress depends on which ESG topics are of value relevance to investors.[3] Without a clear mapping of which ESG topics are relevant to investors, the SEC cannot demand additional reporting requirements on U.S. firms (Grewal et al., 2021). This regulatory progress depends on which ESG topics are of value relevance. Under such circumstances, there is a need to determine whether the thematic content of ESG reports contains information value and, in particular, *what* topics in ESG reports are of relevance.

Prior literature has extensively examined the information value of ESG reports to predict firm performance or explain firm value, with mixed results. On the one hand, the stakeholder theory argues that ESG disclosure is used by managers as a tool to provide information for the various stakeholders (such as, e.g., employees, suppliers, shareholders, customers, banks, regulatory agents), which ultimately increases earnings quality (Kim et al., 2012) and reduces information asymmetry (Cormier et al., 2009; Clarkson et al., 2013; Cormier et al., 2011). In fact, reporting on ESG activities has been shown to improve financial performance (Arendt and Brettel, 2010), increase firm value (Cormier et al., 2009; Reverte, 2014; Clarkson et al., 2020), reduce financial risk (Cormier et al., 2009; Mishra and Modi, 2012; Cormier et al., 2011), enhance access to finance (Cheng et al., 2013), lower the cost of equity (Dhaliwal et al., 2014; Cormier et al., 2009; Reverte, 2011), reduce analysts' forecast dispersion (Aerts et al., 2008; Cormier and Magnan, 2013) and improve analyst forecast accuracy (Dhaliwal et al., 2012; Cormier and Magnan, 2013). More recently, Du and Yu (2020) show that investors react positively to optimistic and readable ESG reports and conclude that *how* an ESG

towards ESG goals. The regulation now requires ESG disclosure from firms in the financial sector with more than 500 employees to update their disclosed product information on sustainability issues and to follow new standards for dealing with sustainability risks or negative sustainability events.

[3]Regulation S-K is a prescribed regulation under the U.S. Securities Act of 1933 that lays out reporting requirements for various SEC filings used by public firms.

report is written significantly reduces information asymmetry. On the other hand, the shareholder theory adopts a sceptical view regarding ESG investments and states that a firm's main goal is to maximize its value (Friedman, 1963). In line with this theory, Clarkson et al. (2013) find marginal evidence on a relationship between cost of capital and voluntary social disclosure, while Richardson and Welker (2001) provide what they refer to as *surprising* evidence regarding a positive relationship between the cost of capital and social disclosure. Similarly, Berkman et al. (2019) find a positive relation between climate risk disclosure and the cost of capital. Against this backdrop, it is likely that the information value of ESG disclosures is more subtle than just assuming that more information is better (Cormier et al., 2009).

This lack of consensus regarding the value relevance of ESG reports may be explained by the fact that prior literature generally assumes the content of ESG reports to be homogeneous and discards to identify *what* is effectively disclosed. Yet, given the lack of audit and regulation surrounding ESG disclosures, ESG reports have remained largely heterogeneous in length and themes discussed. With such flexibility and freedom in the reporting of the content of ESG information, investors are likely to value specific information on various aspects of the firm's ESG performance (e.g., carbon footprint or gender diversity), while attribute less value or even ignore other topics (e.g. nutrition and compliance). In this paper, it is therefore necessary to innovate by relaxing the assumption that ESG reports are an homogeneous collection of topics. Instead of examining *whether* and *how* ESG information is reported, we argue that not all topics have the same value relevance and that some themes will lead to a decrease in information asymmetry, whereas other topics may have limited value relevance.

We rely on unsupervised machine learning methods to map the topics discussed in ESG reports. We then test whether the thematic content of ESG reports is informative in explaining the market reaction around and after the report's release date. The thematic content of a ESG report is defined as the distribution of underlying topics in a document, and is measured as the number of sentences relating to a specific theme. To identify the topics contained in ESG reports, we use a process-based approach

based on the Latent Dirichlet Allocation (LDA) at the sentence-level (sentLDA) introduced by Bao and Datta (2014). This method reasonably assumes that a sentence is the smallest integral unit of text that conveys a complete and meaningful idea (Ivers, 2010), and thus incorporates the information in sentence boundaries while identifying the topic clusters (Bao and Datta, 2014). Consequently, it allows us to identify a topic for each sentence, as opposed to estimating the topic distribution of the entire document. Because the purpose of our paper is to test ESG reports' informativeness and identify what topics are of relevance for investors, rather than simply classifying the documents, sentLDA is particularly suited for the task. As a result, our approach allows for semantic categories to emerge from the data. The topical composition is then used to discover *what* ESG information influences the market at the release of the report.

We map the informational composition of 1,667 ESG reports for 416 firms in the United States between January 2007 and September 2020. We find 30 optimal topic categories and evaluate their semantic validity using both human and machine-based procedures. Altogether, we find that the sentLDA method produces a coherent set of meaningful topics that capture ESG reports' content. Our results show significant diversity in the topics covered in ESG reports, ranging from discussions regarding environmental performance such as climate change and water conservation, to community relations, diversity, safety standards and health, as well as transparency concerns, accountability and work culture. We find that community relations is the most discussed topic in the analyzed collection, followed by diversity, the description of collaboration and the details on safety standards and compliance. Our analyses also reveal a number of interesting topic shifts, pointing to dynamic practices in ESG reporting. The environmental ('climate change', 'supply chain', and 'emission') and social ('safety standards', 'community relations', and 'diversity') aspects of the ESG construct are significantly given more prominence, whereas topics regarding governance appear to be less discussed in recent years. Furthermore, although the main objective of ESG reports is to demonstrate firms' actions and activities in relation to society and environment, we note that still a significant proportion

of the corpus mentions financial concerns and achievements.

We then examine whether the market reacts to the thematic content of ESG reports. While ESG performance has the potential to affect equity pricing, we find that information regarding ESG performance also affects stock prices and influences the level of information asymmetry around the release of the ESG report. We use the *absolute* cumulative abnormal return to estimate investors' reaction around the release of ESG reports and the price impact measure defined by Amihud (2002) around the release of the ESG report as a proxy for information asymmetry. Our tests indicate that, relative to models that include standard firm-specific, text-based control variables and other fixed-effects identified in prior literature, including the thematic content significantly improves the models' capacity to explain the absolute cumulative abnormal return and the Amihud measure around the release of the ESG report. Specifically, our results show that the thematic content increases the adjusted R-square for the model explaining absolute cumulative abnormal returns by 21.940%, which is economically significant.

We also find that not all topics have equal information value in an ESG report. We find that investors react more to ESG reports that discuss environmental matters and less to social and governance topics. In particular, we find that environmental topics tend to decrease information asymmetry, while social topics have a marginal influence on the asymmetry at the release of the report. On the other hand, we find that governance topics tend to be detrimental to the level of information asymmetry around the release of the report. These results confirm that investors are sensitive to the release of ESG reports, but, more importantly, that the ESG report's value relevance depends on the themes discussed in the reports.

We then analyze the market impact of ESG reports' topic diversity. We are interested in understanding whether investors' response to ESG topics depends on whether the document covers a narrow or large array of topics. We find that reports that disclose a more diverse ESG topics typically lead to a stronger stock price reaction and more significantly reduce information asymmetry. This result

suggests that reports that cover a wider array of topics are therefore considered as more informative, i.e. investors consider the report to include more value relevant information. However, cataloging all possible generic ESG topics, regardless of their information value for investors does not further reduce information asymmetry. In fact, we find that this relationship between topic diversity and information value is curvilinear as, beyond a threshold, topics do not contribute to the report's information value any longer.

Several additional analyses further our understanding of the thematic content of ESG reports. First, we analyze how the value relevance of ESG reports has evolved over time. From sub-sample analyses, we find that the informativeness of the documents is particularly notable in more recent periods. Second, we examine the information value of the thematic content of ESG reports across industries. Relying on the Barth et al. (2001) industry classifications, we find substantial heterogeneity in ESG topics' value relevance across industries. Third, our findings related to importance of the topics contained in ESG reports are corroborated by an analysis using bid-ask spread as an alternative proxy for information asymmetry. Fourth, we find that the thematic content of ESG reports helps explain a firm's ESG performance and predict future firm value (e.g. Tobin's Q). Finally, because our sample only includes firms that provide stand-alone ESG reports, our estimation models may be subject to the sample selection bias. We therefore conduct the Heckman two-stage procedure (Heckman, 1979) to account for the endogenous nature of firms' decision to publish an ESG report. Our results remain qualitatively similar.

Our paper makes several contributions to the literature. First, we add to the shareholder and stakeholder theories by providing novel empirical evidence on the value relevance of ESG reports, which is yet to offer adequate conclusions on the role played by ESG disclosures in decreasing information asymmetry. To the best of our knowledge, this paper is the first to show that ESG reports' direct stock market impact around and after the report's release is a function of the thematic content of the document. Prior research typically focuses on *whether* disclosing on ESG reports influences the

firm's cost of capital (Dhaliwal et al., 2010; Cormier et al., 2009; Reverte, 2011), firm value (Clarkson et al., 2020) and analysts' forecasts (Dhaliwal et al., 2012; Cormier and Magnan, 2013; Aerts et al., 2008), or examines *how* an ESG report helps explain the stock market's reaction around the release of the report (Du and Yu, 2020). Our study complements this line of research and highlights that, in addition to *whether* and *how*, *what* is discussed by managers in ESG reports matters for investors and reduces information asymmetry. Taken together, this study suggests that ESG reports are a critical vehicle for imparting value relevant information and substantiates the important role of ESG reports' content in increasing information transparency.

Second, we contribute to prior literature on ESG disclosures by showing that invesors do not value all ESG topics equally (see, e.g., Cormier et al., 2011; Barth and McNichols, 1994; Aerts et al., 2008). We find that it is the selection of value relevant topics that make the ESG report informative, as discussing all possible ESG topics, regardless of their value to investors, is not a strategy that pays off. In particular, we find that environmental topics tend to be more important than social. We also evidence that, relative to environmental and social topics, governance-related activities are less of value relevance for investors. In that sense, we extend prior findings that social disclosure (see, e.g., Cormier et al., 2009, 2011) and environmental disclosures (see, e.g., Barth and McNichols, 1994; Aerts et al., 2008; Cormier et al., 2009; Cormier and Magnan, 2013; Cormier et al., 2011) are informative and mitigate information asymmetry. The machine-learning approach used in this paper significantly departs from prior research. Although prior literature's methods based on human coders or pre-defined lists of words may benefit from increased classification accuracy (see, e.g., Verbeeten et al., 2016; Cormier et al., 2011; Aerts et al., 2008; Cormier et al., 2009), the use of a machine learning-based method allows us to map the content of ESG reports at a more granular level and analyze a larger number of documents.[4] This more granular approach in distinguishing topics also allows us to investigate a more extensive set of dimensions to evidence ESG reports' informativeness. In particular,

[4]Section 4.2 discusses the validation methods used to confirm that the sentLDA provides a valid set of semantically meaningful topics that are reasonably coherent.

we provide evidence in this paper on the diversity of topics discussed in the document and show how topical diversity decreases information asymmetry, up to a given threshold. This approach matches with Hoberg and Lewis (2017), who argue that the flexible nature of disclosure content requires a more extensive set of dimensions along which we can analyze corporate narratives.

Third, our results advance the literature's current understanding of what is effectively contained in ESG reports. In two studies close to ours, Jaworska and Nanda (2018) and Goloshchapova et al. (2019) identify the topics discussed in ESG reports. They apply word-based machine learning methods and provide descriptive evidence on the topics discussed in ESG reports. Similar to our results, they reveal a number of major trends and topic shifts pointing to changing practices of ESG. Topics such as 'people', 'communities', and 'rights' seem to be given more prominence, whereas 'environmental protection' appears to be less discussed. As such, our research is related to but distinct from that of Jaworska and Nanda (2018) and Goloshchapova et al. (2019). We provide similar descriptive evidence regarding the trends in the topics of interest in ESG reports. However, the finality of our study significantly departs from theirs as they do not leverage this information to test the informational value of the thematic content of ESG reports and identify what topics are of value relevance to investors. We thus contribute to this prior literature and take a significant step forward by examining whether the selection of topics discussed influences the market around the release of the report.

Fourth, there is only scarce evidence on the direct stock market impact of ESG reports' release. To our knowledge, the only such work investigating the direct impact of the release of an ESG report on the stock price is that of Du et al. (2017) and Du and Yu (2020), who examine whether ESG reports lead to a stock price reaction on the release date and whether the reports' tone and readability are related to this market reaction. Complementing Du et al. (2017) and Du and Yu (2020), our study therefore provides additional evidence that ESG reports are informative for investors, but more importantly highlights that, controlling for an array of firm-, industry-, year- and linguistic-specific factors, the investor reaction also depends on its thematic content. Our results therefore confirm that

investors do pay attention to the release of such reports and incorporate sustainability information in their stock valuation. Given the persuasive influence of the stock market on managers' decisions (Currim et al., 2012; Du and Yu, 2020), our results provide insight to managers on which topics in ESG reports matter the most. [5]

2. Motivation and literature review

2.1. Does ESG performance influence financial performance?

ESG performance measures can be informative to investors to the extent that ESG activities increase firm value. ESG performance has been shown to affect financial performance through various channels. For instance, prior research shows that the implementation of ESG-oriented practices supports firm performance by reaping a variety of marketing benefits, such as more favorable product evaluation and purchase behavior, higher trust, satisfaction and loyalty (see, e.g., Du et al., 2017; Brammer and Millington, 2008; Carmeli et al., 2007; Kim et al., 2012; Zulu-Chisanga, 2019).[6] Hasan et al. (2016) also show that ESG practices help a firm raise its productivity, while Luo and Du (2015) find evidence that addressing ESG challenges triggers innovation. In addition, a positive approach to ESG practices helps a firm attain legitimacy and the license to operate in local communities, as well as to obtain more positive treatment from the media (Fombrun et al., 2000). As such, Godfrey et al. (2009) show that ESG initiatives act as an insurance policy, which reduces the negative externalities caused by negative events or crises. Furthermore, ESG performance improves communication to shareholders on financial matters (Fieseler, 2011) and encourages more effective corporate governance (Blazovich and Smith, 2011; Jo and Harjoto, 2011, 2012). For instance, Ghosh and Wu (2012) show how analysts consider firms' ESG performance when they make recommendations.

[5] It is beyond the scope of this paper to examine whether the topics are manipulated to manage investors' impressions regarding the firm ESG performance. Future research should investigate whether managers tend to shift topics or omit discussing specific topics when the ESG or financial performance is below expectations.

[6] During the financial crisis of 2008, a firm's ESG efforts have been confirmed to help them successfully re-establish trust between themselves and their investors (Giannarakis and Theotokas, 2011).

There also exist other channels through which ESG practices can affect firms' financial performance and value. For instance, voluntary ESG-oriented behavior can help firms avoid government regulation and, therefore, reduce compliance costs (Dhaliwal et al., 2010). Socially responsible firms also appeal to consumers who care about the corresponding social issues, which leads to superior sales and financial performance (Lev et al., 2009). Taken together, prior research has largely investigated the economic impact of ESG performance and concludes that ESG performance has a positive impact on the firm's financial performance and increases firm value (Du et al., 2017; Fombrun et al., 2000).

2.2. ESG reports – An unaudited, unregulated and multi-stakeholder disclosure

To realize these benefits, an increasing number of firms are disclosing reports that discuss their ESG performance. In fact, ESG reporting has become prevalent in the past 40 years both in academic research and business strategy.[7] The success of ESG reporting is clearly visible in the exponential growth in the number of firms that measure and report ESG information (Cho et al., 2015). A recent survey by KPMG (2020) reports that 96% of the 250 world's largest companies provided sustainable reports in 2020. This focus on ESG reporting is largely motivated by the pressure from stakeholder groups, such as non-governmental organizations, heightened government regulation and increased investor interest in ESG-related information, but mostly because ESG information complements financial reporting (Vitolla et al., 2019, 2020).

However, relative to other corporate disclosures studied in prior management literature, such as the 10-K, CEO letter to shareholders, 8-K or the MD&A (Loughran and Mcdonald, 2016; Boudt and Thewissen, 2019; Arslan-Ayaydin et al., 2016, 2020, 2021), ESG reports are unique because they provide a deeper and larger vision on corporate social performance in critical domains in the firm, such as employee welfare, diversity, community outreach, product safety, the use of energy and the

[7]Starting in 1980s, the notion of corporate social responsibility started to be increasingly used to refer to firms' actions in response to their environmental impact (Carroll and Shabana, 2010). In fact, the first stand-alone reports focused predominantly on environmental issues and were therefore referred to as "environmental reports". Recently, corporations include a wider range of issues in ESG reports (Milne and Gray, 2013) and discuss their impact on the society and their decisions in terms of firm governance.

impact on the environment. In addition, compared to financial reporting that provides information to the investor community and specializes on providing quantitative data, ESG information is primarily reported in the form of textual information regarding the firm's policies, practices and performance in social, environmental and governance domains (Dhaliwal et al., 2010). Another important contrast with mandatory financial information is that ESG reports are voluntary, largely unregulated and do not fit in a widely enforced reporting framework (Perrini and Tencati, 2006; Tschopp and Huefner, 2015).

In spite of significant attempts to enforce an ESG reporting topology (Allen and Ramanna, 2013), managers keep a substantial discretion in whether, what type and how they report ESG-related information. In fact, there exists no template dictating what a ESG report should or should not include. This lack of regulation regarding ESG reporting explains the diverse reporting practices with respect to length, performance indicators and topics contained in the report. In addition, the audit process of such reports is neither comprehensive, nor stringent compared with the verification of corporate annual reports. Even if the report is audited, the scope of the audit relates to the process and less on the information itself (Muslu et al., 2017). The discretionary and largely heterogeneous nature of ESG documents raises the question of whether ESG reports contain credible information, resonating with a longstanding debate in financial accounting between voluntary versus mandatory disclosures (see, e.g., Healy and Palepu, 2001).

2.3. The value relevance of ESG reports

There has been an increasing number of studies on the value relevance of ESG reports for capital markets, with mixed results. While some evidence points towards a negative relationship between ESG reports and information asymmetry, other studies highlight that ESG reports increase the asymmetry of information on capital markets. To explain this relationship, we distinguish two competing theories. On the one hand, the shareholder theory adopts a sceptical view regarding ESG investments and

states that a firm's main goal is to maximize firm value. As such, ESG-related expenditure is considered a poor use of investors' money and this practice deviates from shareholder value maximization (Friedman, 1963). Accordingly, investing in ESG misuses resources that could otherwise help optimize profits for shareholders. As a result a higher ESG performance goes against shareholders' interest and reduces the firm's long term value.

On the other hand, a central prediction of information economics is that the positive impact of ESG performance on firm value would lead managers to provide information relevant to investors voluntarily (Clarkson et al., 2020). The stakeholder theory suggests that firms should go beyond the interests of shareholders and that ESG is essential for corporations in obtaining necessary resources and stakeholder support (Freeman, 1984; Jones, 1995). Because managers are moral actors inclined to exercise their actions toward socially responsible outcomes and to improve information transparency (Wood, 1991), Freeman (1998) argues that ESG disclosure should be seen a means to improve information transparency (Wood, 1991; Dhaliwal et al., 2010).

Three reasons explain this relationship between ESG disclosure and information transparency. First, prior theoretical research in financial accounting argues that the disclosure of information leads to liquid and efficient financial markets, resulting in a lower cost of capital for firms (Grossman and Hart, 1980; Milgrom, 1981; Botosan, 1997; Diamond and Verrecchia, 1991; Graham et al., 2005; Dhaliwal et al., 2010). Second, ESG activities can increase the size of the investor base. Heinkel et al. (2001) argue that ESG oriented investors prefer to exclude firms with low ESG performance from their portfolio. This means that firms with high ESG performance can attract more investors. The larger the size of the investor base, the lower the information asymmetry and the higher the market valuation (Merton, 1987). Third, disclosing on ESG information helps build the firm's reputation as a good social citizen. Cui et al. (2016) theoretically and empirically show that firms with better ESG transparency increase their reputation and are more likely to obtain capital at a lower cost. As such, to realize these benefits, managers need to publish an external ESG report in order to inform

potential employees, customers, and community members of their ESG activities (see, e.g., Clarkson et al., 2020). Taken together, the stakeholder theory argues that ESG disclosure should be positively correlated with information transparency.

The fact that there persists such a theoretical disagreement on the relationship between ESG reporting and corporate performance has been the stimulus for numerous empirical studies over the last 20 years. Some findings are consistent with the stakeholder theory. For instance, Dhaliwal et al. (2010) examine the impact of ESG disclosures on firms' information asymmetry. Based on a sample of firms that initiate the disclosure of ESG reports, they find that the disclosing ESG activities leads to a significant reduction in firms' cost of equity capital. Dhaliwal et al. (2010) also find that the issuance of ESG reports is associated with lower analyst forecast error. Furthemore, Aerts et al. (2008) and Cormier and Magnan (2013) show that enhanced environmental disclosure translates into more precise eanrings forecasts by analysts, which suggests that environmental disclosures decrease information asymmetry. Extending on this evidence, Cormier et al. (2009) investigate the impact of precision attributes of social and human capital disclosure on information asymmetry. They show that quantitative disclosure reduces share price volatility and increases the firm's Tobin's Q. This study on information precision led more recent studies to examine *how* ESG reports' information is disclosed and which linguistic features influence investors' trading decisions. Du and Yu (2020) find that, despite the tendency for managers to obfuscate information when firm performance is poor, the tone of ESG reports has a significant impact on the market reaction at the release of the report. Similarly, they find that a more readable report leads to a more pronounced stock market reaction. Overall, they conclude that *how* ESG reports are written conveys information to the market, reduces information asymmetry and increases firm value.

In contrast, some scholars align with the shareholder theory and argue that disclosing on ESG is costly and could eventually increase the firm's cost of capital. For instance, Chen et al. (2018) examine how disclosure on ESG impacts firm performance. Their analysis exploits China's 2008

mandate requiring firms to disclose ESG activities, using a difference-in-differences design. They find that mandatory ESG reporting firms experience a decrease in profitability subsequent to the mandate. Similarly, Grewal et al. (2019) further examine the equity market reaction to events associated with the passage of a directive in the European Union mandating increased non-financial disclosure. Consistent with Chen et al. (2018), they find a marginal average negative market reaction of –0.79% across all firms, but also that the negative market reaction is concentrated among firms with weak pre-regulation ESG performance and disclosure, which exhibit an average return of –1.54%. In contrast, firms with strong pre-regulation disclosure and performance exhibit an average positive return of 0.52%. Berkman et al. (2019) also find a positive association between climate risk related disclosure and the cost of capital and argue that the long-term nature of environmentally responsible projects contribute to increased risk and increase information asymmetry. Finally, Richardson and Welker (2001) find evidence that the disclosure of social information is detrimental to the cost of capital.

2.4. Unlocking the black box of ESG reports: ESG reports' thematic content and its information value

The conflicting nature of the results obtained in prior literature suggests that the theoretical predictions may be incompletely specified. Rather than accumulating studies and trying to control for an increasing number of variables, another research direction is advisable. In fact, there remains an important gap in clearly understanding the value relevance of ESG reports. While prior research mostly focuses on *whether* (Dhaliwal et al., 2010) or *how* ESG information is disclosed (Du and Yu, 2020), it overlooks the context or types of information conveyed in the document. Yet, because of the lack of a formal control mechanism or standardization, ESG reports are significantly different from each other, which reflects in considerable differences in the report's length. For instance, the AT&T's 2010 sustainability report (ATT, 2010) has 82 pages, while the 2011 report of the NYSE Euronext company contains 40 pages. In addition, given that managers have significant discretion in

the information they choose to disclose, ESG reports significantly differ in their content. A typical ESG report covers a firm's ESG performance along a large array of dimensions, including human resources, environment, community, customers, suppliers, risk management, charitable giving, corporate governance, the company's support of local/international initiatives and cultural development. In addition, within each issue domain, the report covers the firm's actions and performance regarding various aspects of that domain (Perrini and Tencati, 2006). In such context, ignoring the content of ESG reports constitutes a major shortcoming in our understanding of the information value of ESG reports.

We propose that ESG reports' information value needs to be reflected on the basis of the granular concepts (e.g. topics) that it represents. What is value relevant to investors not only depends on *whether* and *how* the information is disclosed, but also on *what* type of information is provided in the report. For instance, Amel-Zadeh and Serafeim (2018) conduct a survey, which shows that, when considering investing in a stock, investors seek information on specific aspects of the firm's ESG performance and weigh this information more depending on the theme of the information. Their descriptive evidence provides indication that the ESG report's information content most likely depends on the thematic content disclosed in the document. Furthermore, Cormier et al. (2011) distinguish between social and environmental disclosures and test whether these disclosures have a substituting or complementing effect in reducing information asymmetry. Based on a pre-established grid, they identify social and environmental disclosure in 137 web disclosures for the year 2005 and evidence that social and environmental disclosure substitute each other in reducing stock market asymmetry. This result is important as it suggests that there may be interacting effects between the themes discussed and that social disclosure reinforces the informativeness of environmental disclosure for the stock market. Based on these results, we expect ESG reports with different content to have different informational value to investors. We therefore need to relax prior literature's assumption that the ESG report is an homogeneous collection of topics (see, e.g., Dhaliwal et al., 2010, 2012; Du and

Yu, 2020; Clarkson et al., 2020). Instead, we argue that ESG reports are clusters of a variety of topics that, individually, have a different impact on investors. Taken together, we expect the thematic content of ESG reports to improve the explanatory power of investors' reaction and information asymmetry around the release of the document to the market, beyond what can be achieved by using quantitative financial metrics and other measures of textual style features identified in prior literature. Specifically, we address the following hypothesis:

Hypothesis 1: Ceteris paribus, ESG reports' thematic content significantly explains investors' reaction and the level of information asymmetry around the release of ESG reports.

The results of this hypothesis offer insights into how and why managers disclose their ESG activities to investors and inform standard setters who are considering possible ESG disclosure requirements. We automatize the identification of topics discussed in a report, which complements previously popular approaches of identifying value relevant information through surveys (Amel-Zadeh and Serafeim, 2018), grids (Cormier and Magnan, 2013), experiments (Martin and Moser, 2016) or keywords (Zhang et al., 2010). If we find evidence consistent with our prediction that the thematic content has a stock market influence around the ESG report release, then the next objective is to identify which topics, specifically, explain this influence. To provide some insights into this issue, we then explore the impact of each topic on the market reaction around the ESG report release.

2.5. Topic diversity in ESG reports

Based on this granular analysis of ESG reports' thematic content, we next examine whether the number of of topics discussed in ESG reports influences its informational value. There is some evidence in prior research suggesting that the number of topics (henceforth; topical diversity) increases information value. For instance, Son et al. (2019) show that discussing more topics in online consumer reviews provides more information about a product's features and functions. Similarly, Azarbonyad

et al. (2015) find a positive relationship between the value of a document and topical diversity. This evidence indicates that, to make ESG reports more informative, managers should craft their reports by incorporating more ESG themes. A report discussing a large number of themes would then contribute to the investor assessment of the firm's future performance, future goals, and risk/opportunity management strategies. As a result, we expect a larger number of topics discussed in ESG reports to be associated with to a stronger stock price reaction at the release of the report and a decrease in the level of information asymmetry. However, given the lack of standardization, it may also be that managers exhaustively catalog all possible generic ESG topics, regardless of their relevance for investors. A laundry list of generic risks has little instructive value and would not increase the information value to investors. We can therefore expect that, while the informativeness of ESG reports initially increases with the number of topics, additional topics, after a treshold, do not contribute to the information value of the document, leading to a curvilinear relationship with investors' reaction and information asymmetry around the ESG report's release.

Hypothesis 2: Ceteris paribus, the diversity of topics discussed in a ESG reports has a concave (convex) relationship with investors' reaction (information asymmetry) around the ESG report's release.

3. Sample selection, text preprocessing, sentLDA alogirthm, variable definitions and research design

3.1. Sample selection and ESG reports text preprocessing

3.1.1. Sample selection

This paper focuses on stand-alone ESG reports provided by firms included in the MSCI database between 2007 and 2020. It is important to note that a part of the ESG information could be included

in annual reports (10-Ks), CEO letters to shareholders or public media. However, the analysis of a stand-alone ESG report allows to isolate the market reaction to the ESG report. Therefore, by looking at isolated ESG reports instead of other financial disclosures, the market's reaction around the release of the ESG report is less likely to be influenced by financial or other news and to only reflect ESG information.

To collect our ESG sample, we proceed in two steps. At each step in the sample selection, any observation with missing data for the required control variables is discarded. In the first step, we put together a sample of firm-year observations with all the required dependent and control variables. Our initial sample includes all the firms included in the MSCI ESG database between 2007 and 2019. The sample contains 35,578 firm/year observations for 6,148 firms. To obtain accounting-based data, we merge the MSCI sample with US COMPUSTAT annual industrial, which results in 23,793 observations. Next, we merge our sample with the stock market CRSP data and get 19,928 observations. When merging with the IBES database to obtain the analyst forecasts, we lose a significant number of observations. With IBES analysts' forecasts, our sample is composed of 7,874 data points. We finally include the geographical and business segments from COMPUSTAT to our database and have 5,956 firm-year observations left. In the second step, we use the sample of 5,956 observations and hand-collect the ESG reports. We obtain ESG reports from various internet sources, including CSRwire.com, CorporateRegister.com, GlobalReporting.org, SocialFunds.com, BusinessWire.com, and corporate websites. This step leads to a sample of 1,789 reports.

To measure the market response to ESG reports' topics around the release of ESG reports, we need the extract date on which the ESG report has been released. We search for the issue dates of the ESG reports on the company's website or on the document. ESG reports with unidentifiable release dates are excluded. To control for other concurrent events, we follow Du and Yu (2020) and check for other major news regarding the firm and eliminate the firm-date observations for which the ESG release dates are within the five-day window around the release of another major corporate event

(M&A, earnings press release or the issue of the 10-K). After these final steps, our sample includes 1,667 reports for 416 unique firms.

3.1.2. Pre-processing of the ESG reports

The raw files in our sample are typically in a PDF format.[8] Therefore, to analyze the text contained in the reports, we used the Tesseract Optical Character Recognition (OCR) engine in R to convert the PDF files to TXT format.[9] Based on these documents, various pre-processing steps on the text are implemented, namely, non-English characters (punctuation and numbers), non-alphanumeric symbols, numbers, repeated space character, one letter words, letters attached with special characters (taking care not to delete words such as o2 and co2), and then lemmatizing the words into their root forms. We then remove stopwords, which are frequent words with little information value, such as "the", "is", "have", "of", "a", "thus".[10] Finally, we follow Huang et al. (2017) and remove company names, websites and tickers to prevent the sentLDA from identifying firms as topics. All these steps increase the interpretability of the topics identified and the quality of the sentLDA output.

3.2. Topic modeling algorithm – Implementing sentLDA for knowledge extraction from ESG reports

To extract information from the ESG texts, we apply a specific iteration of Latent Dirichlet Allocation (LDA) (Blei et al., 2003), an unsupervised machine-learning-based tool that researchers widely use in the information retrieval literature (see e.g., Bellstam et al., 2020; Kaplan and Vakili, 2015; Brown and Hillegeist, 2007; Hoberg and Lewis, 2017). Specifically, we apply the Sentence Latent Dirichlet

[8] In a very few instances, the ESG report was in a HTML file. In such cases, we copy pasted the text in a TXT file.

[9] This step removes all pictures and tables.

[10] The list contains *deictic* words, which are words that cannot be fully understood without additional contextual information. However, using a pre-defined list of stopwords can be problematic as it cannot be readily generalized across knowledge domains. Therefore, instead of using a pre-defined list of stopwords (Nguyen, 2014), we remove the words with a lower conditional entropy based on the following equation:

$$p(d|w) = \frac{p(w,d)}{p(w)} = \frac{n(w,d)}{n(w)},$$

with $n(w,d)$ defined as the number of occurrences (tokens) of a word w in document d (Gerlach et al., 2019).

Allocation (sentLDA) proposed by Bao and Datta (2014). As output, while LDA provides, for each document, the overall probability distribution of topics, sentLDA assigns a single topic to each sentence within the document. As such, sentLDA additionally imposes a constraint that all words in a sentence represent a single topic, i.e. a sentence represents a single idea. Including this sentence boundary allows for a detailed topic mapping within each document and the estimation of which topic is discussed in a given sentence. For our study, this information is especially relevant as our objective is to identify the various topics discussed in each ESG report, and not to classify the documents.

Compared to focus on words in the traditional LDA (Blei et al., 2003), the sentLDA proposed by Bao and Datta (2014) presents several advantages. The original LDA by Blei et al. (2003) is based on the bag-of-words assumption, which states that the order of words in a document does not matter. However, this word-level analysis imposes limitations, as intuitively, a sentence boundary conveys which words should be grouped into the same topic. In line with this intuition, Bao and Datta (2014) evidence that the proposed sentLDA method outperforms competing unsupervised methodologies and provides more meaningful topics, when the objective is to map the semantic composition of documents. Accordingly, the sentLDA model relies on a few assumptions. First, the model assumes that there is a collection of K topics in a given document d and that the list of words in each topic follows a Dirichlet distribution, $\beta_k \sim \text{Dirichlet}(\eta)$. Second, for each document, sentLDA considers a vector of topic proportions drawn from a Dirichlet distribution $\theta_d \sim \text{Dirichlet}(\alpha)$. From these assumptions and a few learning parameters, the sentLDA model categorizes words in each document into K number of topics and assigns a topic to each sentence in a document.

The critical inferential problem is the computation of the posterior distributions of the two hidden variables θ (the topic proportions for each document) and z (the topic assigned to each sentence), based on the model parameters and the observed documents. As the distribution is intractable (Blei et al., 2003), we follow Bao and Datta (2014) and use the Variational Expectation Maximization learning algorithm to approximate the posterior distributions. As such, sentLDA organizes the words

in a collection of documents into the specified number of topics and defines each document as a collection of topic-sentences (for further details on the computation, see Bao and Datta, 2014).

We run the sentLDA algorithm on the corpus of cleaned ESG reports to generate a list of topics. As a probabilistic model, sentLDA assigns weights corresponding to each topic to every word in the vocabulary. Thus, the topics are defined as sequential lists of words based on the topic-weights assigned. sentLDA then allocates each sentence in the ESG report to the highest weighted topic based on the words it contains. The output can be described as follows:

$$Topic_k = TopicWeight_k \cdot Word_z, \tag{1}$$

$$TopicAssignment_S = k | \max_k \sum_{w=1}^{W} TopicWeight_k \cdot Word_w, \tag{2}$$

where k represents the k^{th} topic, z is a word in the total vocabulary, s is a given sentence in a report, W represents the total number of words in sentence S, and w represents the w^{th} word in sentence S. Therefore, for every report, sentLDA provides a vector output of K elements, which describes the distribution of the per-sentence topic allocations. We detail the implementation of sentLDA in Section 4.

3.3. Dependent variables – Investors' reaction and information asymmetry around the ESG report's release

Our models test the market impact ($MarketImpact$) of the thematic content of ESG reports. To assess the market impact, we focus on two variables: (i) investors' reaction and (ii) information asymmetry, both around and after the release of the ESG reports. First, our measure of investors' reaction is the absolute cumulative abnormal return, which measures whether investors react to the release of the ESG report. Second, we proxy information asymmetry with the price impact measure introduced by

Amihud (2002), which measures the illiquidity in the stock market.[11] All our dependent variables are winsorized at the one and 99 percent levels.

3.3.1. Investors' reaction

To assess investors' reaction to the thematic content of ESG reports, we measure the absolute cumulative abnormal return around the release of the ESG report. $|CAR_j|$ represents the absolute cumulative abnormal return around the release date of firm j's ESG report. We consider three periods around the release date: (i) $|CAR_{[-1,+1]}|$, (ii) $|CAR_{[-2,+2]}|$, and (iii) $|CAR_{[+1,+60]}|$. Considering several windows allows us to integrate any leakage of information the days before the report is released, as well as any under/overreaction over the days following the release. For all the periods, abnormal returns are computed based on the three-factor Fama and French model calibrated on the estimation window that starts 315 days before the announcement and ends 62 days before that release date.[12] Furthermore, as we take the absolute values of CAR, the models examine the relationship between the explanatory variables and the price fluctuations around the ESG report's release date, irrespective of the direction of the price movements.

3.3.2. Information asymmetry

Information asymmetry is defined when one party in a transaction has more information than the other (Ross, 1977). Typically, managers have more information about the company than outsiders. Therefore, information asymmetry is prevalent when there are many shareholders or stakeholders, and a high level of complexity in the firm. Although information asymmetry is a pervasive phenomenon in corporations, there is no easy way, nor a definitive way to measure it. Prior literature focuses on trading models (Mohanram and Rajgopal, 2009; Roll et al., 2009). Other work use analysts' coverage,

[11] We also run tests with the bid-ask spread as a robustness check.

[12] We also consider the industry-adjusted CAR ($IndCAR$) to correct for industry heterogeneity. $IndCAR$ is the industry-adjusted return over the days relative to the date of the news. This analysis provides qualitatively similar results. They are available upon request.

or analysts' forecast accuracy (Lee and So, 2017).

To estimate information asymmetry, we rely on the proxy of illiquidity set forth by Amihud (2002) and Goyenko et al. (2009). The illiquidity measure is defined as firm j daily price response divided by the trading volume. If the thematic content of ESG reports reduces information between a firm's insiders and outsiders, the identified topics should influence the transaction costs, and therefore reduce stock illiquidity. Similar to the absolute CAR, we consider three periods (t) around the release dates: (i) $AMH_{[-1,+1]}$, (ii) $AMH_{[-2,+2]}$ and (iii) $AMH_{[+1,+60]}$. For each period, AMH_j is defined as the sum of the daily ratio of absolute returns on dollar volume and represents the information asymmetry around the release date of firm j's ESG report.

3.4. Regression Model

To study the market's reaction to ESG topics, we use the topics obtained from the sentLDA model as input variables in a multi-factor model. The benchmark model is the traditional approach consisting of a linear model in which the market impact at the release of the report are estimated using a set of covariates used in prior literature. This set of covariates is detailed below (see Section 3.4.1) and relates investors' reaction to the ESG reports' style features, firm-specific controls, as well as year- and industry fixed-effects:

$$MarketImpact_j = \alpha + \beta \cdot \text{ESG Report Controls}_j + \beta \cdot \text{Firm Controls}_j \tag{3}$$

$$+ \beta \cdot \text{Industry}_j + \gamma \cdot \text{Year}_j + \varepsilon_j,$$

where ε_j are firm-clustered standard errors, corrected for heteroskedasticity. In our regression model, our dependent variable ($MarketImpact$) is proxied by the price impact measure (AMH) and investors' reaction around the ESG report's release ($|CAR|$) for the same set of empirical tests. As a

measure of thematic content in ESG reports, we include the topics estimated based on the sentLDA ($Topic_{k,j}$) to Equation 3 and define the following model:

$$MarketImpact_j = \alpha + \delta_k \cdot \sum_{k=1}^{K} Topic_{k,j} + \beta \cdot \text{ESG Report Controls}_j + \beta \cdot \text{Firm Controls}_j \quad (4)$$

$$+ \beta \cdot \text{Industry}_j + \gamma \cdot \text{Year}_j + LAMBDA_j + \varepsilon_j,$$

where ε_j are firm-clustered errors, corrected for heteroskedasticity. $Topic_{k,j}$ is defined as the number of sentences referring to topic k in the ESG report of firm j. To test the incremental information value of ESG reports' thematic content, we then compare the two regressions by conducting an ANOVA test comparing the Residuals Sum of Squares (RSS).

Hypothesis 2 examines whether there exists a curvilinear relationship between topic diversity and the market impact at the release of ESG reports. We define the following model:

$$\begin{aligned} MarketImpact_j = &\alpha + \beta \cdot TopicDIVERSITY_j + \beta \cdot TopicDIVERSITY_j^2 \\ &+ \beta \cdot \text{ESG Report Controls}_j + \beta \cdot \text{Firm Controls}_j \\ &+ \beta \cdot \text{Industry Controls}_j + \gamma \cdot \text{Year Controls}_j + LAMBDA_j + \varepsilon_j, \end{aligned} \quad (5)$$

where topic diversity ($TopicDIVERSITY$) is based on the Shannon diversity index. The index is calculated as follows:

$$TopicDIVERSITY_j = -\sum_{k=1}^{30} P_{k,j} \cdot lnP_{k,j}, \quad (6)$$

where, for firm j, P_k is the percentage of sentences in a ESG report dedicated to topic k out of 30 topics (if $P_k = 0$, lnP_k is set to 0) (Shannon, 1948). Intuitively, the index quantifies the uncertainty in predicting the identity of a randomly chosen entity from a given data (for examples of application, see Reguera-Alvarado et al., 2017; Campbell and Mínguez-Vera, 2008). We expect a positive coefficient for $TopicDIVERSITY$ and a negative coefficient for $TopicDIVERSITY^2$ for the investor reaction variables ($|CAR_j|$). On the contrary, for the information asymmetry variables (AMH_j), we expect a negative coefficient for $TopicDIVERSITY$ and a positive coefficient for $TopicDIVERSITY^2$.[13]

3.4.1. Control variables

We define a large set of control variables from prior studies, which exmaine factors that affect the market's reaction around the release of ESG reports (Bushman and Smith, 2003; Choi and Richardson, 2016; Schoenfeld, 2017; Jones, 2007; Du and Yu, 2020). We first control for firm-specific variables, such as firm size and book-to-market ratio. We control for firm size ($SIZE$) as larger firms often have more visibility and receive greater stakeholder pressure to engage in ESG disclosure (Bushman and Smith, 2003). $SIZE$ is the natural logarithm of total assets at the end of the current year. We then control for the firm's book-to-market (BTM), which is defined as the ratio of the firm's book value divided by the firm's size ($SIZE$). In addition to these control variables, we include year and industry fixed-effects (Clarkson et al., 2020; Li, 2006).

We then include four variables to control for a firm's concurrent and past performance. We first control for the firm's return on assets (ROA), which is measured as the income before extraordinary items divided by total assets at the end of the previous year. We then control for the earnings surprise ($SURPRISE$), which is defined as the spread in the analysts' consensus EPS forecast and the actual EPS, divided by the stock price at the end of the previous year. The consensus analyst forecast is defined as the average of the most recent analyst forecasts for the firm in the current year. We also

[13] We expect ESG reports to decrease information asymmetry at first. After a threshold, we expect the information asymmetry to increase again.

include a dummy variable $EPSdecline$ which equals one if the EPS declined in the current year, and zero otherwise. Finally, we add the momentum variable ($MOMENTUM$), which is defined as the one-year cumulative abnormal return from the [-375,-10] trading day window, where the event is the ESG reporting date.

There is also evidence in prior literature that the market's reaction to corporate information depends on a firm's financing needs, stock price liquidity, research and development and firm information environment (Schoenfeld, 2017; Jones, 2007). We include financing activities (FIN), stock liquidity ($LIQUID$), research and development intensity ($R\&D$), analyst following (NoA) and operating complexity ($GEOsegments$ and $BUSsegments$) as additional control variables. Specifically, we calculate FIN as the firm's net debt amount and raised equity capital (i.e., the sum of net sale of common and preferred shares and net long-term debt issuance) during the year scaled by total assets at the year-end. $LIQUID$ is the number of shares traded divided by the total number of shares outstanding for the same year. $R\&D$ represents to the research and development expenses deflated by sales during the year. NoA is measured as the natural logarithm of one plus the number of financial analysts at the end of the same year. Finally, following Huang et al. (2013), we control for the operating complexity of the firm and include the logarithm of the number of geographical ($GEOsegment$) and business segments ($BUSsegment$). Furthermore, we include two variables to control for information uncertainty. In particular, we control for financial leverage (LEV) because firms with constrained financial resources are likely to be more risky and, therefore, lead to more volatility (Choi and Richardson, 2016). LEV is calculated as total debt (short-term plus long-term debt) divided by total assets at the year-end. We also control for earnings uncertainty $VOLATILITY$, which is computed as the standard deviation on ROA for a period of five years.

We also include a series of text-based variables from the ESG report. In addition to controlling for the length of the ESG reports with the logarithm of the total number of words in the document (WC), we follow Du and Yu (2020) and control for the ESG report's readability (FOG). To measure

readability, we use the FOG index as developed by Robert Gunning (Strong, 1986).[14] FOG is measured as the average number of words per sentence added with the percentage of words per sentence. The FOG index estimates the years of schooling that a reader of average intelligence would need to read and understand the text. We also include the tone of the ESG report, $TONE$, which is defined as the spread in the number of positive and negative words, scaled by the total number of words in the ESG report. To identify the positive and negative words, we follow Du and Yu (2020) and use the Loughran and Mcdonald (2011)'s lists of positive and negative words. However, the concern with such lists is that they are customized for the analysis of 10-K documents. We therefore also include an ESG emphasis measure based on the list of words defined by Baier et al. (2020). We define this ESG emphasis measure ($ESGemph$) as the sum of the ESG-specific words from the Baier et al. (2020) list, divided by the total number of words in the ESG report. Following Hope et al. (2016) and Loughran and McDonald (2014), we further include the specificity and uncertainty measures of the text. To estimate the level of specificity, we use a Named Entity Recognition(NER) automated tool (Apache NER) to identify the terms mentioning named entities, such as person, organization, location, percentage, monetary value and date. The specificity of the text, $SPECIFIC$, is measured as the total number of unique named entities detected by NER. The $UNCERTAIN$ variable indicates the percentage of uncertain words in the text based on Loughran and McDonald (2014)'s lists of uncertain words. Finally, we follow Bozanic et al. (2018) who show that disclosures with more forward-looking information are associated with a stronger market reaction. Based on the method adopted by Henry (2008); Henry et al. (2021) and Athanasakou and Hussainey (2014) to identify forward-looking sentences, we include the proportion of forward-looking sentences as a control variable ($FWDLOOK$).

[14] Loughran and McDonald (2014) argue that the FOG index is a weak proxy to assess the readability of a text. They show that the many words considered as complex, may, in fact, be easy to understand in a specific context (e.g. for investors or financial analysts). They therefore recommend using the size of the file as a readability measure. They argue that the longer documents have higher information processing costs and are more difficult to read. For robustness, we replace our FOG proxy for readability by the file's size. Our results remain qualitatively similar.

3.5. Summary statistics

Panel A of Table 1 reports the distribution of the sample by year. Consistent with the trend that a growing number of firms engage in ESG reporting in the recent years, the number of ESG reports increased from 25 in 2007 to a maximum of 262 in 2019. Panel B of Table 1 reports the distribution of the sample by industry based on Barth et al. (2001) industry classifications. The durable manufacturing industry contains the largest number of reports and firms, accounting for 19.916% of all the ESG reports and 18.269% of the firms. The Computers industry has the second largest number of ESG reports and firms. On the other hand, the Mining and Construction industry has the smallest number of ESG reports in our sample, with 1.500% of all the ESG reports and 1.923% of the firms in the sample.

< Insert Table 1 about here >

Before testing our research questions, we first examine whether the stock market reacts to the release of ESG reports by looking at the absolute abnormal returns. Panel A of Table 2 provides the mean and median of cumulative absolute abnormal returns (in percentage). There is a significant investor reaction around the release of the ESG report. The average cumulative absolute abnormal returns during the 1-day and 2-day event windows ($|CAR_{[-1,+1]}|$ and $|CAR_{[-2,+2]}|$) are equal to 1.387% and 1.873%, respectively. All averages are significant at a 99% confidence interval based on two-sided sample t-tests.[15] Our results also show that the release of an ESG report has a long-term impact on the stock price. In fact, we find that the long-term cumulative absolute abnormal returns ($|CAR_{[+1,+60]}|$) is significant at a 99% confidence level and equal to 2.829%. The significant cumulative absolute abnormal returns confirm the results of Du et al. (2017) and suggest that investors do respond to ESG reports, supporting the fact that ESG reports contain information value. To be complete, Table 2 also provides the distribution of the signed cumulative abnormal returns. The

[15]The use of non-parametric Wilcoxon signed-rank tests provide us with consistent results.

average cumulative signed abnormal stock returns across all event windows is insignificant and equals -0.02% for the $CAR_{[-1,+1]}$ and 0.1% for $CAR_{[-2,+2]}$. The $CAR_{[+1,+60]}$ is equal to -0.02%.

Panel A also provides summary statistics for the impact of an ESG report's release on stock illiquidity. The Amihud illiquidity measure $AMH[-1,+1]$ equals 0.374, while $AMH[-2,+2]$ equals 0.222, both significantly different from zero at a 99% interval. In addition, $AMH[+1,+60]$ equals 0.103, which is also different from zero at a 99% confidence level. Overall and in line with Du et al. (2017) and Du and Yu (2020), the insignificant means of signed CAR, together with the significantly positive means of absolute CARs and Amihud illiquidity measures, strongly suggest that there exists a large market impact around the release of ESG reports across the sample firms.

< Insert Table 2 about here >

Panel B of Table 2 reports the descriptive statistics of our control variables. The average ESG report contains 13,999 words and the average FOG index is 23.753, which falls into the category of difficult to read. The FOG index is slighthly higher than Du and Yu (2020) who find a score of 18.68.[16] The mean tone ($TONE$) is equal to 1.065%, which indicates that the ESG reports tend to be on the positive side. We also find that the ESG emphasis in the report is 6.374%. The disclosure also contains 5.281% of forward-looking sentences, while the number of uncertainty words equals 0.477%, out of the total number of words. The average financial performance of firms in our sample (ROA) equals 6.605%, while the ESG score of firms in our sample equals 5.297 ($ESGscore$), which shows a great variation with a standard deviation of 5.400. In addition, the average firm in our sample has a market capitalization ($SIZE$) of $72,455 million, 27.401% leverage (LEV) and 3.530% volatility ($VOLAT$), and is followed by 15.711 financial analysts (NoA). Finally, 6.239% of firms observe a decrease in EPS ($EPSdecline$). The average number of business ($BUSsegment$) segments is 1.495 and geographical segments ($GEOsegment$) is equal 1.746.

[16] In comparison, Loughran and McDonald (2014) find a similar FOG index of 18.89 for 10-K reports, which means that ESG reports tend to be more difficult to read than 10-K reports.

Table 3 presents the Pearson correlations of the main variables for the full sample. The correlation between the firm's financial and ESG performance is equal to 12.5% and significant at a 99% level, which is consistent with prior literature (Hasan et al., 2016; Luo and Du, 2015; Dhaliwal et al., 2010, 2012). We also find that investors react more to the release of ESG reports when their stock is more liquid ($LIQUID$), when they are followed by more financial analysts (NoA) and when their operating complexity is larger ($BUSsegment$ and $GEOsegment$). We also find that the information asymmetry (AMH) is lower for firms that are larger ($SIZE$), more profitable (ROA), highly leveraged (LEV), more liquid ($LIQUID$) and followed by more financial analysts (NoA). In addition, for $AMH[-2,+2]$ and $AMH[+1,+60]$, we find that firms with a lower level of information asymmetry tend to have longer ESG reports (WC). We also conclude that the level of information asymmetry is lower for firms that disclose more forward looking statements ($FWDLOOK$) in their ESG reports.

< Insert Table 3 about here >

4. Topic modeling of ESG reports

Before we analyze the thematic content of ESG reports, we first discuss the implementation of sentLDA, which details the selection of number of topics, and the topic labeling and validation process.

4.1. The number of topics

While sentLDA is an unsupervised method, the number of topics needs to be specified by the researcher. Intuitively, the number of topics of the model affects the interpretation of the results. Defining a number that is too low can lead to topics that are too noisy and broad. On the contrary, setting a number of topics that is too high may lead to topics that are difficult to interpret. Following Zhao et al. (2015), rather than adopting a trials and errors approach to evaluate the most appropriate

model, we use a cross-validated perplexity measure. The perplexity score assesses topic models based on how well a model performs when predicting unobserved documents (Bao and Datta, 2014).[17] Following prior literature (Blei et al., 2003; Huang et al., 2018), we first compute and plot the perplexity scores of the sentLDA models for different numbers of topics ranging between 1 and 100. The perplexity scores diminish with the number of topics, with an improvement that is marginally decreasing. The decrease in perplexity is significant until the number of topics exceeds 30. As a result, we choose 30 as the number of topics in our analyses. We provide a figure illustrating the perplexity scores at different number of topics in the Appendix, Figure A.1. We note however that the optimal number of topics largely depends on the specific samples used and the semantic coherence of the generated topics. For instance, in other studies, Jaworska and Nanda (2018) define 80 topics for their sample of 317 ESG reports by oil companies between 2000 and 2013, while Goloshchapova et al. (2019) specify 50 topics to analyze their 1,122 ESG reports, which consist of European firms between 1999 and 2016. Therefore, in addition to using the perplexity score, we also manually compare the sentLDA outputs based on 50 and 80 topics. We find that the sentLDA output with 30 topics provides a set of semantically coherent topics without generating uninterpretable topics.

4.2. Validation of the sentLDA output

To enable semantic interpretation, we manually label the estimated topics. Thus, to ensure that the assigned labels effectively capturing human comprehension, we undertake a comprehensive procedure. We first generate a list of the highest weighted phrases and sentences for each topic. Specifically, we construct lists of 1,000 sentences per topic based on the weights assigned to their constituting words. Next, we sort the sentences by length and extract the middle tercile (334 sentences) as representative

[17] The perplexity score measures the ability of the LDA model estimated on a training sample to predict the word choices in the test sample. Specifically, for a testing sample (D_{test}) with M documents, the score is equal to:

$$Perplexity = \exp\left(\frac{-\sum_{d=1}^{M} \log p(w_d)}{\sum_{d=1}^{M} N_d}\right)$$

Accordingly, the score is monotonically decreasing in the likelihood of observing the testing data, given the model estimated from the training data (Huang et al., 2018).

sentences of typical length. We also extract the 20 most frequent bigrams (two-word phrases excluding stop words, numbers, and symbols) from the 334 mid-length sentences. These sentences are also sorted based on the cosine similarity between them. We then evaluate the semantic meaning of the top 20 bigrams and the top 100 mid-length sentences based on cosine similarity and assign descriptive labels to each topic.

To illustrate and as shown in Table 4, the topic with the highest-weighted words c(*climate*, *environmental*, *energy*, and *water*) is labelled as *Climate Change*. From evaluating the assigned topic sentences, we find that c(*climate - change*, *risks - opportunities*, *risk - management*, and *environmental - impact*) are the most frequent bigrams. Furthermore, a sample of the highest weighted mid-length sentence reads as follows: "ACE is committed to developing insurance products and risk management services that facilitate market-based solutions to environmental and climate-related issues." As such, these further evaluations provide support for the assigned label. Table 4 provides the word clouds representing the list and weights of the highest weighted words in each topic with their associated labels, while Table 5 provides the labels of each topic and briefly describes the 30 topics.[18] Tables A.1, A.2, and A.3 of the online appendix provide the lists of 20 highest weighted words, the 20 most common bigrams, and the 100 mid-length representative sentences for each topic. This form of evaluation is qualitative and is supported by several studies, including those by Hoberg and Lewis (2017), Dyer et al. (2017), and Brown and Hillegeist (2007).

< Insert Table 4 about here >

< Insert Table 5 about here >

Second, we evaluate the semantic coherence of the topics using *word intrusion* tasks designed by Chang et al. (2009). With the intrusion method, the subject is presented with six randomly ordered words. The subjects are asked to find the word that is out of place, or the intruder. When the set of

[18] We also observe that the sentLDA outcome recognizes the contextual nature of some words by assigning the same word to various topics. For instance, the word *water* is prominently related to the topics such as *Waste Management*, *Emission*, *Conservation*, which highlights the contextual nature of the word.

words minus the intruder makes sense together, the subjects should easily identify the intruder. For instance, for the list of words *library, shelf, page, pen, cover, knife*, the word *knife* is easily identified as the intruder since all the other words refer to a books. In contrast, for a set of words such as *house, doctor, car, cow, plate, pillow*, which lacks coherence, it is difficult to identify the intruder. Following Chang et al. (2009), for each topic, we select the five highest-weighted words and randomly select an intruder word from a pool of highest-weighted words from the other topics. All six words are then shuffled and presented to the subjects. We define model precision (MP_k) of the topic k as the fraction of subjects agreeing with the model:

$$MP = \frac{1}{K}\sum_{k}\frac{1}{S}\sum_{s}\mathbb{1}(i_{k,s} = \omega_k), \tag{7}$$

where $i_{k,s}$ is the intruder word selected by subject s among S subjects, ω_k is the true intruder word, and $\mathbb{1}(\cdot)$ is an indicator function that equals 1 if $(\cdot)$ is true and zero otherwise. MP is simply the average of the corresponding MP over topics. We conduct the experiment in a group of 55 participants on the online platform Prolific. We present them with the list of words, including an intruder word, for each topic and ask them to select the intruder. We obtain an model precision (MP) of 55.062%, which is superior to the performance levels reported by Bao and Datta (2014). This result further supports the evidence for semantic coherence of the estimated topics. The experiment details and results are provided in the Appendix Table A.4 along with a boxplot of the response accuracy in Figure A.2.

As a third evaluation, we graphically examine the similarities between the topics based on the words they emphasize. Figure 1a displays a network graph where each weighted line represents the correlation between adjoining topics based on the weights assigned to each word (for ease of interpretation, correlations under 35% are not displayed). We observe several clusters. For instance,

topics *WaterConservation*, *Conservation* and *EnergyEfficiency* show notable degree of correlations, meaning that the sentences discussing conservation efforts include words relating to energy efficiency-focused information. Likewise, we find clusters between topics concerning the welfare of the workforce and safety, such as *WorkSafety* and *SafetyStandards*, and also between topics related to various corporate governance issues, such as *Transparency*, *Accountability* and *Compliance*. Other topics, such as *ClimateChange* and *EnvironmentalImpact*, also display a degree of overlaps across topics. Since our topic labels are discretionary, these linkages provide additional nuance and support for the interpretation of the topics.

< Insert Figure 1a about here >

Taken together, our evaluation methods suggest that the sentLDA algorithm provides a valid set of semantically meaningful topics that are reasonably coherent and interpretable by human judges. We describe the topics identified in our sample below.

4.3. The thematic content of ESG reports

Table 5 reports the summary statistics of the various topics variables. For ease of interpretation, we allocate the topics to each cluster of the ESG construct it belongs to. The *Environmental* cluster includes the following topics: *ClimateChange*, *Conservation*, *Emission*, *EnergyEfficiency*, *EnvironmentalImpact*, *Recycling*, *RenewableEnergy*, *Supplyhain*, *SustainableFarming*, *GreenTechnology* and *WaterConservation*. To form the *Social* cluster, we group the following topics: *Charity*, *CommunityRelations*, *DisasterRelief*, *Diversity*, *Education*, *Health*, *Nutrition*, *SafetyStandards*, *WorkSafety*. The remaining topics form the *Governance* cluster.

Overall, we find that *CommunityRelations* (29 sentences, 8.759%) is the most frequent topic in ESG reports, followed by *Diversity* (23 sentences, 5.086%). The least discussed topic is *Managemen-*

tApproach, with an average of five sentences (1.313%). Interestingly, although the main objective of ESG reports is to demonstrate firms' actions and activities in relation to society and environment, we note that there still is a proportion seven sentences (2.055%) of the corpus that focuses on financial concerns and achievements. Hence, and contrary to the wider assumptions (Breeze, 2013), ESG reports are not just about ESG activities; they also communicate about issues related to corporate financial performance, strategy and management. The standard deviations of most topics are however substantial in comparison to the mean, indicating much variability in the information composition within ESG reports.

If we aggregate topics in each ESG pillar, we find that the average ESG report displays a fairly equal distribution of topics accross each E, S and G pillar. The average report discusses environmental topics with an average of 149 sentences (34.08%), then social issues (148 sentences, 33.86%), followed by governance topics (126 sentences, 30.31%). Examining the diversity of topics that compose ESG reports, we confirm that the content of ESG reports is fairly distributed across the E, S and G components. Since the measure ranges from zero to 3.40 (log of total number of topics),[19] our sample average of 2.57 indicates that the topic content of ESG reports across each ESG pillar is fairly diverse.

We further examine which topics are more likely to appear in the same ESG report. Figure 1b reports a network graph with each weighted line indicating the degree of co-occurrence between topics within documents (for ease of interpretation, correlations under 35% are not displayed). Interestingly, Figure 1b demonstrates that topics belonging to each core ESG area, which include *Emission* for the environment, *Diversity* for the social pillar and *Transparency* regarding governance issues, occupy a more central area in the network. We also find that a variety of less-discussed topics gravitate around those leading themes, which highlights that the average report discusses a large array of smaller topics, such as *Recycling*, *CustomerService* or $EnergyEfficiency$. We observe, for example, that the topic *WorkCulture* regularly appears with sentences concerning *Diversity* (correlation:

[19] The upper bound indicates that the topics are equally distributed in the ESG report.

68.755%), while sentences discussing the conservation of water (*WaterConservation*) and climate change (*ClimateChange*) are often found in the same reports (correlation: 63.299%). Among all topics, the two that appear most frequently together are the topics *SustainableFarming* and *Nutrition*, which concerns description of food production and consumption (correlation: 77.254%). Other topics tend to differentiate themselves. For instance, topics such as *Financial* and *Health* appear to be less correlated with the leading ESG topics.

< Insert Figure 1b about here >

Despite a fair distribution of the ESG topics across the reports, we find that reporting practices in ESG reports have undergone substantial changes over time. In order to understand the evolving nature of ESG practices, we report how each cluster of the ESG construct (Environment, Social and Governance) evolves over the years. Figure 2 shows a number of major topic shifts, pointing to changing practices in ESG reporting. In particular, we find that, after a peak in 2012, the pillar regarding the environment is steadily given less prominence in ESG reports. By 2020, environmental concerns constitute the second largest topic of discussion, with an average of 30.151% of the sentences. This decrease in environement-related sentences is compensated by the significant increase in the discussion of social-related topics. Although Figure 2 shows that the reference to social matters decreases between 2007 and 2015, topics in the social pillar have significantly gained in importance since 2015. Starting at 42.223% in 2007, topics discussing the workforce, employee benefits, gender issues or communities decreased to 36.354% to increase again to 39.133% in 2020. The upward trajectory in recent years supports the claim that social commitment of a company is gaining greater importance for investors (Breeze, 2013). Several studies in prior literature report a positive influence of firms' social activities on performance through innovation, operational efficiency, increased trust and a better brand image (Hull and Rothenberg, 2008; Arendt and Brettel, 2010; Saha et al., 2020).

< Insert Figure 2 about here >

Despite the increasing consensus of the positive impact of governance initiatives on firm financial performance (Bhagat and Bolton, 2009; Paniagua et al., 2018), topics regarding a firm's compensation methods, relationships between employers and employees, training opportunities, leadership structure and potential conflicts of interest seem to receive less attention in ESG reports. Figure 2 shows that governance topics not only receive less attention than their environmental and social counterparts, but also that the proportion of governance topics remains steady over our sample period. In fact, the discussion of ESG topics oscillate between 26.356% in 2007 and 28.851% in 2013. Since, the average number of sentences regarding governance remained at an average of 30.452% between 2013 and 2020.

5. The informational value of the thematic content of ESG reports and topical diversity

5.1. Thematic content's informational value

Based on the topics identified above, we now examine Hypothesis 1 and investigate the informativeness of the thematic content of ESG reports, by applying the multivariate model defined in Equation 4. To account for outliers, extremely long and short ESG reports, we winsorize the topic variables at 1^{st} and 99^{th} percentiles. We first provide the results from the base models (Equation 3) in Table 6, which include only the controls as explanatory variables. We start with Panel A, which focuses on investors' response to the release of ESG reports. We find that a number of the variables are significantly related to the CAR around the reports' release date and in the long term. For instance, we find that the variables $LIQUID$ show consistent positive relationships with investors' reaction ($|CAR|$) across the different time windows. We also find some evidence that the long term $CAR[+2, +60]$ is negatively correlated with the number of words (WC), which suggests that the longer the ESG report, the lower investors' reaction. We also find that a that the [+1,+60] investor reaction ($|CAR[+1, +60]|$) is significantly and positively associated with the number of uncertainty words ($UNCERTAIN$) and

forward-looking statements ($FWDLOOK$). Additionally, we find that investors respond significantly more as the geographical complexity of the firm increases ($GEOsegment$). Furthermore, we find that the firm's ESG score is only slightly correlated with investors' reaction at the [+1,+60] horizon. The coefficient is negative and statistically significant at a 90% confidence level, which indicates that investors react less to the disclosure of ESG reports of firms whose ESG score is positive. Finally, we find that some of the control variables, namely volatility ($VOLAT$), firm size ($SIZE$), ESG report's readability (FOG) and leverage (LEV) are not statistically significantly in Model (1), (2) and (3). Note that since these dependent variables look at share price movements in absolute terms, the significant negative relationships should not be interpreted as associations with adverse investor reactions, but as weaker share price fluctuations around the release date.

< Insert Table 6 about here >

Panel B of Table 6 focuses on the Amihud information asymmetry measure (AMH). Compared to Panel A, we find that a larger number of statistically significant variables. For instance, we find that the firm's leverage ratio (LEV) is consistently negatively correlated with the level of information asymmetry around the ESG report's release. We also find several of the control variables reduces the information asymmetry, including the size of the firm ($SIZE$), stock's liquidity ($LIQUID$), the number of financial analysts (NoA), R&D spending (RD), and the number of uncertainty words in the ESG report ($UNCERTAIN$). $AMH[-1,+1]$ shows a positive coefficient for firms with higher financial activity (FIN) and for firms with more geographical segments ($GEOsegment$). Our results also point towards a reduction in information asymmetry for firms with a good accounting performance (ROA).

We then report in Panel A of Table 7 the test statistics comparing the base model's performance (Equation 3) to the model that includes both the topics and the controls (Equation 4). The models' performance measures are derived from ANOVA tests comparing the Residuals Sum of Squares (RSS).

Our findings show that the thematic content of ESG reports contains incremental information value to explain investors' reaction around the release of ESG reports. For Model 1 with $|CAR_{[-1,+1]}|$ as the dependent variable, we find that the estimated RSS significantly decreases by 2.80%, suggesting an improvement in the model performance. This decrease is significant at a 95% confidence level. For Model 2, we find a reduction in RSS by 2.34%, which is significant at a 90% confidence interval, while Model 3 shows a highly significant improvement in the model fit of 3.53%. Collectively, these results suggest that content-based information drawn from ESG reports contains incremental information value beyond what can be achieved by the ESG report-specific, firm-specific and other fixed effects identified in prior literature. We also consider the impact of the thematic content in relation to the information asymmetry variables ($AMH_{[-1,+1]}$, $AMH_{[-2,+2]}$ and , $AMH_{[+1,+60]}$), we obtain similar results. Model (3), (4) and (5) show a significant decrease in the models' RSS by 3.04%, 3.16% and 3.18%, respectively. Overall, this evidence highlights that investors significantly respond to the thematic content of ESG reports.

< Insert Table 7 about here >

Although the thematic content of ESG reports contains information value to explain investors' reaction and the level of information asymmetry around ESG reports' release, not all topics have equal information value. In fact, only limited themes in the report are effectively consequential. In Panel B of Table 7, we discuss our main findings regarding which topics are of value relevance for investors. We find that not all topics share the same degree of relevance to investors. While investors appear to deem some topics as important and credible, others seem superfluous with no meaningful impact on any measures of investors' reaction. In particular, regarding the environmental category, we find that investors tend to react to topics regarding climate change or energy efficiency. For instance, Model (2) shows a positive and significant coefficient for $ClimateChange$, $Conservation$, $SustainableFarming$ and $WaterConservation$ topics. Our models also highlight other topics that are of importance to

investors. For instance, we find that the description of supply chain (*SupplyChain*) leads to a positive investor reaction in Model (1) and (3), while investors' reactions are diminished by topics concerning compliance (*GreenTechnology*) in Model (2) and (3). Yet, it is interesting to highlight that topics regarding recycling (*Recycling*), environmental impact (*EnvironmentalImpact*), renewable energy (*RenewableEnergy*) or gas emission (*Emission*) do not statistically influence investors' reaction at the release of the report. Turning to the impact of topics on the level of information asymmetry around ESG reports' release, we find similar results. We find that the topic *ClimateChange* and *Recycling* tend to reduce the level of asymmetry between insiders and outsiders. On the contrary, we find that discussing topics regarding *GreenTechnology* increases the level of information asymmetry.

We observe other trends for the social category. In contrast to the importance of environmental topics, fewer themes are of value relevance for investors in the social category. In fact, we find only two significant topics (*Nutrition* and *Education*), which appear to reduce investor activity. Moreover, several topics, including that concerning firm's charitable activities (*Charity*) or health at the work place (*Health*), do not significantly influence investors' reaction. We find similar conclusions for the impact of social topics on the level of information asymmetry in Panel B.

While the social category has a limited number of significant topics, the category concerning the firm's governance has a larger set of significant topics, but with mostly negative signs. This evidence indicates that investors react significantly less to ESG reports when they discuss specific topics on the firm's transparency or accountability (*Transparency*, *Accountability*). Regarding the Amihud measure, we find that the topics *Accountability*, *Collaboration*, *CustomerService* or *WorkCulture* have a statistically positive coefficient, which indicates that discussing such topics significantly increases information asymmetry around the release of the ESG report.

Overall, our results highlight the fact that there are substantial differences in the value relevance of the topics discussed in the report and that investors tend to react more to ESG reports that discuss environmental matters, while they tend to find the social and governance pillars less informative.

Specifically, including environmental topics alone increases the R-square by 13.595% (from 0.068 to 0.077), while adding the topics regarding social issues to the model decreases the adjusted R-square by 2.380% (from 0.068 to 0.066). Governance topics, however, have a lower economic impact as they improve the model's adjusted R-square by 6.768% (from 0.068 to 0.072).[20] These results are of great interest as they confirm Hypothesis 1 that investors are influenced by *what* information is provided in the ESG report. In addition, while we find that the thematic content of ESG reports helps explain the level of information asymmetry, the impact depends on the topic discussed. We show that information asymmetry decreases with environment-related topics, while social topics have a marginal impact. On the contrary, governance-related topics tend to significantly increase the level information asymmetry.

5.2. The impact of topic diversity on the ESG report's informativeness

Extending on these results, Hypothesis 2 investigates whether the diversity in topics increases the informativeness of ESG reports. We expect to find a curvilinear relationship between topic diversity and the market impact around the release of the ESG report. We run Equation 5 and report the results in Table 8. Panel A first reports the results for the impact of topic diversity on investors' reaction around the release of the report. In line with Hypothesis 2, we find that for Model (1) that there is a positive relationship between $TopicDIVERSITY$ and the absolute three-day cumulative abnormal return around the release of the report ($|CAR_{[-1,+1]}|$). The coefficient is significant at a 95% confidence level. We also find that this relationship is curvilinear, with a significant and negative coefficient for the variable $TopicDIVERSITY^2$, which is significant at a 99% confidence level. We however find no significant relationship for $|CAR_{[-2,+2]}|$ and $|CAR_{[+1,+60]}|$. The results for the impact of ESG reports' release on information asymmetry are substantially stronger. In Panel B, we test the impact of topic diversity on the information asymmetry around the ESG report's release. We find that the coefficient for the variable $TopicDIVERSITY$ is negative for all models and is significant

[20] That is, including the governance-related topics helps explain investors' reaction by 6.768%.

at a 99% confidence level. This negative coefficient indicates that, as the diversity of topics in ESG reports increases, the information asymmetry around the release of the report decreases. However, while the informativeness of ESG reports increases with diversity, additional topics, after a threshold, do not contribute to the information value of the document, which leads to a convex relationship between topic diversity and information asymmetry around the ESG report's release date. These findings suggest that a mere reporting of a catalog of generic ESG topics leads to a weaker market response or a lower reduction in information asymmetry, evidencing investors' preference for more focused ESG reports.

< Insert Table 8 about here >

6. Additional Analysis

Our first results highlight that there exist clear patterns in the topics discussed in ESG reports, along with significant temporal shifts in managers' focus on environmental, social or governance issues. Another important take-away is that the content of such reports is informative and helps explain the market impact of ESG reports around and after their release date. Yet, we find that a limited number of topics lead to a significant stock price reaction at the release of the ESG report. In fact, most topics lead to an insignificant change in stock prices. This evidence strongly contrasts with prior literature, which assumes all topics to be of equal relevance to investors (Dhaliwal et al., 2012; Du and Yu, 2020). We now extend our findings to examine whether: (i) there has been any noticeable change in the market impact of ESG reports' thematic content throughout our sample period, (ii) these changes vary across industries and (iii) the thematic content helps explain future firm value and ESG performance.

6.1. The value relevance of ESG issues over time

As shown in Figure 2, the reporting of ESG practices is a dynamic construct, which has undergone significant changes across our sample period. We can therefore expect the value relevance of ESG topics for investors to vary with time. To identify trends in the value relevance of ESG topics for investors, we divide our data period into following sub-samples, e.g. 2007-2010, 2011-2013, 2014-2016 and 2017-2020. We report our results in Table 9, providing the coefficients that are significant for each dependent variable in each sub-sample.

< Insert Table 9 about here >

Overall, our findings show that the market's reaction to different ESG activities changes over time. In Panel A, we focus on the impact of the thematic content on the share price reaction. While we find that the thematic content of the reports in our first sample (2007–2010) has limited statistical significance in reducing the RSS of our models, the topics in ESG reports increasingly become relevant in the later periods 2011–2013 and 2017–2020. In addition, although the evidence for model improvement is weaker in the last period 2017–2019, we still find that several topics remain significant. In particular, the topics $WaterConservation$ and $Accountability$ are rather consistently significant.

The results for the AMH variables are reported in Part B of Table 9. We observe significant relevance of the thematic content for the years 2007–2010 and 2014–2016. In the earliest period, we find that the topic $EnviromentalImpact$, $WaterConservation$ and $Collaboration$ has consistent significant positive coefficients and that the coefficient for the topic $ClimageChange$ and $CustomerService$ are consistently negative. However, the types of topics relevant in the later periods are different from the topics identified in the first period. Overall, such trends in the value relevance of ESG topics show that juggling the interests of the firm's shareholders with those of a wide stakeholder group is a challenge that requires constant attention.

6.2. The value relevance of ESG issues by industry

The topics highlighted as value relevant are, so far, generic across our sample and are not unique to a particular company's competitive positioning in an industry. We therefore examine how the relationship between ESG reports' thematic content and investors' reaction varies across different industry groups based on the classification of Barth et al. (Barth et al.).[21] In Table 10a and 10b, we report the significant topic coefficients from the regression results derived from industry sub-samples. We find various consistencies and notable differences in the value relevance of ESG topics across industries. For instance, particularly for chemicals, computers and textile industries, there are significant model improvements from including the topic variables in the CAR models. Moreover, in relation to the Amihud illiquidity variables, we find that the ESG topics improve model performance across several industries, including chemicals, durable manufacturers, extractive industries, financial institutions, pharmaceuticals and transportation. Among the relevant topics, we find in Table 10a that *ClimateChange*, *ManagementApproach* and *CommunityRelations* that are associated with *Environmental* and *Governance* topic clusters are consistently significant and inversely correlated with illiquidity, hence associated with lower information asymmetry. Furthermore, we find that some topics are particularly relevant in specific industries, such as $WorkCulture$ in retail, $CustomerService$ in food, *Emission* in extractive industries, and $EnergyEfficiency$ in the transportation industry.

< Insert Table 10a about here >

< Insert Table 10b about here >

6.3. Alternative valuation metrics – ESG performance and firm value

One of the main contributions of this paper is that it concentrates on the direct market impact of ESG reports' thematic content. To measure this impact, we rely on investors' reaction (CAR) and the

[21] Industries 'insurance and real estate', 'mining and construction' and 'others' are not included given the limited number of observations.

price impact measure (AMH) over various periods. This approach significantly departs from prior literature, which mostly focuses on the relationship between ESG reports on future firm value or ESG performance. To further confirm the information value of the thematic content of ESG reports, we re-run Equation 4 and consider alternative dependent variables from prior literature. We first look at the relationship between the ESG thematic content with the firm's future ESG performance. We then examine the relationships between the former and the firm's Tobin's Q at three different time horizons: t (ESG report's release fiscal year), $t+1$ and $t+2$.

Following Clarkson et al. (2020) who examine the impact of ESG reports' readability on ESG performance, we first examine whether the topical composition of ESG reports are indicative of a firm's future ESG performance (i.e. ESG score in year $t+1$). ESG scores are third-party evaluations of a firm's socially responsible engagement, quantifying their efforts to implement environmental, social and governance issues (see e.g., Dorfleitner et al., 2015). Following prior literature (Dhaliwal et al., 2012), we use MSCI ratings provided by MSCI ESG Research as the proxy for overall firm ESG performance. As reported in Table 11, we find that the ESG report's thematic content significantly predicts a firm's ESG score, suggesting a significant link between the documents' content and future ESG performance. Among the environmental topics, we observe that the topics $EnergyEfficiency$ and $EnvironmentalImpact$, $RenewableEnergy$ and $GreenTechnology$ are positively associated with ESG scores. The topic $Education$, $Health$ are also positively linked with the ESG score. Moreover, other topics, including $Collaboration$, $Achievements$ and $Accountability$, are also associated with ESG scores.

< Insert Table 11 about here >

Because each topic is likely to have a different impact on each category of the ESG score, we split the ESG in its three components (E, S and G). Our results remain consistent. Concerning the information value of thematic content, we find that including the ESG topics in model significantly

improves model performance in relation to all three individual scores. However, we find that fewer topics are associated with the governance score.

We then examine whether the topical content of ESG reports are relevant to explain firms' current and future financial performance. Given that disclosure, such as ESG reporting, are known to influence firms' Tobin's Q in the specific contexts of family-firms and small-sized firms (Nekhili et al., 2017; Wu et al., 2021), we proxy a firm's current and future performance with Tobin's Q in fiscal years t, $t+1$ and $t+2$. In Panel B of Table 11, we find that including the topic variables in the model significantly improves the performances of all three models. Moreover, we find that the topics *Accountability*, $WorkSaftety$, $SustainabileFarming$ are positively associated with firm value. However, topics such as $SupplyChain$, $Diversity$ and $Nutrition$ are negatively associated with the outcome variables.

7. Robustness checks

Our main findings that the thematic content in ESG reports has a significant market impact around the release of ESG reports completely align with our initial expectations. However we bear in mind that the exact quantification of this effect depends on the measurement of information asymmetry, as well as on the model specifications used. Therefore, we now test the robustness of our findings.

7.1. Bid-ask spread as additional information asymmetry proxy

We first consider an alternative information asymmetry proxy to establish the robustness of our results. In our main tests, we examine the impact of the thematic content on information asymmetry based on the Amihud price impact measure. Although this proxy is well accepted in the literature, there exists other proxies to measure information asymmetry. Another very popular proxy in the finance literature used to estimate information asymmetry is the bid-ask spread ($BIDASK$). Elbadry et al. (2015) show that adverse selection increases as the entire spread moves widely and operationalize the use of the bid-ask spread as a proxy for information asymmetry. $BIDASK[-1,+1]$ is defined as

the average bid-ask spread around the ESG report's release date. As for the CAR and AMH measures, we estimate the bid-ask spread over several windows around the release date: (i) $BIDASK[-1,+1]$, (ii) $BIDASK[-2,+2]$ and (iii) $BIDASK[+1,+60]$. We re-estimate Equation 4 with $BIDASK$ as the dependent variable. Results in Table 12 indicate that the thematic content of ESG reports significantly explains the $BIDASK$ around the ESG report's release. In fact, we find that the RSS of our models significantly decreases after the inclusion of topics in Equation 3. To the extent that the $BIDASK$ measures information asymmetry (Elbadry et al., 2015), these results are suggestive of the thematic content being informative to investors.

< Insert Table 12 about here >

7.2. Addressing sample selection bias

Because our sample only includes firms that provide stand-alone ESG reports, the OLS estimation may be subject to the sample selection bias. We therefore conduct the Heckman two-stage procedure (Heckman, 1979) to account for the endogenous nature of firms' decision to publish an ESG report. Specifically, we follow Du and Yu (2020) and estimate the following first-stage Probit model:

$$\begin{aligned} DISCLOSURE_j =& \alpha + \beta \cdot \text{Firm Controls}_j \\ &+ \beta \cdot \text{Industry}_j + \gamma \cdot \text{Year}_j \\ &+ \beta_{10} MKTSHARE_j + \beta_{12} CAPX_j + \beta_{13} FFIN_j + \varepsilon_j, \end{aligned} \tag{8}$$

where $DISCLOSURE$ is a categorical variable taking 1 as a value if the firm releases a ESG report for year t and zero otherwise. In addition to the control variables that are not derived from the ESG reports, we include additional variables to model the selection process. Namely, we include market share ($MKTSHARE$), which is computed as the firm fraction of sales within its industry, capital

expenditure ($CAPX$) normalized by total asset, and financial opacity ($FFIN$), which is computed from an accruals measure, equaling one if the absolute value of a firm's scaled accruals averaged over the past three years is higher than the corresponding industry-year mean (Dhaliwal et al., 2010; Muslu et al., 2017; Shi and Zhang, 2011).[22] $MKTSHARE$, $CAPX$ and $FFIN$ are included in Equation 8, but excluded from the second stage models. These variables impose important exclusion restrictions on the second stage estimation. The results of the first-stage regression are reported in Table A.5 in Appendix. We find that $DISCLOSURE$ is significantly associated with all three variables. In the second stage (Equation 4 and 5), we add the inverse Mills ratio ($LAMBDA$), computed from Model 8 as an additional control variable.[23] Our results are reported in Table 13. We find that, after including $LAMBDA$ to the equation, the conclusions remain qualitatively similar to those presented in Tables 7.

8. Conclusion

We contribute to prior literature by focusing on the informativeness of ESG reports. ESG reports are now a prominent part of corporate disclosure, as ESG criteria become ever more important in investors' capital allocation process considerations (Eccles et al., 2011). However, ESG reports are characterized by limited regulation, which raises doubts about the role and informativeness of such unverifiable voluntary disclosures. Given the heterogeneity of these documents, in the types of information and their composition, it remains an empirical challenge to examine and understand how investors respond to these disclosures. While some evidence in prior research highlights the role of ESG reports in mitigating information asymmetry among investors (Cormier et al., 2009; Dhaliwal

[22] The accruals measure is defined as:

$$ACCR \quad = \frac{\delta CA + \delta CASH - \delta CL - \delta STD - Dep}{TA}, \tag{9}$$

where $ACCR$ is accruals computed using consecutive change in the balance sheet, CA the changes in total current assets, $CASH$ the changes in cash and short-term investments, CL the changes in total current liabilities, STD the changes un debt in current liabilities, DEP the depreciation and amortization expenses from the income statement and Avass the average total assets (Shi and Zhang, 2011).

[23] The sample size of the second stage regression reduces to 1,425 given the data unavailability for exclusion restriction variables: $MKTSHARE$, $CAPX$ and $FFIN$.

et al., 2014; Reverte, 2011), there are studies showing that the disclosure of ESG reports increases information asymmetry among investors (Richardson and Welker, 2001; Berkman et al., 2019). Given this backdrop, we echo the longstanding debate between stakeholder versus shareholder theory and provide timely evidence on the role played by ESG reports on the stock market. Relying on machine learning methods (Bao and Datta, 2014), we depart from prior literature (Du and Yu, 2020; Cormier et al., 2011) and focus on what is contained in the ESG reports at a more granual level and, instead of studying the how the information is disclosed, we analyze the direct impact of the thematic content of ESG reports on investors' reaction and information asymmetry.

We use a hand-collected dataset of ESG reports during the period 2007 and 2020. Relying on this sample, we provide a taxonomy of prominent ESG topics that firms disclose, and as such, evidence the thematic differences across ESG documents. We also show that, while evaluating the value relevance of ESG reports to investors, accounting for informational differences across documents is crucial, as we observe substantial heterogeneity in topics relevance. Crucially, we also find that not all topics share the same value relevance and that there exist substantial differences in the association of each topic with investors' reaction even within the same ESG cluster – Environment, Social and Governance. We also find that reports that disclose a more diverse ESG topics typically lead to a stronger stock price reaction and more significantly reduce information asymmetry. However, cataloging all possible generic ESG topics, regardless of their information value for investors does not further reduce information asymmetry as, beyond a threshold, topics do not contribute to the report's information value any longer. We further observe significant differences through various sub-sample analyses, namely among ESG reports released in various periods, and reports from firms in different industries. Together, this study illustrates that investors are not only sensitive to whether and how information in ESG reports is disclosed, but also to *what* themes are discussed in the report.

We note that, similar to prior literature, our results are descriptive and illustrate associations rather than causal relationships. Nonetheless, as the prevalence of ESG reports grows and the concerning

regulatory landscape evolves, our results are central to investors, academics and regulators alike. Based on the automated and replicable classification of topics using a machine learning-based method, we highlight how regulators can draw a better understanding of what topics are discussed in the ESG reports, and what type of information are value relevant to investors. Moreover, the study directs to several avenues for future research. First, future research can expand the scope of the study by examining the content of ESG disclosures internationally and extend the analyses initiated by Aerts et al. (2008) and Hummel et al. (2022) on non-financial disclosures at an international level. Second, relying on topic modeling, future studies could explore various research questions that concern specific types of information revealed in ESG reports. For instance, sentLDA can be a useful tool to identify the precise sentences that concern a firm's renewable energy operations, allowing researchers to construct text-based measures of firm's renewable-energy-related operations and commitments. Third, given managers' freedom in disclosing ESG information, managers may choose to obfuscate their bad financial performance by disclosing specific themes or topics in their report, while omitting other topics. In line with the obfuscation theory by Schrand and Walther (2000), an interesting avenue for future research would be to investigate whether managers tend to manage the topics disclosed in ESG reports.

References

Aerts, W., D. Cormier, and M. Magnan (2008). Corporate environmental disclosure, financial markets and the media: An international perspective. *Ecological Economics 64*(3), 643–659.

Allen, A. and K. Ramanna (2013). Towards an understanding of the role of standard setters in standard setting. *Journal of Accounting and Economics 55*(1), 66–90.

Amel-Zadeh, A. and G. Serafeim (2018). Why and how investors use esg information: evidence from a global survey. *Financial Analysts Journal 74*(3), 87–103.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets 5*(1), 31–56.

Arendt, S. and M. Brettel (2010). Understanding the influence of corporate social responsibility on corporate identity, image, and firm performance. *Management Decision 48*(10), 1469–1492.

Arslan-Ayaydin, Ö., N. Bishara, J. Thewissen, and W. Torsin (2020). Managerial career concerns and the content of corporate disclosures: An analysis of the tone of earnings press releases. *International Review of Financial Analysis 72*, 101598.

Arslan-Ayaydin, Ö., K. Boudt, and J. Thewissen (2016). Managers set the tone: Equity incentives and the tone of earnings press releases. *Journal of Banking & Finance 72*, S132–S147.

Arslan-Ayaydin, Ö., J. Thewissen, and W. Torsin (2021). Disclosure tone management and labor unions. *Journal of Business Finance & Accounting 48*(1-2), 102–147.

Athanasakou, V. and K. Hussainey (2014). The perceived credibility of forward-looking performance disclosures. *Accounting and Business Research 44*(3), 227–259.

ATT (2010). Meet the possibility economy2010 att sustainability report.

Azarbonyad, H., F. Saan, M. Dehghani, M. Marx, and J. Kamps (2015). Are topically diverse documents also interesting? In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pp. 215–221. Springer.

Baier, P., M. Berninger, and F. Kiesel (2020). Environmental, social and governance reporting in annual reports: A textual analysis. *Financial Markets, Institutions & Instruments 29*(3), 93–118.

Bao, Y. and A. Datta (2014). Simultaneously discovering and quantifying risk types from textual risk disclosures. *Management Science 60*(6), 1371–1391.

Barth, M. E., W. H. Beaver, and W. R. Landsman. Relative valuation roles of equity book value and net income as a function of financial health. *Journal of Accounting and Economics 25*(1), 1–34.

Barth, M. E., W. H. Beaver, and W. R. Landsman (2001). The relevance of the value relevance literaturefor financial accounting standard setting:another view. *Journal of Accounting and Economics 31*, 77–104.

Barth, M. E. and M. F. McNichols (1994). Estimation and market valuation of environmental liabilities relating to superfund sites. *Journal of Accounting Research 32*, 177.

Bellstam, G., S. Bhagat, and J. A. Cookson (2020). A text-based analysis of corporate innovation. *Management Science 67*(7), 3985–4642.

Berkman, H., J. Jona, and N. S. Soderstrom (2019). Firm-specific climate risk and market valuation. *SSRN Electronic Journal*.

Bhagat, S. and B. J. Bolton (2009). Corporate governance and firm performance. *Journal of Corporate Finance 14*, 257–273.

Blazovich, J. L. and L. M. Smith (2011). Ethical corporate citizenship: Does it pay? pp. 127–163.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent dirichlet allocation. *Journal of Machine Learning Research 3*, 993–1022.

Botosan, C. (1997). Disclosure level and the cost of capital. *The Accounting Review 72*(3).

Boudt, K. and J. Thewissen (2019). Jockeying for position in ceo letters: impression management and sentiment analytics: impression management and sentiment analytics. *Financial Management 48*(1), 77–115.

Bozanic, Z., D. T. Roulstone, and A. V. Buskirk (2018, feb). Management earnings forecasts and other forward-looking statements. *Journal of Accounting and Economics 65*(1), 1–20.

Brammer, S. and A. Millington (2008, dec). Does it pay to be different? an analysis of the relationship between corporate social and financial performance. *Strategic Management Journal 29*(12), 1325–1343.

Brandon, R. G., P. Krueger, and P. S. Schmidt (2021, sep). ESG rating disagreement and stock returns. *Financial Analysts Journal 77*(4), 104–127.

Breeze, R. (2013). *Corporate discourse*, Volume 1. A&C Black.

Brown, S. and S. A. Hillegeist (2007). How disclosure quality affects the level of information asymmetry. *Review of Accounting Studies 12*, 443–477.

Bushman, R. M. and A. Smith (2003). Transparency, financial accounting information, and corporate governance. *Economic Policy Review 9*, 65–87.

Campbell, K. and A. Mínguez-Vera (2008). Gender diversity in the boardroom and firm financial performance. *Journal of Business Ethics 83*(3), 435–451.

Carmeli, A., G. Gilat, and D. A. Waldman (2007, sep). The role of perceived organizational performance in organizational identification, adjustment and job performance. *Journal of Management Studies 44*(6), 972–992.

Carrasco-Monteagudo, I. and I. Buendía-Martínez (2013). Corporate social responsibility: a crossroad between changing values, innovation and internationalisation. *European Journal of International Management 2 7*(3), 295–314.

Carroll, A. B. and K. M. Shabana (2010, mar). The business case for corporate social responsibility: A review of concepts, research and practice. *International Journal of Management Reviews 12*(1), 85–105.

Chang, J., J. L. Boyd-Graber, S. Gerrish, and C. Wang (2009). Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems 32*, 288–296.

Chen, Y.-C., M. Hung, and Y. Wang (2018). The effect of mandatory csr disclosure on firm profitability and social externalities: Evidence from china. *Journal of Accounting and Economics 65*(1), 169–190.

Cheng, B., I. Ioannou, and G. Serafeim (2013). Corporate social responsibility and access to finance. *Strategic Management Journal 35*(1), 1–23.

Cho, C. H., G. Michelon, D. M. Patten, and R. W. Roberts (2015). CSR disclosure: the more things change...? *Accounting, Auditing & Accountability Journal 28*(1), 14–35.

Choi, J. and M. Richardson (2016). The volatility of a firm's assets and the leverage effect. *Journal of Financial Economics 121*(2), 254–277.

Christensen, D. M., G. Serafeim, and A. Sikochi (2021, apr). Why is corporate virtue in the eye of the beholder? the case of ESG ratings. *The Accounting Review 97*(1), 147–175.

Clarkson, P. M., X. Fang, Y. Li, and G. Richardson (2013). The relevance of environmental disclosures: Are such disclosures incrementally informative? *Journal of Accounting and Public Policy 32*(5), 410–431.

Clarkson, P. M., J. Ponn, G. D. Richardson, F. Rudzicz, A. Tsang, and J. Wang (2020). A textual analysis of US corporate social responsibility reports. *Abacus 56*(1), 3–34.

Cormier, D., W. Aerts, M.-J. Ledoux, and M. Magnan (2009). Attributes of social and human capital disclosure and information asymmetry between managers and investors. *Canadian Journal of Administrative Sciences 26*(1), 71–88.

Cormier, D., M. Ledoux, and M. Magnan (2009). The use of web sites as a disclosure platform for corporate performance. *International Journal of Accounting Information Systems 10*(1), 1–24.

Cormier, D., M. Ledoux, and M. Magnan (2011). The informational contribution of social and environmental disclosures for investors. *Management Decision 49*(8), 1276–1304.

Cormier, D. and M. Magnan (2013). The economic relevance of environmental disclosure and its impact on corporate legitimacy: An empirical investigation. *Business Strategy and the Environment 24*(6), 431–450.

Cui, J., H. Jo, and H. Na (2016). Does corporate social responsibility affect information asymmetry? *Journal of Business Ethics 148*(3), 549–572.

Currim, I. S., J. Lim, and J. W. Kim (2012). You get what you pay for: The effect of top executives' compensation on advertising and r&ampd spending decisions and stock market return. *Journal of Marketing 76*(5), 33–48.

Davis, G. and C. Searcy (2010). A review of canadian corporate sustainable development reports. *Journal of Global Responsibility 1*(2), 316–329.

Dhaliwal, D., O. Z. Li, A. Tsang, and Y. G. Yang (2014). Corporate social responsibility disclosure and the cost of equity capital: The roles of stakeholder orientation and financial transparency. *Journal of Accounting and Public Policy 33*(4), 328–355.

Dhaliwal, D. S., O. Li, A. Tsang, and Y. Yang (2010). Voluntary nonfinancial disclosure and the cost of equity capital: the initiation of corporate social responsibility reporting. *SSRN Electronic Journal*.

Dhaliwal, D. S., S. Radhakrishnan, A. Tsang, and Y. G. Yang (2012). Nonfinancial disclosure and analyst forecast accuracy: International evidence on corporate social responsibility disclosure. *The Accounting Review 87*(3), 723–759.

Diamond, D. W. and R. E. Verrecchia (1991, sep). Disclosure, liquidity, and the cost of capital. *The Journal of Finance 46*(4), 1325–1359.

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

Dorfleitner, G., G. Halbritter, and M. Nguyen (2015). Measuring the level and risk of corporate responsibility–an empirical comparison of different esg rating approaches. *Journal of Asset Management 16*(7), 450–466.

Du, S. and K. Yu (2020). Do corporate social responsibility reports convey value relevant information? evidence from report readability and tone. *Journal of Business Ethics 172*(2), 253–274.

Du, S., K. Yu, C. Bhattacharya, and S. Sen (2017). The business case for sustainability reporting: Evidence from stock market reactions. *Journal of Public Policy & Marketing 36*(2), 313–330.

Dyer, T., M. Lang, and L. Stice-Lawrence (2017). The evolution of 10-k textual disclosure: Evidence from latent dirichlet allocation. *Journal of Accounting and Economics 64*(2-3), 221–245.

Eccles, R., I. Ioannou, and G. Serafeim (2011). The impact of corporate culture of sustainability on corporate behavior and performance. *SSRN Electronic Journal*.

Elbadry, A., D. Gounopoulos, and F. Skinner (2015). Governance quality and information asymmetry. *Financial Markets, Institutions &amp Instruments 24*(2-3), 127–157.

Fieseler, C. (2011). On the corporate social responsibility perceptions of equity analysts. *Business Ethics: A European Review 20*(2), 131–147.

Fombrun, C. J., N. A. Gardberg, and J. M. Sever (2000). The reputation quotient sm: A multi-stakeholder measure of corporate reputation. *Journal of Brand Management 7*(4), 241–255.

Freeman, R. (1984). Strategic management: A stakeholder approach.

Freeman, R. E. (1998, dec). A stakeholder theory of the modern corporation. pp. 125–138.

Friedman, M. (1963). Capitalism and freedom. *Ethics 74*(1), 70–72.

Ghosh, D. and A. Wu (2012). The effect of positive and negative financial and nonfinancial performance measures on analysts' recommendations. *Behavioral Research in Accounting 24*(2), 47–64.

Giannarakis, G. and I. Theotokas (2011, jan). The effect of financial crisis in corporate social responsibility performance. *International Journal of Marketing Studies 3*(1).

Godfrey, P. C., C. B. Merrill, and J. M. Hansen (2009). The relationship between corporate social responsibility and shareholder value: An empirical test of the risk management hypothesis. *Strategic Management Journal 30*(4), 425–445.

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

Governance Accountability Institute, Inc. (2021). Navigating the way to sustainability.

Goyenko, R. Y., C. W. Holden, and C. A. Trzcinka (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics 92*(2), 153–181.

Graham, J. R., C. R. Harvey, and S. Rajgopal (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics 40*(1-3), 3–73.

Grewal, J., C. Hauptmann, and G. Serafeim (2021). Material sustainability information and stock price informativeness. *Journal of Business Ethics 171*(3), 513–544.

Grewal, J., E. J. Riedl, and G. Serafeim (2019). Market reaction to mandatory nonfinancial disclosure. *Management Science 65*(7), 3061–3084.

Grossman, S. J. and O. D. Hart (1980). Takeover bids, the free-rider problem, and the theory of the corporation. *The Bell Journal of Economics 11*(1), 42.

Hasan, I., N. Kobeissi, L. Liu, and H. Wang (2016). Corporate social responsibility and firm financial performance: The mediating role of productivity. *Journal of Business Ethics 149*, 671–688.

Healy, P. M. and K. G. Palepu (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics 31*(1-3), 405–440.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153.

Heinkel, R., A. Kraus, and J. Zechner (2001). The effect of green investment on corporate behavior. *The Journal of Financial and Quantitative Analysis 36*(4), 431.

Henry, E. (2008). Are investors influenced by how earnings press releases are written? *The Journal of Business Communication 45*(4), 363–407.

Henry, E., J. Thewissen, and W. Torsin (2021). International earnings announcements: Tone, forward-looking statements, and informativeness. *European Accounting Review 0*(0), 1–35.

Hoberg, G. and C. Lewis (2017). Do fraudulent firms produce abnormal disclosure? *Journal of Corporate Finance 43*, 58–85.

Hope, O.-K., D. Hu, and H. Lu (2016). The benefits of specific risk-factor disclosures. *Review of Accounting Studies 21*(4), 1005–1045.

Huang, A. H., R. Lehavy, A. Y. Zang, and R. Zheng (2018). Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science 64*(6), 2833–2855.

Huang, X., S. H. Teoh, and Y. Zhang (2013). Tone management. *The Accounting Review 89*(3), 1083–1113.

Hull, C. E. and S. Rothenberg (2008). Firm performance: the interactions of corporate social performance with innovation and industry differentiation. *Strategic Management Journal* (29), 781–789.

Hummel, K., S. Mittelbach-Hörmanseder, C. H. Cho, and D. Matten (2022). Corporate social responsibility disclosure: a topic-based approach. *Accounting and Business Research 0*(0), 1–38.

Ivers, M. (2010). *Random House Guide to Good Writing*. Ballantine Books.

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

Jo, H. and M. A. Harjoto (2011). Corporate governance and firm value: The impact of corporate social responsibility. *Journal of Business Ethics 103*(3), 351–383.

Jo, H. and M. A. Harjoto (2012). The causal effect of corporate governance on corporate social responsibility. *Journal of Business Ethics 106*(1), 53–72.

Jones, D. A. (2007). Voluntary disclosure in r&d-intensive industries. *Contemporary Accounting Research 24*(2), 489–522.

Jones, T. M. (1995, apr). Instrumental stakeholder theory: A synthesis of ethics and economics. *The Academy of Management Review 20*(2), 404.

Kaplan, S. and K. Vakili (2015). The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal 36*(10), 1435–1457.

Kim, Y., M. S. Park, and B. Wier (2012). Is earnings quality associated with corporate social responsibility? *The Accounting Review 87*, 761–796.

Kleine, A. and M. Von Hauff (2009). Sustainability-driven implementation of corporate social responsibility: Application of the integrative sustainability triangle. *Journal of Business Ethics 85*(3), 517–533.

KPMG (2020). The kpmg survey of sustainability reporting 2020.

Lee, C. M. and E. C. So (2017). Uncovering expected returns: Information in analyst coverage proxies. *Journal of Financial Economics 124*(2), 331–348.

Lev, B., C. Petrovits, and S. Radhakrishnan (2009). Is doing good good for you? how corporate charitable contributions enhance revenue growth. *Strategic Management Journal*, 182–200.

Li, F. (2006). Do stock market investors understand the risk sentiment of corporate annual reports? *SSRN Electronic Journal*.

Loughran, T. and B. Mcdonald (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. *Journal of Finance 66*(1), 35–65.

Loughran, T. and B. McDonald (2014). Measuring readability in financial disclosures. *Journal of Finance 69*(4), 1643–1671.

Loughran, T. and B. Mcdonald (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research 54*(4), 1187–1230.

Luo, X. and S. Du (2015). Exploring the relationship between corporate social responsibility and firm innovation. *Marketing Letters 26*, 703–714.

Martin, P. R. and D. V. Moser (2016). Managers' green investment disclosures and investors' reaction. *Journal of Accounting and Economics 61*(1), 239–254.

Merton, R. C. (1987, jul). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance 42*(3), 483–510.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics 12*(2), 380.

Milne, M. J. and R. Gray (2013). W(h)ither ecology? the triple bottom line, the global reporting initiative, and corporate sustainability reporting. *Journal of Business Ethics 118*(1), 13–29.

Mishra, S. and S. B. Modi (2012). Positive and negative corporate social responsibility, financial leverage, and idiosyncratic risk. *Journal of Business Ethics 117*(2), 431–448.

Mohanram, P. and S. Rajgopal (2009). Is PIN priced risk? *Journal of Accounting and Economics 47*(3), 226–243.

Muslu, V., S. Mutlu, S. Radhakrishnan, and A. Tsang (2017). Corporate social responsibility report narratives and analyst forecast accuracy. *Journal of Business Ethics 154*(4), 1119–1142.

Nekhili, M., H. Nagati, T. Chtioui, and C. Rebolledo (2017). Corporate social responsibility disclosure and market value: Family versus nonfamily firms. *Journal of Business Research 77*, 41–52.

Paniagua, J., R. Rivelles, and J. Sapena (2018). Corporate governance and financial performance: The role of ownership and board structure. *Journal of Business Research 89*, 229–234.

Perrini, F. and A. Tencati (2006). Sustainability and stakeholder management: The need for new corporate performance evaluation and reporting systems. *Business Strategy and the Environment 15*(5), 296–308.

Reguera-Alvarado, N., P. De Fuentes, and J. Laffarga (2017). Does board gender diversity influence financial performance? evidence from spain. *Journal of Business Ethics 141*(2), 337–350.

Reverte, C. (2011). The impact of better corporate social responsibility disclosure on the cost of equity capital. *Corporate Social Responsibility and Environmental Management 19*(5), 253–272.

Reverte, C. (2014). Corporate social responsibility disclosure and market valuation: evidence from spanish listed firms. *Review of Managerial Science 10*(2), 411–435.

Richardson, A. J. and M. Welker (2001). Social disclosure, financial disclosure and the cost of equity capital. *Accounting, Organizations and Society 26*(7-8), 597–616.

Roll, R. W., E. S. Schwartz, and A. Subrahmanyam (2009). O/s: The relative trading activity in options and stock. *SSRN Electronic Journal*.

Ross, S. A. (1977). The determination of financial structure: The incentive-signalling approach. *The Bell Journal of Economics 8*(1), 23.

Saha, R., R. Cerchione, R. Singh, and R. Dahiya (2020). Effect of ethical leadership and corporate social responsibility on firm performance: A systematic review. *Corporate Social Responsibility and Environmental Management 27*(2), 409–429.

Schoenfeld, J. (2017). The effect of voluntary disclosure on stock liquidity: New evidence from index funds. *Journal of Accounting and Economics 63*(1), 51–74.

Schrand, C. M. and B. R. Walther (2000). Strategic benchmarks in earnings announcements: The selective disclosure of prior-period earnings components. *The Accounting Review 75*(2), 151–177.

Securities and Exchange Commission (2016). Business and financial disclosure requirements in regulation s-k.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal 27*(3), 379–423.

Shi, L. and H. Zhang (2011). Can the earnings fixation hypothesis explain the accrual anomaly? *Review of Accounting Studies 17*(1), 1–21.

Son, J., A. Negahban, and D. T Chiang (2019). Topic diversity of online consumer reviews and its effect on review helpfulness. *Conference: Twenty-fifth Americas Conference on Information Systems*.

Strong, R. A. (1986). Using gunning's fog index with term papers and outside reading lists. *Journal of Financial Education* (15), 63–67.

Threlfall, R., A. King, J. Shulman, and W. Bartels (2020). The time has come: The KPMG Survey of Sustainability Reporting 2020. Technical Report 11th edition, KPMG.

Tschopp, D. and R. J. Huefner (2015). Comparing the evolution of csr reporting to that of financial reporting. *Journal of Business Ethics 127*(3), 565–577.

Verbeeten, F., R. Gamerschlag, and K. Möller (2016). Are CSR disclosures relevant for investors? empirical evidence from germany. *Management Decision 54*(6), 1359–1382.

Vitolla, F., N. Raimo, M. Rubino, and A. Garzoni (2019). How pressure from stakeholders affects integrated reporting quality. *Corporate Social Responsibility and Environmental Management 26*(6), 1591–1606.

Vitolla, F., A. Salvi, N. Raimo, F. Petruzzella, and M. Rubino (2020). The impact on the cost of equity capital in the effects of integrated reporting quality. *Business Strategy and the Environment 29*(2), 519–529.

Wood, D. J. (1991). Corporate social performance revisited. *The Academy of Management Review 16*(4), 691.

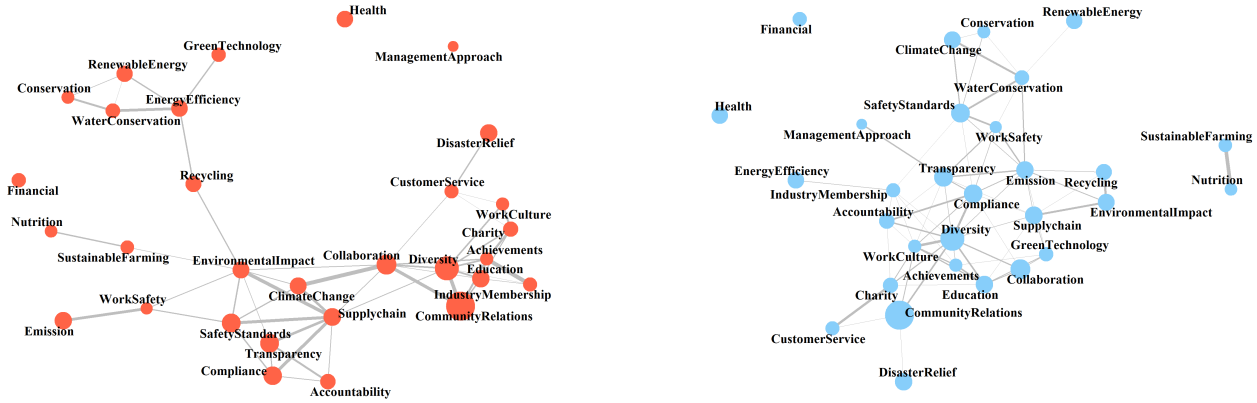
Wu, T., L. He, and F. Zhang (2021). Endogenous discounting, investment and tobin'sq. *The North American Journal of Economics and Finance 55*, 101315.

Zhang, Y., R. Jin, and Z.-H. Zhou (2010). Understanding bag-of-words model: A statistical framework. *International Journal of Machine Learning and Cybernetics 1*(1-4), 43–52.

Zhao, W., J. J. Chen, R. Perkins, Z. Liu, W. Ge, Y. Ding, and W. Zou (2015). A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC Bioinformatics 16*(S13).

Zulu-Chisanga, S. (2019, dec). Does CSR enhance young bank customers' satisfaction and loyalty in a developing economy? the mediating role of trust. *Journal of Accounting and Finance in Emerging Economies 5*(2), 325–342.

Figure 1.: **Network graphs on topic interlinkages**



(a) Word correlation among topics

(b) Co-occurrence of topics across documents

Note: Figure 1(a) presents a network graph illustrating the correlation between the ESG topics with respect to the weights assigned to the constituting words. Figure 1(b) presents a network graph illustrating the correlation between the ESG topics with respect to their co-occurrence across documents. The size of the nodes are proportioned to represent the number of sentences assigned to the given topic in the total sample.

Figure 2.: **Evolution of ESG thematic content between 2007 and 2020**

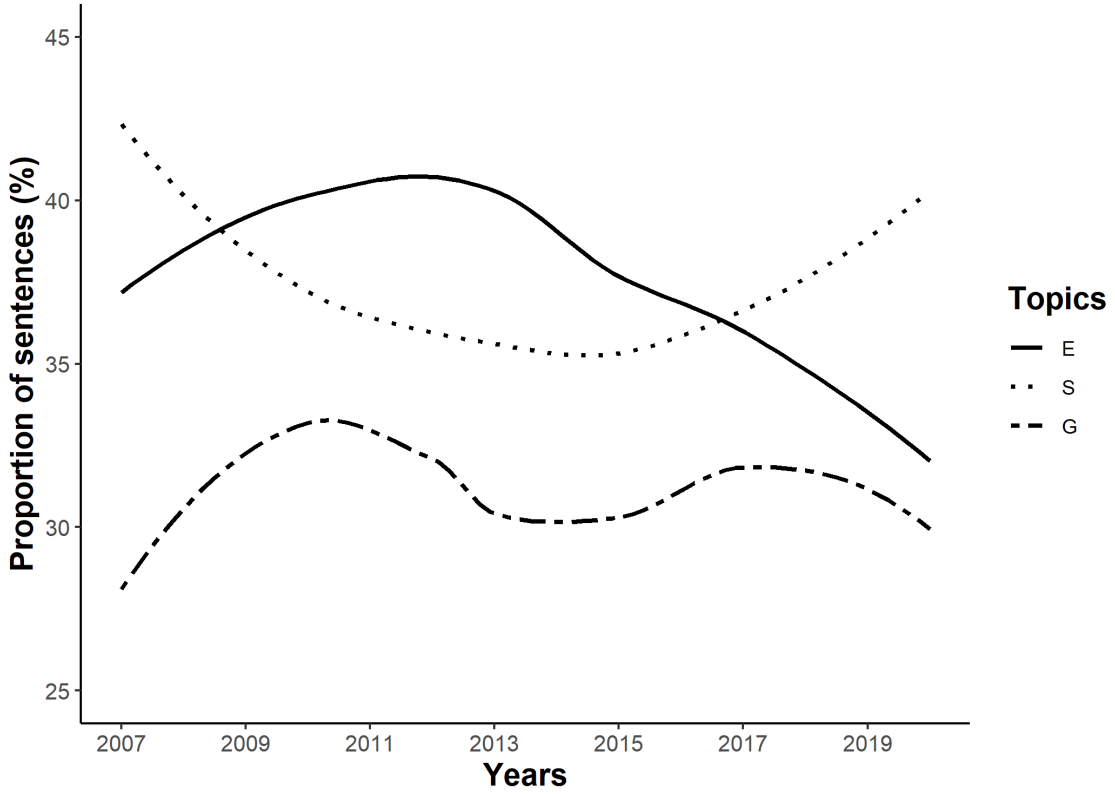


Table 1.: **Sample distribution**

Panel A: Sample distribution by year

Year	Frequency	Percent	Cumulative frequency	Cumulative percent
2007	25	1.500	25	1.500
2008	43	2.579	68	4.079
2009	48	2.879	116	6.959
2010	76	4.559	192	11.518
2011	76	4.559	268	16.077
2012	111	6.659	379	22.735
2013	131	7.858	510	30.594
2014	146	8.758	656	39.352
2015	158	9.478	814	48.830
2016	143	8.578	957	57.409
2017	172	10.318	1129	67.726
2018	213	12.777	1342	80.504
2019	262	15.717	1604	96.221
2020	63	3.779	1667	100.000

Panel B: Sample distribution by industry

Industry	No. of ESG reports	(%)	No. of firms	(%)
Chemicals	72	4.319	17	4.087
Computers	232	13.917	58	13.942
Durable manufacturers	332	19.916	76	18.269
Extractive industries	61	3.659	16	3.846
Financial institutions	186	11.158	65	15.625
Food	109	6.539	16	3.846
Insurance and real estate	51	3.059	24	5.769
Mining and Construction	25	1.500	8	1.923
Others	2	0.120	1	0.240
Pharmaceuticals	94	5.639	16	3.846
Retail	182	10.918	38	9.135
Services	74	4.439	26	6.250
Textiles, printing and publishing	66	3.959	18	4.327
Transportation	101	6.059	15	3.606
Utilities	80	4.799	22	5.288
Total	1,667	100	416	100

Note: Panel A and B of this table presents the sample distribution by year and industry respectively. The industry classifications are based on Barth et al.(1988).

Table 2.: **Summary statistics of ESG report sample**

	Mean	Median	St. Dev.	Q1	Q3
Dependent Variables					
(1) $\text{CAR}_{[-1,+1]}$	−0.0002	−0.00000	0.019	−0.010	0.010
(2) $\text{CAR}_{[-2,+2]}$	0.001	0.001	0.026	−0.012	0.014
(3) $\text{CAR}_{[+1,+60]}$	−0.0002	−0.0004	0.039	−0.021	0.020
(4) $\lvert CAR_{[-1,+1]} \rvert$	1.387***	0.998	1.346	0.352	1.963
(5) $\lvert CAR_{[-2,+2]} \rvert$	1.873***	1.301	1.839	0.514	2.608
(6) $\lvert CAR_{[+1,+60]} \rvert$	2.829***	2.054	2.682	0.809	3.931
(7) $\text{AMH}_{[-1,+1]}$	0.374***	0.130	0.741	0.050	0.332
(8) $\text{AMH}_{[-2,+2]}$	0.222***	0.080	0.412	0.033	0.196
(9) $\text{AMH}_{[+1,+60]}$	0.103***	0.038	0.191	0.015	0.090
Independent Variables					
(10) MOMENTUM	−0.012	−0.004	0.484	−0.230	0.184
(11) SIZE#	72,455.910	17,983.510	237,729.300	6,700.158	48,113.500
(12) ROA%	6.605	5.922	6.281	2.769	9.764
(13) LEV%	27.401	26.083	16.127	14.962	37.925
(14) VOLAT%	3.530	2.028	14.753	1.051	3.761
(15) FIN%	−1.967	−1.730	7.719	−4.987	0.789
(16) LIQUID	2.213	1.793	1.536	1.326	2.646
(17) RD%	4.289	0.474	7.112	0.000	5.396
(18) NoA#	15.711	15	7.182	11	20
(19) WC#	13,999.910	10,000	13,473.300	4,428.5	19,014
(20) FOG	23.753	22.894	5.969	21.341	24.733
(21) TONE%	1.065	1.049	0.782	0.640	1.473
(22) UNCERTAIN%	0.477	0.438	0.272	0.337	0.564
(23) SPECIFIC%	2.920	2.628	1.642	2.225	3.171
(24) FWDLOOK%	5.281	4.532	4.068	2.694	7.042
(25) ESGemph%	6.374	6.372	1.591	5.492	7.171
(26) ESGscore	5.297	5.400	2.228	3.730	6.900
(27) EPSdecline	0.062	0.000	0.242	0.000	0.000
(28) SURPRISE	0.001	0.0004	0.006	−0.00003	0.001
(29) BUSsegment#	1.496	1.386	1.064	0.000	2.441
(30) GEOsegment#	1.746	1.946	1.038	1.099	2.565

Note: This table presents the summary statistics (mean, median, standard deviation, 1st quartile and 3rd quartile) of our dependent variables, as well as for the set of covariates in our model. For the mean values of CARs, we report the two-sided sample t-test results indicating their distinction from 0, where significance levels 10 percent, 5 percent, and 1 percent are denoted with *, ** and ***, respectively.
# denotes that we use the natural logarithmic values in the regression models.

Table 3.: **Correlation matrix**

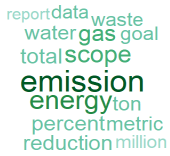
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) $\lvert CAR_{[-1,+1]} \rvert$													
(2) $\lvert CAR_{[-2,+2]} \rvert$	0.628***												
(3) $\lvert CAR_{[+1,+60]} \rvert$	0.440***	0.594***											
(4) $AMH_{[-1,+1]}$	-0.057**	-0.080***	-0.090***										
(5) $AMH_{[-2,+2]}$	-0.062**	-0.078***	-0.093***	0.965***									
(6) $AMH_{[+1,+60]}$	-0.072***	-0.084***	-0.097***	0.942***	0.981***								
(7) MOMENTUM	0.020	-0.012	-0.015	0.008	0.010	0.006							
(8) SIZE#	-0.005	-0.047*	-0.053**	-0.116***	-0.125***	-0.125***	-0.019						
(9) ROA%	-0.018	-0.012	0.005	-0.057**	-0.069***	-0.078***	-0.036	-0.145***					
(10) LEV%	-0.020	0.020	-0.013	-0.064***	-0.067***	-0.076***	-0.019	-0.060**	-0.029				
(11) VOLAT%	-0.022	-0.012	0.001	0.029	0.022	0.019	0.001	-0.042*	-0.032	0.014			
(12) FIN%	-0.045*	-0.022	-0.021	0.027	0.039	0.042*	0.020	0.034	-0.303***	0.125***	-0.003		
(13) LIQUID	0.180***	0.168***	0.164***	-0.044*	-0.055**	-0.061**	0.014	-0.063**	-0.114***	0.071***	0.127***	-0.018	
(14) RD%	0.024	0.004	0.045*	-0.092***	-0.095***	-0.098***	-0.001	-0.080***	0.127***	-0.103***	0.078***	-0.065***	0.011
(15) NoA#	0.074***	0.055**	0.052**	-0.371***	-0.397***	-0.397***	-0.007	0.222***	0.145***	-0.104***	-0.025	-0.113***	0.209***
(16) WC#	0.003	-0.016	-0.068***	-0.159***	-0.170***	-0.169***	0.023	0.096***	-0.004	-0.013	-0.020	-0.018	-0.086***
(17) FOG	-0.039	-0.017	-0.010	0.005	0.021	0.028	0.027	-0.004	-0.031	-0.036	-0.003	0.020	-0.053**
(18) TONE%	-0.032	0.004	0.037	-0.035	-0.037	-0.032	-0.013	0.020	0.046*	-0.037	-0.042*	-0.030	0.014
(19) UNCERTAIN%	0.013	0.019	0.007	-0.015	-0.022	-0.030	0.009	0.092***	-0.123***	0.018	0.016	0.026	-0.019
(20) SPECIFIC%	-0.016	-0.003	0.022	-0.024	-0.022	-0.023	0.022	-0.003	0.020	-0.057**	-0.015	-0.023	-0.026
(21) FWDLOOK%	0.019	0.015	0.023	-0.039	-0.048*	-0.046*	-0.030	0.015	-0.053**	0.055**	-0.002	0.030	-0.007
(22) ESGemph%	0.036	0.042*	0.019	0.044*	0.041*	0.041*	0.013	-0.061**	0.027	0.040	0.054**	-0.079***	-0.026
(23) ESGscore	0.005	-0.033	-0.056**	-0.070***	-0.077***	-0.071***	0.004	-0.108***	0.125***	-0.020	0.006	-0.043*	-0.097***
(24) EPSdecline	-0.030	-0.028	-0.017	0.009	-0.002	-0.007	-0.001	0.100***	0.280***	-0.043*	-0.079***	-0.081***	-0.004
(25) SURPRISE	-0.007	-0.013	-0.015	0.006	0.008	0.006	-0.028	-0.037	-0.019	0.018	0.042*	0.011	0.070***
(26) BUSsegment	0.057**	0.067***	0.085***	-0.099***	-0.100***	-0.104***	0.034	-0.146***	0.029	-0.079***	-0.014	0.013	-0.062**
(27) GEOsegment	0.081***	0.062**	0.097***	-0.066***	-0.067***	-0.071***	-0.010	-0.182***	0.157***	-0.017	0.030	-0.032	-0.005
	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)
(15) NoAsuperscript	0.154***												
(16) WC	0.006	0.148***											
(17) FOG	0.076***	-0.025	-0.035										
(18) TONE%	-0.017	0.072***	-0.177***	0.015									
(19) UNCERTAIN%	-0.076***	-0.002	0.148***	-0.035	-0.216***								
(20) SPECIFIC%	-0.018	-0.057**	-0.305***	0.029	-0.083***	-0.224***							
(21) FWDLOOK%	-0.033	-0.003	0.163***	0.266***	0.021	0.017	-0.114***						
(22) ESGemph%	0.094***	-0.058**	-0.080***	0.149***	-0.069***	-0.023	0.064***	-0.074***					
(23) ESGscore	0.134***	0.058**	0.171***	0.022	-0.033	0.005	0.000	-0.020	0.129***				
(24) EPSdecline	-0.001	0.077***	-0.008	-0.027	-0.007	0.033	0.012	0.015	-0.040	-0.010			
(25) SURPRISE	0.032	0.016	-0.004	-0.010	0.019	-0.042*	0.010	-0.041*	-0.012	0.011	0.003		
(26) BUSsegment	0.191***	0.009	0.057**	-0.001	-0.039	-0.050**	0.063**	0.081***	0.027	-0.012	-0.032	0.069***	
(27) GEOsegment	0.363***	0.024	0.008	0.006	-0.034	-0.057**	0.009	-0.001	0.103***	0.124***	-0.061**	-0.005	0.310***

Note: This table presents the correlation matrix between the dependent variables and the control variables. The table shows Pearson correlation coefficients with significance levels 10 percent, 5 percent, and 1 percent denoted with *, ** and ***, respectively.

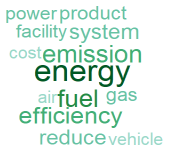
Table 4.: **ESG reports' wordclouds**

(1) ClimateChange

(2) Conservation



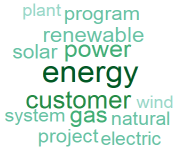
(3) Emission



(4) EnergyEfficiency

(5) EnvironmentalImpact

(6) Recycling



(7) RenewableEnergy

(8) Supplychain

(9) SustainableFarming

(10) GreenTechnology

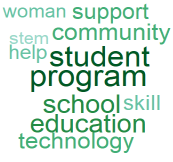
(11) WaterConservation

(12) Charity

(13) CommunityRelations

(14) DisasterRelief

(15) Diversity



(16) Education

(17) Health

(18) Nutrition

(19) SafetyStandards

(20) WorkSafety

(21) Accountability

(22) Achievements

(23) IndustryMembership

(24) Collaboration

(25) Compliance

(26) CustomerService

(27) Financial

(28) ManagementApproach

(29) Transparency

(30) WorkCulture

Table 5.: **Summary statistics of ESG topics**

	Description	Mean	Median	St. Dev.	Q1	Q3
Environmental		148.727	89	175.729	40	184.5
(1) ClimateChange	Initiatives concerning climate change	13.880	5.000	28.307	1.000	14.000
(2) Conservation	Various environmental conservation initiatives	8.479	1.000	24.898	0.000	5.000
(3) Emission	Emissions of greenhouse gases and energy consumption	16.461	10	19.866	3	22
(4) EnergyEfficiency	Reduction and efficiency of energy usage	14.552	7.000	24.056	2.000	16.000
(5) EnvironmentalImpact	Environmental impact and footprint	15.548	6	24.376	1	17
(6) Recycling	Reducing consumption of raw material	13.665	7.000	19.307	1.000	17.000
(7) RenewableEnergy	Alternative energy sources	12.855	1.000	53.615	0.000	5.000
(8) Supplychain	Sourcing of raw materials and distribution	16.886	6.000	29.709	0.000	19.000
(9) SustainableFarming	Emissions and sustainability issues concerning farming	9.043	0.000	29.035	0.000	2.000
(10) GreenTechnology	Efficient and environment-friendly technology	10.773	3.000	19.892	0.000	11.000
(11) WaterConservation	Efficient use and conservation of water	11.174	4.000	17.866	1.000	13.000
Social		147.539	105	141.320	46	203
(12) Charity	Contributing to social issues	11.931	7.000	15.083	2.000	15.000
(13) CommunityRelations	Activities enhancing social relations	29.069	24.000	23.448	12.000	39.000
(14) DisasterRelief	Contributions specific to disaster relief	16.198	9.000	20.621	2.000	22.000
(15) Diversity	Reducing discrimination through fair representation of people	23.815	14	27.593	3	34.5
(16) Education	Promoting eduation-related issues	15.919	6.000	27.573	1.000	18.000
(17) Health	On health, whether physiological or psychological	14.337	0.000	44.619	0.000	4.000
(18) Nutrition	Consumption of local food and food securiity	8.043	0	29.232	0	2
(19) SafetyStandards	Safety standards implemented to protect employees and community	18.563	7.000	29.536	1.000	22.000
(20) WorkSafety	Information on worker safety and injury	7.572	4	10.680	0	10
Governance		125.556	83	145.153	40	167
(21) Accountability	Communicating accountability and details on responsible individuals	12.893	8.000	15.088	2.000	19.000
(22) Achievements	Communicating firm and individual achievements	8.877	6.000	10.795	2.000	12.000
(23) IndustryMembership	Membership in industry groups	10.359	7.000	11.692	2.000	14.000
(24) Collaboration	Increasing collaboration among employees and with stakeholders	19.445	8.000	30.067	2.000	23.000
(25) Compliance	Compliance to guidelines and abiding by ethical code of conduct	18.563	11.000	21.715	2.000	27.000
(26) CustomerService	Relating to customer support and utility	9.901	2.000	19.373	0.000	10.000
(27) Financial	Financial indicators and figures	6.954	2.000	13.652	0.000	7.000
(28) ManagementApproach	Management approach and corporate values	5.637	0	13.422	0	4
(29) Transparency	Disclosure of information	18.277	11	21.173	3	25
(30) WorkCulture	Corporate culture, explicitly described	9.280	4	12.747	0	12
$TotalTopics$		421.822	305	378.442	162	574
$TopicDIVERSITY$		2.570	2.633	0.377	2.388	2.812

Note: This table presents the summary statistics (mean, median, standard deviation, 1st quartile and 3rd quartile) of the topic variables, i.e. the number of document sentences assigned to a given topic. The table also reports the summary statistics of the total sentence count and $TopicDIVERSITY$.

Table 6.: **Base Model results – Control Variables**

	Panel A – Market Response			**Panel B – Information Asymmetry**		
	$\lvert CAR_{[-1,+1]}\rvert$	$\lvert CAR_{[-2,+2]}\rvert$	$\lvert CAR_{[+1,+60]}\rvert$	$\text{AMH}_{[-1,+1]}$	$\text{AMH}_{[-2,+2]}$	$\text{AMH}_{[+1,+60]}$
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	**1.980*****	**2.842*****	**5.927*****	**3.867*****	**2.292*****	**1.664*****
	(0.611)	(0.836)	(1.221)	(0.282)	(0.150)	(0.108)
MOMENTUM	0.056	−0.054	−0.104	−0.018	−0.010	−0.010
	(0.067)	(0.091)	(0.133)	(0.031)	(0.016)	(0.012)
SIZE_{ln}	0.048	0.007	−0.103	**−0.219*****	**−0.134*****	**−0.101*****
	(0.033)	(0.046)	(0.067)	(0.015)	(0.008)	(0.006)
ROA%	−0.003	−0.006	−0.001	−0.004	**−0.003***	**−0.003****
	(0.006)	(0.009)	(0.013)	(0.003)	(0.002)	(0.001)
LEV%	−0.002	0.001	−0.001	**−0.003****	**−0.001****	**−0.001*****
	(0.002)	(0.003)	(0.005)	(0.001)	(0.001)	(0.000)
VOLAT%	−0.003	−0.003	−0.001	−0.001	−0.001	−0.001
	(0.002)	(0.003)	(0.005)	(0.001)	(0.001)	(0.000)
FIN%	−0.004	−0.001	−0.001	**0.004***	**0.003****	**0.002****
	(0.005)	(0.006)	(0.009)	(0.002)	(0.001)	(0.001)
LIQUID	**0.153*****	**0.164*****	**0.203*****	**−0.054*****	**−0.035*****	**−0.028*****
	(0.026)	(0.036)	(0.053)	(0.012)	(0.006)	(0.005)
RD%	0.004	0.006	**0.033****	**−0.009****	**−0.005*****	**−0.004*****
	(0.008)	(0.011)	(0.015)	(0.004)	(0.002)	(0.001)
NoA_{ln}	0.059	0.157	**0.331***	**−0.312*****	**−0.173*****	**−0.123*****
	(0.087)	(0.118)	(0.173)	(0.040)	(0.021)	(0.015)
WC_{ln}	−0.045	−0.080	**−0.225*****	−0.026	−0.014	−0.008
	(0.038)	(0.053)	(0.077)	(0.018)	(0.009)	(0.007)
FOG	**−0.015***	−0.001	−0.011	−0.003	0.000	0.000
	(0.009)	(0.012)	(0.017)	(0.004)	(0.002)	(0.002)
TONE%	**−0.115****	−0.065	0.056	−0.037	**−0.026****	**−0.016***
	(0.051)	(0.070)	(0.102)	(0.023)	(0.013)	(0.009)
UNCERTAIN%	−0.009	0.288	**0.541***	**−0.140***	**−0.096****	**−0.083*****
	(0.155)	(0.213)	(0.311)	(0.072)	(0.038)	(0.028)
SPECIFIC%	−0.025	−0.071	−0.094	**−0.062*****	**−0.037*****	**−0.026*****
	(0.037)	(0.051)	(0.074)	(0.017)	(0.009)	(0.007)
FWDLOOK%	**0.019***	0.012	**0.034***	0.004	0.001	0.001
	(0.010)	(0.013)	(0.019)	(0.004)	(0.002)	(0.002)
ESGemph%	0.029	**0.056***	0.041	−0.005	−0.006	−0.004
	(0.024)	(0.032)	(0.047)	(0.011)	(0.006)	(0.004)
ESGscore	0.001	−0.027	**−0.059***	−0.001	−0.000	0.001
	(0.016)	(0.021)	(0.031)	(0.007)	(0.004)	(0.003)
EPSdecline	−0.014	−0.008	−0.007	**0.009***	0.004	0.003
	(0.011)	(0.015)	(0.023)	(0.005)	(0.003)	(0.002)
SURPRISE	−3.556	−7.848	−12.575	0.882	0.698	0.415
	(5.385)	(7.372)	(10.765)	(2.485)	(1.326)	(0.955)
BUSsegment_{ln}	0.011	0.064	**0.120***	−0.005	0.001	−0.001
	(0.036)	(0.050)	(0.073)	(0.017)	(0.009)	(0.006)
GEOsegment_{ln}	0.016	−0.060	0.028	**0.036***	**0.022****	**0.015***
	(0.043)	(0.058)	(0.085)	(0.020)	(0.010)	(0.008)
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Num. obs.	1667	1667	1667	1667	1667	1667
Adj. R^2	0.068	0.064	0.063	0.346	0.397	0.408

Note: The table reports the results from the estimated coefficients from base model (Eq.3) with only the control variables. As a goodness-of-fit measure, adjusted R^2 are provided. *, ** and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-sided t-tests.

Table 7.: **The thematic content of ESG reports and investors' reaction**

	Market Response			Information Asymmetry		
	$\lvert CAR_{[-1,+1]}\rvert$ Model 1	$\lvert CAR_{[-2,+2]}\rvert$ Model 2	$\lvert CAR_{[+1,+60]}\rvert$ Model 3	$AMH_{[-1,+1]}$ Model 4	$AMH_{[-2,+2]}$ Model 5	$AMH_{[+1,+60]}$ Model 6
Panel A: Model Comparison – *RSS*						
Base Model (Eqn 1)	2730.683	5118.017	88282.268	581.617	165.462	0.943
Models with Topic Variables (Eqn 2)	2640.348	4951.455	86570.573	563.964	160.232	0.913
Difference	**−90.335*****	**−166.562*****	1711.695	**−17.653****	**−5.23*****	**−0.03****
Panel B: 30 Topic Categories						
Environmental						
(1) ClimateChange	0.003	**0.005***	0.014	**−0.001***	**−0.001***	−0.000
	(0.002)	(0.003)	(0.011)	(0.001)	(0.000)	(0.000)
(2) Conservation	0.002	**0.005****	0.001	−0.000	−0.000	−0.000
	(0.002)	(0.002)	(0.010)	(0.001)	(0.000)	(0.000)
(3) Emission	−0.004	−0.003	0.015	0.000	0.000	0.000
	(0.003)	(0.004)	(0.015)	(0.001)	(0.001)	(0.000)
(4) EnergyEfficiency	0.001	0.004	−0.011	−0.000	−0.000	−0.000
	(0.002)	(0.003)	(0.011)	(0.001)	(0.000)	(0.000)
(5) EnvironmentalImpact	0.001	0.000	−0.012	−0.001	−0.000	−0.000
	(0.002)	(0.003)	(0.013)	(0.001)	(0.001)	(0.000)
(6) Recycling	−0.001	−0.001	−0.006	−0.001	**−0.001***	−0.000
	(0.003)	(0.004)	(0.015)	(0.001)	(0.001)	(0.000)
(7) RenewableEnergy	−0.001	−0.001	0.005	−0.000	0.000	0.000
	(0.001)	(0.001)	(0.006)	(0.000)	(0.000)	(0.000)
(8) Supplychain	**0.004****	0.002	**0.023****	−0.001	−0.000	−0.000
	(0.002)	(0.002)	(0.010)	(0.001)	(0.000)	(0.000)
(9) SustainableFarming	−0.003	**0.005***	0.011	−0.001	−0.001	−0.000
	(0.002)	(0.003)	(0.011)	(0.001)	(0.000)	(0.000)
(10) GreenTechnology	**−0.005****	**−0.007****	−0.004	0.002	**0.001****	**0.000****
	(0.003)	(0.004)	(0.015)	(0.001)	(0.001)	(0.000)
(11) WaterConservation	**0.006****	**0.001****	0.012	0.001	0.000	0.000
	(0.003)	(0.004)	(0.018)	(0.001)	(0.001)	(0.000)
Social						
(12) Community	0.000	−0.004	0.008	**−0.003***	−0.001	−0.000
	(0.004)	(0.005)	(0.020)	(0.002)	(0.001)	(0.000)
(13) CommunityRelations	−0.001	−0.004	0.002	−0.001	−0.001	−0.000
	(0.002)	(0.003)	(0.011)	(0.001)	(0.000)	(0.000)
(14) DisasterRelief	−0.003	−0.002	−0.002	−0.001	−0.001	−0.000
	(0.002)	(0.003)	(0.012)	(0.001)	(0.001)	(0.000)
(15) Diversity	−0.000	0.003	−0.008	**−0.002***	−0.001	**−0.000****
	(0.002)	(0.003)	(0.013)	(0.001)	(0.001)	(0.000)
(16) Education	−0.000	−0.002	**−0.023****	**0.002***	0.001	0.000
	(0.002)	(0.003)	(0.011)	(0.001)	(0.000)	(0.000)
(17) Health	0.001	−0.001	−0.005	−0.000	0.000	0.000
	(0.001)	(0.001)	(0.006)	(0.000)	(0.000)	(0.000)
(18) Nutrition	0.003	0.003	**−0.026****	0.000	0.000	0.000
	(0.002)	(0.003)	(0.011)	(0.001)	(0.000)	(0.000)
(19) SafetyStandards	−0.001	0.000	−0.010	0.001	0.001	0.000
	(0.002)	(0.003)	(0.011)	(0.001)	(0.000)	(0.000)
(20) WorkSafety	0.003	0.002	−0.013	0.001	0.001	0.000
	(0.005)	(0.006)	(0.027)	(0.002)	(0.001)	(0.000)
Governance						
(21) Accountability	**−0.007****	**−0.012****	−0.003	**0.004****	**0.002****	**0.000***
	(0.003)	(0.005)	(0.019)	(0.002)	(0.001)	(0.000)
(22) Achievements	0.007	**−0.010***	0.024	−0.001	−0.001	−0.000
	(0.004)	(0.006)	(0.024)	(0.002)	(0.001)	(0.000)
(23) IndustryMembership	−0.003	−0.003	0.016	−0.002	−0.001	−0.000
	(0.004)	(0.005)	(0.023)	(0.002)	(0.001)	(0.000)
(24) Collaboration	0.001	0.001	0.008	**0.001***	**0.001***	**0.000*****
	(0.002)	(0.002)	(0.009)	(0.001)	(0.000)	(0.000)
(25) Compliance	0.000	0.002	0.007	−0.001	−0.001	−0.000
	(0.003)	(0.003)	(0.014)	(0.001)	(0.001)	(0.000)
(26) CustomerService	0.002	0.004	0.010	**0.003****	**0.002****	**0.000****
	(0.002)	(0.003)	(0.014)	(0.001)	(0.001)	(0.000)
(27) Financial	0.001	0.005	0.003	−0.001	−0.001	−0.000
	(0.003)	(0.004)	(0.017)	(0.001)	(0.001)	(0.000)
(28) ManagementApproach	−0.003	−0.006	−0.017	−0.002	−0.001	−0.000
	(0.003)	(0.004)	(0.017)	(0.001)	(0.001)	(0.000)
(29) Transparency	**−0.006****	**−0.006***	−0.002	−0.000	0.000	−0.000
	(0.003)	(0.004)	(0.015)	(0.001)	(0.001)	(0.000)
(30) WorkCulture	−0.001	0.000	0.012	**0.004****	**0.002****	0.000
	(0.004)	(0.006)	(0.024)	(0.002)	(0.001)	(0.000)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed industry and year effects	Yes	Yes	Yes	Yes	Yes	Yes
Num. obs.	1667	1667	1667	1667	1667	1627
Adj. R^2	0.082	0.078	0.174	0.353	0.405	0.475

Note: The table presents the results on the relationship between ESG topic and i) market reaction and ii) information asymmetry. In Panel A, we report the results from the ANOVA tests comparing the sum of squared residuals (RSS) of models with topic category variables (Eq. 4) and without (Eq. 3). Panel B includes the estimated coefficients concerning the various ESG report topics' relationships with cumulative abnormal returns (Market Response) and Amihud illiquidity scores (Information Asymmetry) measured between time horizons: [-1,+1],[-2,+2],[+1,+60]. As a goodness-of-fit measure, adjusted R^2 are provided. *, ** and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-sided t-tests. The ESG topics variables are described in Table 5.

Table 8.: **The Diversity in ESG reports thematic content and investors' reaction**

	Market Response			**Information Asymmetry**		
	$\lvert CAR_{[-1,+1]}\rvert$ Model 1	$\lvert CAR_{[-2,+2]}\rvert$ Model 2	$\lvert CAR_{[+1,+60]}\rvert$ Model 3	$\text{AMH}_{[-1,+1]}$ Model 4	$\text{AMH}_{[-2,+2]}$ Model 5	$\text{AMH}_{[+1,+60]}$ Model 6
Panel A: Model Comparison – *RSS*						
Base Model (Eqn 1)	2732.965	5115	88278.485	582.775	165.626	0.942
Models with Topic Variables (Eqn 2)	2713.964	5104.946	88221.529	577.695	163.81	0.934
Difference	$\mathbf{-19.002}^{***}$	10.054	56.956	$\mathbf{-5.081}^{***}$	$\mathbf{-1.816}^{***}$	$\mathbf{-0.009}^{***}$
Panel B: 30 Topic Categories						
TopicDIVERSITY	$\mathbf{1.060}^{**}$	0.446	−0.424	$\mathbf{-0.745}^{***}$	$\mathbf{-0.427}^{***}$	$\mathbf{-0.030}^{***}$
	(0.455)	(0.624)	(2.593)	(0.210)	(0.112)	(0.009)
TopicDIVERSITY2	$\mathbf{-0.275}^{***}$	−0.141	0.210	$\mathbf{0.145}^{***}$	$\mathbf{0.081}^{***}$	$\mathbf{0.006}^{***}$
	(0.098)	(0.135)	(0.560)	(0.045)	(0.024)	(0.002)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed industry and year effects	Yes	Yes	Yes	Yes	Yes	Yes
Num. obs.	1667	1667	1667	1667	1667	1627
Adj. R^2	0.073	0.066	0.173	0.349	0.402	0.472

Note: The table reports the results concerning the relationship between topical diversity and i) investor reaction and ii) information asymmetry. In Panel A, we report the results from the ANOVA tests comparing the sum of squared residuals (RSS) of models with topic category variables (Eq. 4) and without (Eq. 3). Panel B includes the estimated coefficients concerning the various ESG report topics' relationships with cumulative abnormal returns (Market Response) and Amihud illiquidity scores (Information Asymmetry) measured between time horizons: [-1,+1],[-2,+2],[+1,+60]. As a goodness-of-fit measure, adjusted R^2 are provided. *, ** and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-sided t-tests.

Table 9.: **ESG reports' thematic content and their market impact by period**

	Panel A – Market Response						Panel B – Information Asymmetry					
	$\lvert CAR_{[-1,+1]}\rvert$ RSS		$\lvert CAR_{[-2,+2]}\rvert$ RSS		$\lvert CAR_{[+1,+60]}\rvert$ RSS		$AMH_{[-1,+1]}$ RSS		$AMH_{[-2,+2]}$ RSS		$AMH_{[+1,+60]}$ RSS	
Years 2007-2010 n 192												
Base Model (Eqn 1)	347.584		796.102		9130.251		1.931		0.667		0.004	
Models with Topic Variables (Eqn 2)	282.982		629.903		7696.79		1.43		0.475		0.004	
Difference	64.602		166.199		1433.46		**−0.502****		**−0.192*****		0.004	
Significant Coefficients	(-) *Accountability**	(+) *Education** *Compliance****	(-) *Conservation*** *DisasterRelief***	(+) *Compliance****	(-)	(+)	(-) *ClimateChange*** *Health** *CustomerService***	(+) *EnvironmentalImpact*** *WaterConservation*** *Collaboration**	(-) *ClimateChange*** *RenewableEnergy*** *Supplychain*** *IndustryMembership** *CustomerService**	(+) *EnvironmentalImpact**** *WaterConservation***** *Collaboration*****	(-) *ClimateChange*** *IndustryMembership*** *CustomerService***	(+) *WaterConservation*** *Collaboration*****
Years 2011-2013 n 318												
Base Model (Eqn 1)	365.393		812.489		16662.949		49.7		15.568		0.146	
Models with Topic Variables (Eqn 2)	305.666		633.293		13635.639		43.869		13.644		0.146	
Difference	**−59.727****		**−179.196*****		**−3027.31*****		5.831		1.924		0.146	
Significant Coefficients	(-) *DisasterRelief** *ManagementApproach*****	(+) *Conservation***** *Supplychain***** *IndustryMembership**	(-) *EnvironmentalImpact** *CommunityRelations*** *SafetyStandards*** *ManagementApproach*****	(+) *Conservation***** *Supplychain**	(-) *GreenTechnology** *ManagementApproach***	(+) *Emission***** *IndustryMembership**	(-)	(+) *GreenTechnology** *Charity*** *SafetyStandards***	(-) *Recycling***	(+) *Charity**	(-) *Recycling** *Supplychain**	(+) *Collaboration***
Years 2014-2016 n 447												
Base Model (Eqn 1)	660.691		1326.779		28313.93		112.165		31.088		0.192	
Models with Topic Variables (Eqn 2)	603.372		1211.73		26487.282		97.274		26.782		0.192	
Difference	57.319		115.048		1826.648		**−14.891*****		**−4.306*****		0.192	
Significant Coefficients	(-) *Transparency**	(+) *Supplychain*** *WaterConservation*****	(-) *Transparency**	(+) *Conservation***	(-)	(+) *Supplychain*** *WaterConservation**	(-) *RenewableEnergy*** *SustainableFarming** *Charity***** *IndustryMembership***	(+) *CustomerService*****	(-) *RenewableEnergy*** *SustainableFarming** *Charity***** *IndustryMembership**	(+) *Diversity** *CustomerService*****	(-) *SustainableFarming** *Charity*** *DisasterRelief**	(+) *CustomerService*****
Years 2017-2020 n 710												
Base Model (Eqn 1)	1334.741		2199.552		40389.199		405.866		117.459		0.621	
Models with Topic Variables (Eqn 2)	1238.511		2099.48		39087.863		386.846		111.334		0.621	
Difference	**−96.231****		100.072		1301.336		19.02		6.125		0.621	
Significant Coefficients	(-) *Emission** *RenewableEnergy** *Accountability*** *IndustryMembership***	(+) *ClimateChange*** *WaterConservation*** *Collaboration**	(-) *Accountability***	(+) *ClimateChange** *Supplychain**	(-)	(+)	(-)	(+) *Accountability** *WorkCulture**	(-) *WaterConservation***	(+) *WorkCulture**	(-) *WaterConservation*****	(+) *SafetyStandards** *Accountability** *Collaboration**

Note: The table reports the main model results based on time period sub-samples. The table includes results from the ANOVA tests comparing the sum of squared residuals (RSS) of models with topic category variables (Eq. 4) and without (Eq. 3). The table also lists the significant topic variables in each sub-sample analysis. *, ** and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-sided t-tests.

Table 10a.: **ESG report topics and investors' reaction by industry**

	Panel A – Market Response						**Panel B – Information Asymmetry**					
	$\lvert CAR_{[-1,+1]}\rvert$		$\lvert CAR_{[-2,+2]}\rvert$		$\lvert CAR_{[+1,+60]}\rvert$		$\text{AMH}_{[-1,+1]}$		$\text{AMH}_{[-2,+2]}$		$\text{AMH}_{[+1,+60]}$	
	RSS		*RSS*		*RSS*		*RSS*		*RSS*		*RSS*	
Chemicals n 72												
Base Model (Eqn 1)	69.703		106.28		2207.211		24.402		4.592		0.034	
Models with Topic Variables (Eqn 2)	36.485		50.046		431.306		6.93		1.27		0.034	
Difference	33.217		56.234		**−1775.905****		17.472		**−3.322***		0.034	
Significant Coefficients	(-)	(+)	(-) *EnergyEfficiency**	(+)	(-) *Conservation**** *Recycling**** *RenewableEnergy** *CommunityRelations**	(+) *Financial**	(-) *Diversity*** *Accountability*** *IndustryMembership*** *Transparency**	(+) *DisasterRelief** *WorkSafety*** *Compliance**	(-)	(+) *GreenTechnology** *DisasterRelief**	(-)	(+) *GreenTechnology***
Computers n 232												
Base Model (Eqn 1)	480.209		712.02		17976.939		53.774		15.059		0.061	
Models with Topic Variables (Eqn 2)	399.593		573.942		14377.058		46.94		12.871		0.061	
Difference	80.616		**−138.078***		**−3599.881***		6.835		2.188		0.061	
Significant Coefficients	(-) *IndustryMembership** *Transparency***	(+) *EnvironmentalImpact**	(-) *RenewableEnergy***	(+) *Diversity**	(-) *Diversity** *Education** *Nutrition**** *ManagementApproach***	(+) *SustainableFarming*** *CommunityRelations*** *CustomerService** *Financial** *WorkCulture***	(-)	(+) *SustainableFarming** *Collaboration***	(-)	(+) *SustainableFarming** *Collaboration***	(-) *EnergyEfficiency*** *Charity** *WorkSafety**	(+) *Achievements** *IndustryMembership*** *Collaboration****
Durable manufacturers n 332												
Base Model (Eqn 1)	445.591		1058.456		16809.114		98.679		29.923		0.19	
Models with Topic Variables (Eqn 2)	394.798		948.783		14713.708		80.363		23.695		0.19	
Difference	50.793		109.673		2095.406		**−18.316*****		**−6.228*****		0.19	
Significant Coefficients	(-) *SafetyStandards*** *Financial** *Transparency**	(+) *Emission*** *Compliance***	(-) *DisasterRelief**	(+) *ClimateChange** *CommunityRelations*** *Compliance**	(-) *Achievements** *Financial***	(+) *Conservation*** *IndustryMembership***	(-) *ClimateChange*** *Recycling**** *Diversity** *Nutrition*** *Achievements** *ManagementApproach****	(+) *Emission**** *WaterConservation** *Accountability****	(-) *ClimateChange*** *Recycling**** *Diversity** *Nutrition*** *Achievements*** *ManagementApproach****	(+) *Emission**** *SustainableFarming** *WaterConservation*** *Accountability****	(-) *ClimateChange*** *EnergyEfficiency** *Recycling*** *Nutrition*** *Achievements*** *ManagementApproach****	(+) *Emission**** *SustainableFarming*** *WaterConservation*** *Accountability****
Extractive industries n 61												
Base Model (Eqn 1)	105.967		144.187		3502.953		0.975		0.319		0.004	
Models with Topic Variables (Eqn 2)	32.26		67.475		585.047		0.105		0.033		0.004	
Difference	73.707		76.712		2917.906		**−0.87***		**−0.287***		0.004	
Significant Coefficients	(-)	(+)	(-)	(+)	(-) *Conservation** *Supplychain** *GreenTechnology*** *ManagementApproach**	(+) *RenewableEnergy** *Health** *Transparency*** *WorkCulture**	(-) *Emission*** *CommunityRelations*** *Achievements*** *Financial***	(+) *RenewableEnergy** *Charity** *Accountability** *Compliance*** *ManagementApproach**	(-) *Emission*** *CommunityRelations*** *Achievements*** *Financial**	(+) *Charity** *Compliance*** *CustomerService** *ManagementApproach**	(-) *Emission** *Achievements***	(+) *Compliance**
Financial institutions n 186												
Base Model (Eqn 1)	238.337		311.281		5759.136		90.868		25.081		0.126	
Models with Topic Variables (Eqn 2)	192.064		253.316		4508.594		49.589		13.047		0.126	
Difference	46.273		57.965		1250.542		**−41.279*****		**−12.034*****		0.126	
Significant Coefficients	(-) *Conservation*** *SafetyStandards** *Financial**	(+) *ClimateChange**	(-)	(+) *WorkCulture***	(-) *ClimateChange*** *EnvironmentalImpact**	(+) *RenewableEnergy*** *Achievements**	(-) *EnergyEfficiency**** *WaterConservation*** *Charity** *CommunityRelations*** *Health*** *Nutrition*** *Compliance**	(+) *SafetyStandards*** *WorkSafety**** *CustomerService***	(-) *WaterConservation** *CommunityRelations**** *Health*** *Nutrition****	(+) *SustainableFarming** *SafetyStandards** *WorkSafety**** *CustomerService***	(-) *Nutrition****	(+) *WorkSafety***
Food n 109												
Base Model (Eqn 1)	75.058		159.899		3688.053		5.15		1.134		0.002	
Models with Topic Variables (Eqn 2)	49.974		125.03		2636.881		3.298		0.744		0.002	
Difference	25.084		34.869		1051.173		1.851		0.39		0.002	
Significant Coefficients	(-) *Emission** *Transparency**	(+)	(-)	(+)	(-) *Achievements***	(+)	(-) *CustomerService***	(+) *Recycling** *Health**	(-) *Nutrition** *CustomerService***	(+) *Recycling**	(-) *EnergyEfficiency*** *Charity** *Education** *Nutrition***	(+) *ClimateChange** *Recycling*** *DisasterRelief*** *WorkCulture****
Pharmaceuticals n 94												
Base Model (Eqn 1)	57.962		162.921		4050.247		1.261		0.372		0.001	
Models with Topic Variables (Eqn 2)	32.569		100.498		1859.375		0.374		0.147		0.001	
Difference	25.393		62.423		**−2190.873***		**−0.886*****		**−0.224****		0.001	
Significant Coefficients	(-)	(+) *Education**	(-) *Financial**	(+) *Education***	(-) *EnergyEfficiency** *WaterConservation** *IndustryMembership**	(+) *EnvironmentalImpact**** *Education**** *Accountability***	(-) *EnvironmentalImpact*** *GreenTechnology**** *Charity**	(+) *Conservation**** *Collaboration*** *Financial** *ManagementApproach**	(-) *EnvironmentalImpact*** *GreenTechnology**	(+) *Conservation****	(-) *EnvironmentalImpact*** *Recycling** *Health**	(+) *Conservation***
Retail n 182												
Base Model (Eqn 1)	307.554		838.298		13032.499		61.783		15.151		0.084	
Models with Topic Variables (Eqn 2)	241.753		652.852		10203.511		49.633		11.925		0.084	
Difference	65.801		185.447		2828.988		12.149		3.226		0.084	
Significant Coefficients	(-) *Accountability*** *ManagementApproach***	(+)	(-)	(+)	(-) *WorkCulture***	(+) *RenewableEnergy** *Collaboration**	(-) *Diversity***	(+) *WorkCulture**	(-) *Diversity***	(+) *WorkCulture**	(-) *Conservation** *Emission** *Diversity****	(+) *Compliance*** *WorkCulture**

Table 10b.: **ESG report topics and long-term investor reaction by industry**

	Panel A – Market Response						Panel B – Information Asymmetry					
	$\frac{\lvert CAR_{[-1,+1]}\rvert}{RSS}$		$\frac{\lvert CAR_{[-2,+2]}\rvert}{RSS}$		$\frac{\lvert CAR_{[+1,+60]}\rvert}{RSS}$		$\frac{AMH_{[-1,+1]}}{RSS}$		$\frac{AMH_{[-2,+2]}}{RSS}$		$\frac{AMH_{[+1,+60]}}{RSS}$	
Services n 74												
Base Model (Eqn 1)	87.778		131.412		3532.041		24.727		5.701		0.034	
Models with Topic Variables (Eqn 2)	31.955		29.061		1457.127		11.654		2.459		0.034	
Difference	55.823		**−102.352****		2074.914		13.073		3.241		0.034	
	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)
Significant Coefficients		*Emission** *SustainableFarming**	*Conservation** *EnergyEfficiency** *Diversity*** *WorkSafety*** *Collaboration**** *WorkCulture***	*Emission***** *Education***	*ManagementApproach**	*Emission***		*RenewableEnergy** *CommunityRelations**		*RenewableEnergy**	*Recycling*** *SafetyStandards*** *Achievements***	*ClimateChange*** *Conservation*** *Emission** *RenewableEnergy** *Charity** *Health****
Textiles, printing and publishing n 66												
Base Model (Eqn 1)	98.907		181.098		2913.637		16.769		4.184		0.024	
Models with Topic Variables (Eqn 2)	18.138		33.391		1024.289		4.87		1.283		0.024	
Difference	**−80.769***		**−147.707***		1889.348		11.899		2.901		0.024	
	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)
Significant Coefficients	*Conservation** *CommunityRelations*** *ManagementApproach***		*CommunityRelations*** *ManagementApproach**				*Supplychain***					
Transportation n 101												
Base Model (Eqn 1)	138.062		258.935		4066.75		12.655		2.553		0.013	
Models with Topic Variables (Eqn 2)	84.62		149.041		2454.94		5.668		1.142		0.013	
Difference	53.441		109.894		1611.81		**−6.987****		**−1.411****		0.013	
	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)
Significant Coefficients	*Emission*** *GreenTechnology***	*EnergyEfficiency*** *CustomerService**	*Emission***	*EnergyEfficiency** *Achievements****	*WorkSafety**	*WaterConservation***	*EnergyEfficiency***	*ClimateChange** *WorkCulture***	*EnergyEfficiency*** *Supplychain**	*WorkCulture***	*EnergyEfficiency*** *Supplychain**	*ClimateChange** *Recycling** *CommunityRelations** *IndustryMembership** *WorkCulture***
Utilities n 80												
Base Model (Eqn 1)	69.725		122.604		1980.007		2.119		0.614		0.01	
Models with Topic Variables (Eqn 2)	30.932		56.119		792.75		1.149		0.354		0.01	
Difference	38.793		66.485		1187.256		0.969		0.26		0.01	
	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)
Significant Coefficients	*EnvironmentalImpact** *Supplychain****	*SustainableFarming** *Accountability*** *WorkCulture**	*Supplychain*** *IndustryMembership****	*ClimateChange*** *Recycling** *SustainableFarming*** *WorkCulture**	*EnergyEfficiency*** *Achievements**	*Nutrition***					*Recycling*** *WorkSafety** *ManagementApproach***	*ClimateChange*** *DisasterRelief**

Note: The table reports the main model results based on industry sub-samples. The table includes results from the ANOVA tests comparing the sum of squared residuals (RSS) of models with topic category variables (Eq. 4) and without (Eq. 3). The table also lists the significant topic variables in each sub-sample analysis. *, ** and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-sided t-tests.

Table 11.: **ESG report topics relationship with ESG score and firm performance**

	ESG Rating				Firm Performance		
	$ESGScore$ Model 1	$EScore$ Model 2	$SScore$ Model 3	$GScore$ Model 4	$TobinQ_t$ Model 5	$TobinQ_{t+1}$ Model 6	$TobinQ_{t+2}$ Model 7
Panel A: Model Comparison – *RSS*							
Base Model (Eqn 1)	6997.493	6545.378	3793.284	4512.105	1881.6	821.611	597.211
Models with Topic Variables (Eqn 2)	6100.172	6045.864	3470.196	4389.368	1799.386	767.342	539.384
Difference	**−897.321*****	**−499.514*****	**−323.087*****	**−122.737****	**−82.214*****	**−54.269*****	**−57.827*****
Panel B: 30 Topic Categories							
Environmental							
(1) ClimateChange	0.004	−0.001	−0.000	−0.001	0.001	−0.001	−0.003
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
(2) Conservation	0.001	**−0.010*****	−0.002	**0.008*****	−0.001	−0.001	−0.000
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(3) Emission	0.006	0.006	−0.001	−0.002	0.001	**0.005***	**0.007****
	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
(4) EnergyEfficiency	**0.007****	0.003	**0.009*****	0.003	0.000	−0.001	−0.002
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(5) EnvironmentalImpact	**0.013*****	**0.008****	**0.008*****	0.003	0.003	0.004	0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
(6) Recycling	**−0.017*****	−0.004	**−0.014*****	−0.003	0.002	0.001	0.001
	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
(7) RenewableEnergy	**0.003****	**0.006*****	0.002	**−0.002***	0.000	0.002	0.002
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
(8) Supplychain	−0.001	**−0.012*****	0.001	−0.001	**−0.003****	**−0.003***	**−0.003***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
(9) SustainableFarming	0.002	0.002	−0.003	−0.001	**0.004*****	**0.005****	**0.006*****
	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
(10) GreenTechnology	**0.009****	**−0.008****	**0.009*****	0.004	−0.004	−0.002	−0.004
	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
(11) WaterConservation	−0.004	−0.002	0.002	−0.001	−0.003	−0.004	−0.005
	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Social							
(12) Charity	0.008	−0.000	**0.007***	0.000	0.001	0.002	0.005
	(0.005)	(0.005)	(0.004)	(0.005)	(0.003)	(0.004)	(0.005)
(13) CommunityRelations	−0.001	−0.002	0.001	0.004	−0.001	−0.000	−0.001
	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
(14) DisasterRelief	0.003	0.002	−0.000	−0.003	−0.001	0.003	0.003
	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
(15) Diversity	**0.009*****	0.005	−0.000	0.003	−0.002	−0.003	**−0.008*****
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
(16) Education	**−0.012*****	0.000	−0.001	**−0.006*****	0.000	−0.002	−0.001
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(17) Health	**−0.004*****	0.002	**−0.004*****	**−0.004*****	0.001	0.001	0.000
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
(18) Nutrition	**0.006****	0.000	**0.008*****	−0.001	**−0.004*****	**−0.006*****	**−0.007*****
	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
(19) SafetyStandards	−0.003	−0.004	−0.000	0.001	−0.002	−0.002	−0.001
	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
(20) WorkSafety	**−0.013***	−0.005	0.006	−0.001	**0.007***	0.006	0.003
	(0.007)	(0.007)	(0.005)	(0.006)	(0.004)	(0.004)	(0.005)
Governance							
(21) Accountability	**−0.016*****	−0.007	**−0.007***	−0.004	**0.011*****	**0.010****	0.007
	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)	(0.004)	(0.005)
(22) Achievements	**0.041*****	**0.018*****	**0.017*****	0.000	**−0.011*****	−0.007	**−0.013****
	(0.006)	(0.006)	(0.005)	(0.005)	(0.004)	(0.004)	(0.005)
(23) IndustryMembership	**−0.017*****	**−0.011***	**−0.017*****	−0.005	−0.005	−0.001	0.002
	(0.006)	(0.006)	(0.005)	(0.005)	(0.003)	(0.004)	(0.005)
(24) Collaboration	**0.013*****	**0.007*****	**0.004****	0.001	0.001	0.001	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
(25) Compliance	**0.011*****	**0.007***	0.001	0.005	0.002	−0.000	0.005
	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
(26) CustomerService	**−0.009****	**−0.018*****	**−0.007****	0.000	0.001	−0.004	−0.004
	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
(27) Financial	**−0.011****	−0.006	**0.011*****	−0.004	−0.001	0.002	0.001
	(0.005)	(0.005)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
(28) ManagementApproach	−0.001	0.000	−0.002	0.000	−0.000	−0.000	−0.002
	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)
(29) Transparency	0.001	**0.010****	−0.004	−0.003	−0.001	0.001	−0.001
	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
(30) WorkCulture	**−0.024*****	0.001	−0.007	0.000	0.001	0.000	0.002
	(0.006)	(0.006)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed industry and year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. obs.	1667	1667	1667	1667	1597	871	643
Adj. R^2	0.227	0.238	0.268	0.195	0.593	0.475	0.467

Note: The table presents the results concerning the relationships between ESG topics and i) ESG score and ii) firm performance. In Panel A, we report the results from the ANOVA tests comparing the sum of squared residuals (RSS) of models with topic category variables (Eq. 4) and without (Eq. 3). Panel B includes the estimated coefficients concerning the various ESG report topics' relationships with cumulative abnormal returns (Market Response) and Amihud illiquidity scores (Information Asymmetry) measured between time horizons: [-1,+1],[-2,+2],[+1,+60]. As a goodness-of-fit measure, adjusted R^2 are provided. *, ** and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-sided t-tests. The ESG topics variables are described in Table 5.

Table 12.: **ESG report topics and alternative measure of Information Asymmetry (Bid-Ask)**

	Information Asymmetry		
	$BIDASK_{[-1,+1]}$ Model 1	$BIDASK_{[-2,+2]}$ Model 2	$BIDASK_{[+1,+60]}$ Model 3
Panel A: Model Comparison – *RSS*			
Base Model (Eqn 1)	18560.161	18037.419	11776.744
Models with Topic Variables (Eqn 2)	17902.176	17421.04	11239.245
Difference	**−657.985*****	**−616.379*****	**−537.5*****
Panel B: 30 Topic Categories			
Environmental			
(1) ClimateChange	**−0.011****	**−0.012****	**−0.008***
	(0.005)	(0.005)	(0.004)
(2) Conservation	0.006	0.003	0.005
	(0.005)	(0.005)	(0.004)
(3) Emission	−0.005	−0.006	−0.002
	(0.007)	(0.007)	(0.006)
(4) EnergyEfficiency	**−0.012****	**−0.009***	**−0.007***
	(0.005)	(0.005)	(0.004)
(5) EnvironmentalImpact	−0.004	−0.004	−0.001
	(0.006)	(0.006)	(0.005)
(6) Recycling	−0.000	0.001	−0.005
	(0.007)	(0.007)	(0.006)
(7) RenewableEnergy	0.002	0.002	**0.004***
	(0.003)	(0.003)	(0.002)
(8) Supplychain	0.006	0.005	**0.008****
	(0.004)	(0.004)	(0.004)
(9) SustainableFarming	−0.006	−0.005	−0.005
	(0.005)	(0.005)	(0.004)
(10) GreenTechnology	**0.021*****	**0.019*****	**0.019*****
	(0.007)	(0.007)	(0.005)
(11) WaterConservation	−0.001	0.002	−0.007
	(0.008)	(0.008)	(0.007)
Social			
(12) Charity	0.000	−0.001	−0.002
	(0.009)	(0.009)	(0.007)
(13) CommunityRelations	**−0.009***	−0.008	−0.005
	(0.005)	(0.005)	(0.004)
(14) DisasterRelief	−0.004	−0.004	−0.005
	(0.006)	(0.005)	(0.004)
(15) Diversity	0.008	0.006	0.001
	(0.006)	(0.006)	(0.005)
(16) Education	−0.001	0.000	−0.002
	(0.005)	(0.005)	(0.004)
(17) Health	−0.003	−0.002	−0.001
	(0.003)	(0.003)	(0.002)
(18) Nutrition	0.006	0.005	0.004
	(0.005)	(0.005)	(0.004)
(19) SafetyStandards	**0.009***	**0.010***	0.006
	(0.005)	(0.005)	(0.004)
(20) WorkSafety	**−0.026****	**−0.021***	**−0.017***
	(0.012)	(0.012)	(0.010)
Governance			
(21) Accountability	−0.012	−0.009	−0.004
	(0.009)	(0.009)	(0.007)
(22) Achievements	0.018	0.017	0.014
	(0.011)	(0.011)	(0.009)
(23) IndustryMembership	−0.011	−0.013	−0.008
	(0.010)	(0.010)	(0.008)
(24) Collaboration	0.002	0.002	0.004
	(0.004)	(0.004)	(0.003)
(25) Compliance	−0.002	−0.005	−0.002
	(0.007)	(0.006)	(0.005)
(26) CustomerService	**0.012***	**0.012***	**0.016*****
	(0.007)	(0.006)	(0.005)
(27) Financial	0.004	0.004	0.002
	(0.008)	(0.008)	(0.006)
(28) ManagementApproach	0.013	**0.015***	0.009
	(0.008)	(0.008)	(0.006)
(29) Transparency	0.006	0.011	0.006
	(0.007)	(0.007)	(0.005)
(30) WorkCulture	−0.003	−0.007	0.004
	(0.011)	(0.011)	(0.009)
Controls	Yes	Yes	Yes
Fixed industry and year effects	Yes	Yes	Yes
Num. obs.	1667	1667	1627
Adj. R^2	0.308	0.315	0.394

Note: The table presents the results with alternative information asymmetry variable: bid-ask spread. In Panel A, we report the results from the ANOVA tests comparing the sum of squared residuals (RSS) of models with topic category variables (Eq. 4) and without (Eq. 3). Panel B includes the estimated coefficients concerning the various ESG report topics' relationships with cumulative abnormal returns (Market Response) and Amihud illiquidity scores (Information Asymmetry) measured between time horizons: [-1,+1],[-2,+2],[+1,+60]. As a goodness-of-fit measure, adjusted R^2 are provided. *, ** and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-sided t-tests. The ESG topics variables are described in Table 5.

Table 13.: **The thematic content of ESG reports and investors' reaction (Heckman Selection - Second Stage)**

	Market Response			Information Asymmetry		
	$\lvert CAR_{[-1,+1]}\rvert$ Model 1	$\lvert CAR_{[-2,+2]}\rvert$ Model 2	$\lvert CAR_{[+1,+60]}\rvert$ Model 3	$AMH_{[-1,+1]}$ Model 4	$AMH_{[-2,+2]}$ Model 5	$AMH_{[+1,+60]}$ Model 6
Panel A: Model Comparison – *RSS*						
Base Model (Eqn 1)	2311.757	4508.987	79820.43	412.736	112.951	0.665
Models with Topic Variables (Eqn 2)	2206.387	4328.274	77984.478	399.922	109.59	0.646
Difference	**−105.37*****	**−180.713*****	1835.952	**−12.813***	**−3.361***	0.019
Panel B: 30 Topic Categories						
Environmental						
(1) ClimateChange	**0.004***	**0.006****	**0.022***	−0.001	−0.001	−0.000
	(0.002)	(0.003)	(0.013)	(0.001)	(0.000)	(0.000)
(2) Conservation	0.003	**0.006****	−0.004	−0.001	−0.000	−0.000
	(0.002)	(0.003)	(0.011)	(0.001)	(0.000)	(0.000)
(3) Emission	−0.003	−0.000	0.009	0.000	0.000	0.000
	(0.003)	(0.004)	(0.017)	(0.001)	(0.001)	(0.000)
(4) EnergyEfficiency	0.001	0.003	−0.008	−0.000	−0.000	−0.000
	(0.002)	(0.003)	(0.012)	(0.001)	(0.000)	(0.000)
(5) EnvironmentalImpact	0.002	0.000	−0.012	−0.001	−0.001	−0.000
	(0.002)	(0.003)	(0.014)	(0.001)	(0.001)	(0.000)
(6) Recycling	−0.001	−0.000	−0.011	−0.000	−0.001	−0.000
	(0.003)	(0.004)	(0.016)	(0.001)	(0.001)	(0.000)
(7) RenewableEnergy	−0.001	−0.001	0.000	0.000	0.000	0.000
	(0.001)	(0.002)	(0.006)	(0.000)	(0.000)	(0.000)
(8) Supplychain	**0.004****	0.003	**0.022****	0.000	0.000	−0.000
	(0.002)	(0.002)	(0.010)	(0.001)	(0.000)	(0.000)
(9) SustainableFarming	**0.003***	**0.005***	0.012	−0.001	−0.001	−0.000
	(0.002)	(0.003)	(0.012)	(0.001)	(0.000)	(0.000)
(10) GreenTechnology	−0.004	−0.005	−0.003	**0.003*****	**0.002****	**0.000****
	(0.003)	(0.004)	(0.016)	(0.001)	(0.001)	(0.000)
(11) WaterConservation	0.005	−0.002	0.015	0.002	0.001	0.000
	(0.003)	(0.005)	(0.020)	(0.001)	(0.001)	(0.000)
Social						
(12) Charity	0.003	−0.001	0.032	−0.000	−0.000	0.000
	(0.004)	(0.006)	(0.027)	(0.002)	(0.001)	(0.000)
(13) CommunityRelations	−0.001	−0.003	0.008	−0.001	−0.001	−0.000
	(0.002)	(0.003)	(0.013)	(0.001)	(0.000)	(0.000)
(14) DisasterRelief	**−0.004***	−0.003	−0.009	−0.000	−0.000	−0.000
	(0.002)	(0.003)	(0.013)	(0.001)	(0.001)	(0.000)
(15) Diversity	−0.000	0.004	−0.007	**−0.002****	**−0.001***	**−0.000***
	(0.002)	(0.003)	(0.014)	(0.001)	(0.001)	(0.000)
(16) Education	0.000	−0.003	−0.020	**0.002***	**0.001***	0.000
	(0.002)	(0.003)	(0.013)	(0.001)	(0.000)	(0.000)
(17) Health	0.001	−0.001	−0.004	0.000	0.000	−0.000
	(0.001)	(0.002)	(0.007)	(0.000)	(0.000)	(0.000)
(18) Nutrition	0.003	0.003	**−0.027****	0.001	0.000	0.000
	(0.002)	(0.003)	(0.012)	(0.001)	(0.000)	(0.000)
(19) SafetyStandards	−0.000	0.001	−0.011	0.001	0.000	0.000
	(0.002)	(0.003)	(0.012)	(0.001)	(0.000)	(0.000)
(20) WorkSafety	0.001	−0.001	−0.005	−0.001	−0.001	−0.000
	(0.005)	(0.007)	(0.029)	(0.002)	(0.001)	(0.000)
Governance						
(21) Accountability	**−0.012*****	**−0.013****	−0.009	0.002	0.001	0.000
	(0.004)	(0.005)	(0.023)	(0.002)	(0.001)	(0.000)
(22) Achievements	**−0.011****	**−0.014****	0.021	−0.000	−0.000	0.000
	(0.005)	(0.007)	(0.028)	(0.002)	(0.001)	(0.000)
(23) IndustryMembership	−0.004	−0.004	0.024	0.000	0.000	0.000
	(0.004)	(0.006)	(0.026)	(0.002)	(0.001)	(0.000)
(24) Collaboration	−0.003	−0.003	0.003	−0.000	−0.000	0.000
	(0.002)	(0.003)	(0.012)	(0.001)	(0.000)	(0.000)
(25) Compliance	0.001	0.003	0.006	0.001	0.000	0.000
	(0.003)	(0.004)	(0.017)	(0.001)	(0.001)	(0.000)
(26) CustomerService	0.003	**0.010***	**0.041***	**−0.003***	−0.001	−0.000
	(0.004)	(0.005)	(0.022)	(0.002)	(0.001)	(0.000)
(27) Financial	0.003	0.004	−0.008	−0.000	−0.000	−0.000
	(0.003)	(0.005)	(0.019)	(0.001)	(0.001)	(0.000)
(28) ManagementApproach	−0.004	−0.007	−0.008	−0.002	−0.001	−0.000
	(0.003)	(0.005)	(0.020)	(0.001)	(0.001)	(0.000)
(29) Transparency	**−0.005***	−0.006	−0.002	−0.001	−0.000	−0.000
	(0.003)	(0.004)	(0.016)	(0.001)	(0.001)	(0.000)
(30) WorkCulture	−0.003	−0.005	−0.002	**0.005****	**0.002****	**0.000***
	(0.005)	(0.007)	(0.028)	(0.002)	(0.001)	(0.000)
IMR	**−0.759****	**−0.780***	**3.722****	**0.816*****	**0.488*****	**0.045*****
	(0.295)	(0.413)	(1.753)	(0.126)	(0.066)	(0.005)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed industry and year effects	Yes	Yes	Yes	Yes	Yes	Yes
Num. obs.	1425	1425	1425	1425	1425	1391
Adj. R^2	0.098	0.083	0.169	0.400	0.451	0.519

Note: The table presents the results for the main models derived from implementing a two-staged Heckman selection model. In Panel A, we report the results from the ANOVA tests comparing the sum of squared residuals (RSS) of models with topic category variables (Eq. 4) and without (Eq. 3). Panel B includes the estimated coefficients concerning the various ESG report topics' relationships with cumulative abnormal returns (Market Response) and Amihud illiquidity scores (Information Asymmetry) measured between time horizons: [-1,+1],[-2,+2],[+1,+60]. As a goodness-of-fit measure, adjusted R^2 are provided. *, ** and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-sided t-tests. The ESG topics variables are described in Table 5.