Topics That Matter: ESG Reports and Information Asymmetry

James Thewissen
Shanghai University &
University of Louvain
james.thewissen@uclouvain.be

Prabal Shrestha
University of Louvain
prabal.shrestha@uclouvain.be

Özgür Arslan-Ayaydin
University of Illinois Chicago
orslan@uic.edu

Yan Beibei
Shanghai University
yanbeibei@shu.edu.cn

ABSTRACT

This study examines how the thematic content of Environmental, Social, and Governance (ESG) reports relates to information asymmetry in capital markets. Analyzing 7,565 ESG reports from 1,715 U.S. firms between 1998 and 2023 using sentence-level topic modeling (sentLDA), we identify 30 distinct disclosure topics. Using the dispersion in analysts' forecasts as a proxy for information asymmetry, the results indicate that environmental themes are associated with lower levels of analyst disagreement, social topics with higher disagreement, while governance-related themes show limited association. These findings highlight that the informativeness of ESG reports depends not only on whether firms choose to disclose ESG information, but also on *what* they choose to disclose. In addition, we find that first-time reports are more informative than subsequent ones and that topic diversity exhibits a concave relationship with informativeness, suggesting diminishing returns from a broader thematic coverage. Finally, we find that ESG reports' informativeness is greater when analyst coverage is greater, institutional ownership is lower, ESG performance is higher, and third-party ESG ratings show greater convergence.

Highlights

(1) Using sentence-level topic modeling, we identify 30 distinct and recurring ESG disclosure

topics.

(2) The specific topics disclosed in ESG reports influence their informational value for capital markets, as indicated by analysts' earnings forecast dispersion

(3) Environmental topics are associated with reduced analyst forecast dispersion, social topics increase it, and governance topics have negligible effects.

(4) Topic diversity exhibits an inverted-U relation with informativeness (dispersion falls initially, then rises when coverage becomes overly broad).

(5) First-time ESG reports are more informative than repeat reports, with stronger effects when analyst coverage is higher and institutional ownership is lower; they are also more pronounced for firms with higher ESG performance and greater convergence in ESG ratings.

Keywords: ESG reports; Topic modeling; Information asymmetry.

Acknowledgements: This work was supported by the UIC College of Business Administration FacultyResearchGrant and the Belgian National Fund for Scientific Research (FNRS grant number 40020540, 40026264, and 40034134). We are grateful to Brandon Cline (Editor) and the anonymous reviewer for their thoughtful comments and suggestions, which have significantly improved this manuscript. We also thank participants at the 2023 MFA Conference, LouRIM and LIDAM seminars at UCLouvain, BMI Seminar Series, and the 2023 Seminar Series at Shanghai University for their helpful feedback. Finally, we are thankful to Elizabeth Walsh, Karine Cerrada, Yves de Rongé, Renée Adams, Andriy Bodnaruk, Joan Farre-Mensa and Valentine de Callataÿ for their valuable insights.

1. Introduction

Capital markets have been instrumental in encouraging the prominence of Environmental, Social, and Governance (ESG) reporting by increasingly incorporating ESG factors into investment decision-making (Eccles, Ioannou, & Serafeim, 2012). According to the Government Accountability Institute (Governance Accountability Institute, Inc., 2024), ESG reporting among S&P 500 firms rose from fewer than 20% in 2011 to nearly 99% in 2023. However, in the absence of standardized reporting guidelines, ESG disclosures vary widely in structure, level of detail and thematic focus (Berg, Kölbel, & Rigobon, 2022). This variability complicates analysts' ability to interpret the implications of ESG reports for capital markets, particularly given the presence of alternative information sources, such as ESG ratings, which may already capture aspects of a firm's sustainability profile (Bofinger, Heyden, & Rock, 2022; Christensen, Serafeim, & Sikochi, 2021; Kimbrough, Wang, Wei, & Zhang, 2024). As a result, despite their widespread adoption, important questions remain about what ESG reports actually convey and whether their content is informative for market participants. This study responds to these concerns by unpacking the thematic composition of ESG reports and evaluating how differences in their content affect information asymmetry in capital markets, as proxied by analyst forecast dispersion.

While prior research extensively examines whether ESG reports relate to lower information asymmetry in capital markets, the evidence remains inconclusive. On the one hand, numerous studies in information economics and signaling theory suggest that disclosing ESG information enhances transparency and reduces uncertainty for capital markets. For example, empirical evidence links ESG reporting to improved earnings quality (Y. Kim, Park, & Wier, 2012), decreased ESG ratings disagreement (Kimbrough et al., 2024), lower information asymmetry (Clarkson, Fang, Li, & Richardson, 2013; Cormier, Aerts, Ledoux, & Magnan, 2009), and reduced cost of capital (Dhaliwal, Li, Tsang, & Yang, 2014; Reverte, 2011). On the other hand, literature on the theory of impression management and symbolic legitimacy argues that ESG disclosures may be used strategically, prioritizing form over substance (e.g. green- and bluewashing). From this perspective, ESG reporting may contribute to investor confusion or even amplify uncertainty. Recent studies support this view and show that such disclosures can increase investor disagreement (Christensen et al., 2021; Kimbrough et al., 2024) or exert only marginal influence on financial outcomes (Bofinger et al., 2022; Clarkson et al., 2013; Richardson & Welker, 2001). These conflicting findings challenge the notion that "more disclosure is always better" and point to a more complex dynamic between ESG reporting and information asymmetry.

Collectively, this lack of consensus highlights a critical gap in the existing literature regarding the actual thematic content of ESG reports. Most empirical studies implicitly treat ESG reports as homogeneous documents, overlooking the potentially divergent effects of different disclosure topics. Yet, in the absence of standardized audit or regulatory oversight, firms retain significant discretion over the length, structure, and emphasis of their ESG communications. Consequently, ESG reports vary widely in the topics they address, from carbon emissions and climate risk to employee wellbeing or philanthropic initiatives. Capital markets may attribute greater informational value to some of these topics (e.g., decarbonization strategies) than others (e.g., compliance rhetoric or symbolic social programs), while certain disclosures may even introduce "noise" or be perceived as greenwashing or bluewashing, thereby exacerbating rather than reducing information asymmetry.

Viewing ESG disclosures as uniform may therefore lead to an oversimplification of the fundamentally diverse types of information they contain, potentially explaining the conflicting empirical findings observed in prior research. Together, this highlights that ESG reports' informativeness for capital markets may depend not merely on whether they are disclosed, but critically on their thematic content, specifically, *what* firms choose to disclose. This dimension remains underexplored by prior literature. We address that gap by systematically unpacking the thematic composition of ESG reports and investigating whether the association between ESG disclosures and information asymmetry varies with the topics discussed.

We rely on unsupervised machine learning methods to map the thematic content discussed in ESG reports. The thematic content of an ESG report is defined as a vector of sentence counts relating the number of sentences to each specific topic. 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 & Datta, 2014). The topical composition is then used to identify which topics are associated with lower measured information asymmetry following the release of the report.

To map the informational composition of ESG reports, we hand collect a large sample of 7,565 stand-alone reports for 1,715 firms listed on the two largest American stock exchanges, the New York Stock Exchange (NYSE) and NASDAQ, between January 1998 and April 2023. 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 also reveal a number of interesting topic shifts, pointing to dynamic practices in ESG reporting. Through the years, the environmental ('climate change', 'supply chain', and 'emission') and social ('safety standards', 'community relations', and 'diversity') aspects of

ESG are significantly given more prominence, whereas topics regarding governance appear to be less discussed.

We then examine whether the relation between ESG reporting and information asymmetry differs across report topics. Using the dispersion of financial analysts' forecasts as a reflection of the degree of disagreement among informed market participants about a firm's future performance, we find that variations in thematic content help explain differences in information asymmetry after the ESG report's release. Our analyses show that incorporating the thematic content significantly enhances the model's ability to explain analysts' earnings forecast dispersion compared with models that rely solely on standard firm-level and text-based control variables, ESG performance metrics, and other fixed effects identified in prior literature. Furthermore, we show that the informational value of ESG reports is dynamic and that first-time ESG reporters tend to be more informative relative to repeating ESG reports. In addition, we find that ESG reports' thematic content is significantly more informative for firms with higher ESG performance, convergence in third-party ESG ratings, and analyst coverage. Furthermore, the relationship is stronger for firms with lower institutional ownership, and when the ESG report is not released within five days of the most recent earnings announcement. Overall, these results show that the value relevance of ESG reports significantly depends on *what* is discussed, and that this relationship is shaped by firms' disclosure history and contextual factors.

Our analysis further highlights that ESG themes differ in their association with information asymmetry. Environmental disclosures are associated with lower forecast dispersion, likely because they tend to involve standardized and financially material information (e.g., emissions data, energy use, or supply chain impact) that provide clear signals to analysts. In contrast, social disclosures (e.g., community engagement, employee well-being, or philanthropy) often rely on qualitative narratives and lack verifiable metrics, thereby increasing interpretive ambiguity and analyst disagreement. This result is in line with the concerns raised by Jason

(2020), who emphasizes that the ’S’ in ESG suffers from inconsistent reporting and a lack of standardization, making it challenging for investors to assess social performance effectively. Governance-related themes appear to have no significant effect, possibly because key governance structures are already disclosed through mandatory financial reportings and are less variable across firms (see, e.g., Holder-Webb & Wood, 2008; Xin, Thewissen, Tsang, & Yan, 2025). These findings align with prior research suggesting that the informational value of disclosure is contingent on its clarity, materiality, and standardization (Cookson & Niessner, 2020). Moreover, this result is consistent with arguments from the sociology literature, which theorizes that a plurality of evaluations is likely to occur in newly emerging fields where rules and norms for evaluation are less developed (Lamont, 2012). Collectively, these insights indicate that not all ESG information reduces information asymmetry equally. Its effectiveness depends on what is disclosed and how interpretable that content is to market participants.

Based on the identified topics in ESG reports, we next examine whether the breadth of thematic coverage, i.e., topic diversity, relates to information asymmetry. Drawing on the premise that greater disclosure enhances transparency (Cormier et al., 2009), ESG reports that address a broader range of ESG themes may offer more comprehensive insights into firm activities. However, due to the subjective and often qualitative nature of ESG information, excessive breadth may hinder comprehension and dilute the report’s informativeness (Christensen et al., 2021). Beyond a certain threshold, expanding disclosure to cover an overly broad set of topics, regardless of their decision relevance, may introduce noise rather than clarity. Accordingly, we hypothesize a convex relationship between topic diversity and information asymmetry, where initial increases in diversity enhance transparency, but further expansion ceases to reduce forecast dispersion. Empirically, our findings confirm this pattern: while topic diversity is initially associated with lower analysts’ forecast dispersion, its marginal benefit diminishes and eventually plateaus as more topics are added.

This paper contributes to the literature by demonstrating that ESG reports are a critical vehicle for conveying value-relevant information and by documenting that their thematic content plays an important role in enhancing information transparency. First, to the best of our knowledge, prior research typically focuses on *whether* disclosing on ESG reports influences the firm's cost of capital (Cormier et al., 2009; Dhaliwal, Li, Tsang, & Yang, 2011; Reverte, 2011), firm value (Clarkson et al., 2020) and analysts' forecasts (Aerts, Cormier, & Magnan, 2008; Cormier & Magnan, 2013; Dhaliwal, Radhakrishnan, Tsang, & Yang, 2012). This paper is the first to show that *what* is disclosed in an ESG report is associated with capital markets' information asymmetry. As such, our study emphasizes that, beyond examining *whether* the release of an ESG report impacts capital markets, the specific content discussed by managers in these reports (the *what*) also plays a crucial role in shaping the firm's information environment.

Second, we show that not all ESG topics are equally informative (see, e.g., Aerts et al., 2008; Barth & McNichols, 1994; Cormier, Ledoux, & Magnan, 2011). Our results indicate that the selection of value-relevant topics is what makes an ESG report informative and that simply including a wide array of ESG themes, regardless of their materiality to investors, does not enhance its information value. These findings highlight the importance of a granular, topic-level approach when evaluating ESG disclosures. Rather than treating ESG content as homogeneous or relying solely on pillar-level aggregates (E, S and G), we analyze a broader set of dimensions to assess informativeness. Additionally, we show that topical diversity enhances information transparency, but only up to a point, beyond which additional topics no longer reduce information asymmetry. This approach is consistent with Hoberg and Lewis (2017), who argue that the flexible and discretionary nature of narrative disclosure calls for multidimensional frameworks to evaluate its content.

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, Poon, Pritchard, and Reed (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. Although our methodological approach shares similarities with Jaworska and Nanda (2018) and Goloshchapova et al. (2019), we substantially extend this literature by employing a much larger dataset and leveraging it to test the informational value of ESG reports' thematic content.[1] As such, our study contributes to this stream of research by moving beyond descriptive analysis to empirically test whether the selection of ESG topics is linked to capital markets.[2]

Finally, the objectives of this study align with the current momentum toward regulatory standardization in ESG reporting. In the European Union, effective as of January 2025, the Corporate Sustainability Reporting Directive (CSRD) requires large firms to provide detailed and standardized ESG disclosures. In contrast, the U.S. Securities and Exchange Commission (SEC) has taken a more cautious approach, emphasizing the need to determine "which, if any, sustainability disclosures are important to an understanding of a registrant's business and financial condition" (Grewal, Hauptmann, & Serafeim, 2021). However, in the context of evolving regulatory standards, the lack of systematic identification of the thematic content in ESG reports, and corresponding evidence of its market relevance, poses a major obstacle to

[1] Jaworska and Nanda (2018) base their analysis on 317 reports produced between 2000 and 2013 by 21 major oil companies. Goloshchapova et al. (2019) focus on a UK sample of 641 reports and a European sample of 481 reports from 1999 to 2016.

[2] It is beyond the scope of this paper to examine whether ESG topics are strategically manipulated to shape investors' perceptions of firm performance. Future research could investigate whether managers selectively emphasize or omit certain themes when ESG or financial performance is weak, or when concerns about greenwashing are more likely to arise.

effective rulemaking. Against this background, the paper addresses this gap by empirically identifying which ESG topics meaningfully reduce information asymmetry. Given the scale and pace of regulatory change, such access is essential to encourage policy-relevant research on ESG disclosure practices, including investigations into greenwashing, strategic topic selection, and variation in disclosure practices across industries and jurisdictions.

2. Literature Review & Hypothesis Development

2.1. Are ESG reports informative to capital markets?

In response to the increasing strategic relevance of ESG activities for customer loyalty, product evaluations, employee satisfaction, operational efficiency, and innovation capacity (Brammer & Millington, 2008; Giannarakis & Theotokas, 2011; Hasan, Kobeissi, Liu, & Wang, 2016; Y. Kim et al., 2012; X. Luo & Du, 2015; Zulu-Chisanga, 2019), the majority of firms today issue standalone ESG reports to disclose their environmental, social, and governance practices. These reports differ fundamentally from mandatory financial disclosures. ESG reports remain mostly voluntary, unregulated, and not subject to a standardized reporting framework (Perrini & Tencati, 2006; Tschopp & Huefner, 2015). Unlike 10-K filings, earnings calls, or MD&As, which are designed primarily for investors and subject to regulatory oversight, ESG reports are directed toward multiple stakeholders and contain mostly narrative, non-financial information (Dhaliwal et al., 2011). Managers retain significant discretion over whether to disclose, what to include, and how to structure their disclosures. Audits of ESG reports, if conducted at all, often focus more on internal processes than on the substantive truthfulness of the information (Muslu, Mutlu, Radhakrishnan, & Tsang, 2017). As a result, ESG reports are highly heterogeneous, both in terms of length and content. This flexibility may allow firms to tailor disclosures to their strengths, but it also raises concerns about information quality, comparability, and credibility. These concerns resonate with longstanding debates in the accounting literature over the effectiveness of voluntary disclosure in reducing information asymmetry (Healy & Palepu, 2001).

However, the extent to which ESG reports reduce information asymmetry remains an open question. While ESG disclosures are intended to enhance transparency and demonstrate firms' commitment to sustainability, their credibility is often questioned, particularly when disclosures lack standardized formats or third-party verification. This has raised growing concerns about greenwashing and blue-washing, where companies selectively highlight positive ESG practices while omitting material risks or negative outcomes. Unlike other forms of voluntary disclosure, such as management forecasts or earnings calls, which are subject to regulatory oversight and audit mechanisms, ESG reports typically consist of unaudited, narrative-driven content that falls outside formal assurance processes. This limited oversight gives managers considerable discretion in framing ESG narratives, increasing the risk that such reports are used more to cultivate reputational capital than to provide decision-useful information (Berg et al., 2022; Christensen et al., 2021; Kimbrough et al., 2024). As a result, financial analysts and other market participants may view ESG reports with skepticism, which can reduce their effectiveness in mitigating information asymmetry.

This perception is further reinforced by the growing reliance on alternative sources of ESG data, such as third-party ratings, which are often viewed as more objective or data-driven. However, the value of such alternative sources is not always straightforward. For instance, ESG ratings frequently diverge due to differences in methodology, weighting schemes, and scope (Berg et al., 2022; Christensen et al., 2021), making it difficult for investors to form consistent judgments about a firm's sustainability profile. In this context, ESG reports offer incremental value by helping investors interpret and contextualize these signals. They allow firms to explain their strategic ESG positioning, provide forward-looking insights, and emphasize internally valued themes that might not be captured by external ratings. Furthermore, theoretically, disclosure models show that even biased or strategically framed communication can reduce information asymmetry when investors account for disclosure incentives (Dye, 1985; Verrecchia, 2001). From this perspective, ESG reports may serve as a signaling device, particularly in areas where performance is multidimensional or contested.

Empirical research mirrors these conceptual tensions. Several studies find that ESG disclosure reduces forecast error, lowers the cost of equity capital, and improves analyst accuracy (Aerts et al., 2008; Cormier & Magnan, 2013; Dhaliwal et al., 2011). Others investigate how the form of disclosure influences informativeness. For instance, Du and Yu (2020) show that despite a tendency for managers to obfuscate when performance is weak, the tone and readability of ESG reports significantly shape investor reactions at the time of disclosure. However, other studies report null or adverse effects. For example, Christensen et al. (2021) and Kimbrough et al. (2024) show that ESG disclosure actually increases ESG ratings disagreement, while Chen, Hung, and Wang (2018) show that mandated ESG reporting in China reduces firm profitability. Furthermore, Grewal, Riedl, and Serafeim (2019) show that EU's non-financial disclosure mandates triggered negative market reactions, albeit mainly for firms with weak ESG performance. Similarly, Bofinger et al. (2022) find that ESG performance may not always create value and could even lead to misvaluation, while Berkman, Jona, and Soderstrom (2019) and Richardson and Welker (2001) find that ESG disclosures increase cost of capital, particularly when linked to long-horizon or complex environmental risks.

2.2. *Opening the black box of ESG reports*

2.2.1. ESG reports' thematic content and its information value

The mixed findings reported in prior studies suggest that existing theoretical frameworks may not fully capture the mechanisms underlying ESG disclosure outcomes. This study adopts a complementary perspective by examining alternative dimensions of ESG report informativeness. While prior research mostly focuses on *whether* (Dhaliwal et al., 2011) or *how* ESG information is disclosed (Du & Yu, 2020), the context or types of information conveyed in the document remains largely overlooked. Yet, because of the lack of a formal control mechanism or standardization, ESG reports are significantly different from each other, which is reflected in considerable differences in their length. For instance, the AT&T's 2010 sustainability report has 82 pages, while the 2011 report of the NYSE Euronext company contains 40 pages. In addition, as 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 such a context, ignoring the content of ESG reports constitutes a major shortcoming in our understanding of the information value of ESG reports.

We propose that the degree of information asymmetry relates to ESG reports' thematic content, namely their topic composition. For example, 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 according to its thematic category. 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 information asymmetry. These results are important as they suggest that the informativeness of ESG disclosures significantly depends on its content. In turn, we need to relax prior literature's assumption that the ESG report is an homogeneous collection of topics (see, e.g., Clarkson et al., 2020; Dhaliwal et al., 2011, 2012; Du & Yu, 2020). Instead, we argue that ESG reports are clusters of a variety of topics that, individually, have a different information value.

Hypothesis 1: Ceteris paribus, a report's thematic content is associated with lower information asymmetry after the release of the ESG report.

In this paper, we automatize the identification of topics discussed in a report, which complements previously popular approaches of identifying value relevant information through surveys (Amel-Zadeh & Serafeim, 2018), grids (Cormier & Magnan, 2013), experiments (Martin & Moser, 2016) or keywords (Zhang, Jin, & Zhou, 2010). If we find evidence consistent with our prediction that the thematic content is associated with reduced information asymmetry after the ESG report release, then the next objective is to identify which topics, specifically, explain this relationship.

2.2.2. What is of value-relevance in ESG reports?

It is not *ex ante* obvious which specific ESG topics are perceived by capital markets as informative. Although prior literature suggests that greater disclosure tends to reduce information asymmetry (Hope, 2003; Lang & Lundholm, 1996), this relationship may not apply uniformly across all categories of ESG content. ESG disclosures often lack standardized frameworks and are characterized by considerable subjectivity, thereby increasing the scope for divergent interpretation. As Cookson and Niessner (2020) argue, a disclosure that introduces complexity or ambiguity, rather than clarity, can paradoxically exacerbate information asymmetry. For example, metrics related to carbon emissions or board composition are generally considered financially material and are supported by quantifiable, comparable indicators. In contrast, disclosures on charitable contributions, regulatory compliance, or employee well-being frequently rely on narrative formats and unverifiable claims, rendering them more difficult to integrate into financial models and increasing the likelihood of interpretive disagreement among analysts.

This challenge is particularly pronounced in the social dimension of ESG. As Jason (2020) observes, the large presence of environmental data in ESG investing does not necessarily reflect its intrinsic importance, but rather its greater measurability. The analysis of Jason (2020) includes the view of an investment manager who emphasizes this point: “Planet isn’t necessarily more important than people, it’s just easier to measure. Investors like measuring things that they can put into their models, and carbon is easy to quantify.” Furthermore, empirical evidence further supports this measurement asymmetry: a 2021 survey conducted by BNP Paribas finds that 51% of institutional investors considered the social pillar the most difficult to assess and incorporate into investment strategies due to a lack of standardized, comparable data (BNP Paribas Securities Services, 2021).

Complementing this observation, insights from the sociology of valuation suggest that information asymmetry is more likely to emerge in domains where evaluative standards are underdeveloped or contested (Lamont, 2012). Disclosures addressing novel, loosely defined, or non-standardized ESG themes, particularly within the “S” dimension, are therefore more likely to prompt heterogenous assessments, resulting in greater dispersion in analysts’ forecasts. These considerations put forth the importance of treating ESG disclosures as differentiated rather than uniform, and of examining how the specific nature and structure of disclosed topics condition their informational relevance to capital markets.

Hypothesis 2: The association between ESG reports’ thematic content and information asymmetry varies systematically across environmental, social, and governance topics.

In this manner, this study adopts a two-step approach. First, we examine whether, on average, the thematic content of ESG reports helps reduce information asymmetry after the report’s release. Second, recognizing that not all topics may be equally informative, we investigate whether the association between thematic content and information asymmetry varies across environmental, social, and governance disclosures. Given the heterogeneity of ESG disclosures and the absence of standardized reporting structures, we do not impose *ex ante* assumptions regarding the specific topics discussed in ESG reports. Instead, we employ a data-driven approach using sentence-level Latent Dirichlet Allocation (sentLDA), an unsupervised machine learning method, to identify the latent thematic structure of ESG disclosures. Following topic extraction, we classify the resulting themes into environmental, social, and governance categories based on manual review of dominant words, bi-grams, and representative sentences and examine their relationship with information asymmetry.

2.3. *Topic diversity and the Informational value of first-time ESG reports*

While ESG reports' thematic content is likely to be a key determinant of their informativeness, specific features of the disclosure itself may further shape its impact on information asymmetry. In particular, the diversity of topics discussed within a report may enhance informativeness by providing a broader representation of firm's activities, although excessive breadth could dilute the relevance of the information conveyed. In addition, the novelty of the disclosure may condition its informational impact: first-time ESG reports, by introducing previously unavailable information, are likely to have greater implications for information asymmetry compared to subsequent reports. Building on these considerations, we next examine the roles of topic diversity and reporting novelty in explaining the informativeness of ESG disclosures.

2.3.1. *Topic diversity in ESG reports*

Based on this granular analysis of ESG reports' thematic content, we next focus on the relationship between information asymmetry and the overall information coverage of ESG reports, as measured by their topic diversity. Apart from the extent of the information on each topic, topic diversity in the document represents the overall information richness – the variety of information that is covered in the document. A high topic diversity in the ESG disclosures represents an equitable distribution of the various prominent and common ESG topics, quantifying the disclosure's topic coverage. In contrast, a low diversity score indicates a narrow focus and a lack of informational richness.

Prior research provides evidence of the association between increased topic diversity and greater information value. For instance, Y. Kim and Son (2022) show that discussing more topics in online consumer reviews provides more information about a product's features and functions. Similarly, Azarbonyad et al. (2017) find a positive relationship between the topical diversity and the value of a document. These results indicate that ESG reports incorporating greater ESG themes are more informative to investors. A report discussing many themes is more likely to contribute to the investor's assessment of the firm's future performance, goals and risk/opportunity management strategies. Consequently, we expect that ESG reports with greater topic diversity is associated with a stronger stock price reaction and a greater reduction in the level of information asymmetry at the release of the report.

However, while a disclosure that covers a breadth of topics signals the firm's diverse efforts on ESG issues, simply covering a laundry list of topics is likely to lead to a lack of specificity and substance. Therefore, such communication of ESG information inadvertently risks being perceived as uninformative. Extant prior studies on corporate communication, such as Miihkinen (2012) and Hope, Hu, and Lu (2016), document the importance of depth and specificity in firm disclosures, and hence, argue the cost of the lack thereof. For instance, Christensen et al. (2021) suggest that, although greater disclosure is typically expected to reduce information asymmetry, the subjective nature of ESG information is likely to be associated with the cost of a higher disagreement, as disclosure expands opportunities for different interpretations of information. Based on this evidence, we therefore expect that the extensive information coverage of ESG reports with greater topic diversity may also lead to information overload and cause hindrance to comprehension (Chapman, Reiter, White, & Williams, 2019; Christensen et al., 2021; Edmunds & Morris, 2000). As such, while the informativeness of ESG reports initially increases with topic diversity, after a threshold, greater diversity may not contribute to the information value of the document but instead lead to a curvilinear relationship.

Hypothesis 3: Ceteris paribus, the diversity of topics discussed in a ESG report has a convex relationship with information asymmetry.

2.3.2. First-time versus repeated reporters

Although the thematic content of ESG reports is generally expected to be informative to investors, its value is unlikely to be static. As such, we argue that the informativeness of ESG disclosures is a dynamic construct that depends on whether a firm is issuing such a report for the first time or has done so in prior years. Prior research suggests that ESG reports tend to exhibit limited year-over-year variation, often retaining similar language and structure across reporting cycles (Dhaliwal et al., 2012; Guidry & Patten, 2010). Consequently, the initial release of a stand-alone ESG report may represent a meaningful change in a firm's disclosure environment (Wang & Li, 2016). Moreover, the decision to publish a first-time ESG report may signal increased commitment to sustainability practices and suggest a firm's capacity to allocate discretionary resources to such initiatives (Dhaliwal et al., 2012; McGuire, Sundgren, & Schneeweis, 1988). This implies that the first ESG report may communicate information that is more novel, credible, or value-relevant than that contained in subsequent reports. Consistent with this view, Guidry and Patten (2010) find that first-time high-quality reports are associated with significant improvements in reputational ratings, while Dhaliwal et al. (2012) show that initial ESG disclosures can reduce firms' cost of capital. Motivated by these findings, we extend Hypothesis 1 by examining whether the thematic content of a firm's first ESG report has a stronger effect on reducing information asymmetry than successive ESG disclosures.

Hypothesis 4: Ceteris paribus, the thematic content of first-time ESG reports provide significantly more information value compared to successive ESG reports.

3. Sample selection, text preprocessing, sentLDA algorithm, 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 listed on the NYSE and the NASDAQ between 1998 and 2023. We focus on stand-alone ESG reports not only because of their growing prevalence across the broader economy, but also because, as highlighted by Cho, Guidry, Hageman, and Patten (2012) and Unerman and O'Dwyer (2007), such reports enhance organizational accountability and serve as a powerful means of assessing a firm's social and environmental impact. In addition, Wang and Li (2016) posit that stand-alone ESG reports tend to provide more detailed ESG information than those that are integrated into the annual reports or filings to meet regulatory requirements.

To collect our sample of ESG reports, we proceed in two steps. At each step in the sample selection, any observation with missing data for the required control variables is discarded.

(1) In the first step, we hand-collect all the ESG reports from the firms listed on the NYSE and NASDAQ. We obtain 8,185 ESG reports between 1998 and 2023. If a report is found on the company's website, we rename it using its company's ticker, with the fiscal year of the release separated by a hyphen (e.g. "INTC-2015"). If the report is not found on the company website, we use Google and search for the report based on a list of keywords. We first mention the company name, the country of the company's headquarters and the fiscal year, "+ ESG CSR report", enclosed by inverted commas. For instance: "Intel-USA-2015+ESG CSR REPORTS". In addition to the company's website, we rely on external sources to identify the ESG report, such as CSRwire.com, CorporateRegister.com, GlobalReporting.org, SocialFunds.com, or BusinessWire.com. Such websites allow us to obtain ESG reports from firms that are delisted, bankrupt or merged, which therefore minimizes the survivorship bias in our sample.

(2) To obtain accounting-based data, we merge the hand-collected sample with US

COMPUSTAT annual industrial, which results in 7,983 observations. Next, we merge our sample with the stock market CRSP data and the Thomson Financial Spectrum Database to obtain institutional investment data. Our sample is then reduced to 7,783 observations. After merging with the IBES database to obtain analysts' forecast variables, our final sample is composed of 7,565 firm-year observations between 1998 and 2023 for 1,715 firms, of which 610 firms are listed on the NASDAQ and 1,105 on the NYSE.

To measure analysts' forecast dispersion after the ESG report release, we extract the date of release of the company's report using the company's own website, or we look it up on Google using the same algorithmic methodology previously mentioned, but complementing with "news" at the end of the query. Alternatively, if we do not find the information online, we use the PDF metadata, which includes the report's release date, the last modification or the report's author.

To incorporate firm-level ESG characteristics, we further merge data from MSCI, LSEG (formerly Refinitiv), and Bloomberg, which are used to construct (i) ESG ratings, (ii) ESG ratings disagreement, and (iii) CO_2 emissions scores. Due to incomplete and non-overlapping coverage across providers, merging with MSCI, LSEG and Bloomberg leads to a significant decrease in the sample size. Specifically, merging our data with MSCI emission score ($EmissionsScore_{MSCI}$) yields 4,807 observations, merging with $ESGscore_{MSCI}$ yields 6,170 observations, LSEG 5,177, and Bloomberg 3,812. Constructing the ESG ratings disagreement measure, which requires observations from all three sources, further reduces the sample to 3,068 firm-years. To mitigate this reduction in sample size, we adopt a progressive estimation strategy: the topic modeling algorithm and baseline models are estimated on the full sample, and ESG ratings, disagreement, and emissions measures are progressively included in the analyses using the corresponding sub-samples.

3.1.2. Pre-processing of the ESG reports

The raw files in our sample are typically in a PDF format.[3] 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 (Thewissen, Thewissen, Torsin, & Arslan-Ayaydin, 2022).[4] Based on these documents, various pre-processing steps on the text are implemented (Arslan-Ayaydin, Bishara, Thewissen, & Torsin, 2020; Boudt & Thewissen, 2019; Henry, Thewissen, & Torsin, 2021). Namely, we remove non-English characters (punctuation and numbers), non-alphanumeric symbols, numbers, repeated space character, one-letter words, and letters attached with special characters (taking care not to delete words such as o2 and co2); and then we lemmatize 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".[5] Finally, we follow Huang, Lehavy, Zang, and Zheng (2018) 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, Ng, & Jordan, 2003), an unsupervised machine-learning-based tool that researchers widely use in the information retrieval literature (see e.g., Bellstam, Bhagat, & Cookson, 2020; Brown & Hillegeist, 2007; Hoberg & Lewis, 2017; Kaplan & Vakili, 2015). Specifically, we apply the Sentence Latent Dirichlet Allocation (sentLDA) proposed by Bao

---

3 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.

4 This step removes all pictures and tables.

5 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, 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 of a word w in document d.

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 for the identification of the topic discussed in each 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 the 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 & 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:

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

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

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. Regression model

To test Hypothesis 1 and 2 concerning the relationship between ESG reports' thematic content and information asymmetry, we use the topics obtained from the sentLDA model as input variables in a multi-factor model. First, the benchmark model is the traditional approach consisting of a linear model in which the analysts' dispersion following the report's release is estimated using a set of covariates used in prior literature. This set of covariates capture firm-specific controls, ESG reports' stylistic features, ESG ratings-related variables, as well as year- and industry fixed-effects:

$$ADisperse_j = \alpha + \beta.Firm\ Controls_j + \beta.ESG\ Report\ Controls_j + \beta.ESG\ Rating\ Controls_j + \beta.Industry_j + \gamma.Year_j + \varepsilon_j, \tag{3}$$

where ε_j are firm-clustered standard errors, corrected for heteroskedasticity. The dependent variable is $ADisperse_j$, following the definition of Diether, Malloy, and Scherbina (2002). For each ESG report released by firm j for fiscal year t, we compute $ADisperse_j$ as the standard deviation of analysts' earnings forecasts for fiscal year t+1, issued after the ESG report release date. This value is then scaled by the absolute value of the mean forecast to normalize across firms. This measure captures post-disclosure disagreement among analysts regarding the firm's future earnings. To reduce the influence of extreme values, we winsorize the dependent variable at the 1^{st} and 99^{th} percentiles.

Next, we include the topic variables based on the sentLDA ($ESGTopicVariable_{k,j}$) to Equation 3 as a measure of thematic content in ESG reports, and define the following model:

$$ADisperse_j = \alpha + \sum_{k=1}^{K} \delta_k.ESGTopicVariable_{k,j} + \beta.Firm\ Controls_j + \beta.ESG\ Report\ Controls_j + \beta.ESG\ Rating\ Controls_j + \beta.Industry_j + \gamma.Year_j + \varepsilon_j, \tag{4}$$

where ε_j are firm-clustered errors, corrected for heteroskedasticity. $ESGTopicVariable_{k,j}$ 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).

Furthermore, Hypothesis 3 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:

$$ADisperse_j = \alpha + \beta.TopicDIVERSITY_j + \beta.TopicDIVERSITY^2 + Firm\ Controls_j + \beta.ESG\ Report\ Controls_j + \beta.ESG\ Rating\ Controls_j + \beta.Industry_j + \gamma.Year_j + \varepsilon_j, \quad (5)$$

where topic diversity (*TopicDIVERSITY*) represents the number of unique topics in the report. We expect a negative coefficient for *TopicDIVERSITY* and a positive coefficient for $TopicDIVERSITY^2$ for analysts' dispersion.[6]

Finally, for Hypothesis 4, we examine the difference in the information value of the thematic content of repeated versus first-time reporters. To do this, we create two sub-samples based on whether the ESG report is a first-time or a repeated reporter. For each sub-sample, we then estimate Equation 3 and 4 and compare the goodness of fit based on the ANOVA test.

3.3.1. Control variables

We define a large set of control variables from prior studies, which examine factors that affect the analysts' forecast dispersion. We first control for firm-specific variables. We include *SIZE*, which is the natural logarithm of total assets at the end of the current year. *ROA* represents the income before extraordinary items divided by total assets at the end of the previous year. *LEV* represents financial leverage and is calculated as total debt (short-term plus long-term debt) divided by total assets at the year-end (Y. Luo, Ni, & Thewissen, 2024; Tong et al., 2025). We also control for earnings uncertainty with *VOLAT*, representing the standard deviation on *ROA* for a period of five years. Furthermore, the firm-specific model controls include financing activities (*FIN*), stock liquidity (*LIQUID*), research and development intensity (*R&D*), and analyst following (*NoA*). *FIN* is measured as the sum of firm's net debt amount and raised equity capital (i.e., the sum of net sales of common and preferred shares, and net long-term debt

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

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 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 (Arslan-Ayaydin, Chen, Ni, & Thewissen, 2022; Jamie Yixing Tong, 2025).

Furthermore, we include several variables to capture trends in firm performance, such as momentum (*MOMENTUM*), earnings decline (*EPSdeline*), earnings surprise (*SURPRISE*). The momentum variable is defined as the one-year cumulative abnormal return from the [-375,-10] trading day window, where the event is the ESG reporting date. *EPSdecline* equals one if the EPS declined in the current year, and zero otherwise. *SURPRISE* 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 then add counts of the firm's geographic (*GEOsegments*) and business segments (*BUSsegments*) to control for operating complexity. We also add percentage ownership of institutional investors (*INSTinvests*).

Since ESG reports are released voluntarily, their informational impact may be influenced by the timing of other corporate disclosures, particularly quarterly earnings announcements. To account for potential confounding effects, we include a set of control variables related to the release timing of ESG reports. The first variable, 1*stESGReport*, is a binary indicator equal to one if the report is the firm's first ESG disclosure within the sample period. The second variable, *DaysAfterEPS*, measures the number of calendar days between the ESG report release and the firm's most recent quarterly earnings announcement. Finally, 5*DaysAfterEPS* is an indicator equal to one if the ESG report was released within a five-day window after a quarterly earnings announcement. These variables allow us to control for the possibility that ESG disclosures coincide with or are strategically timed around earnings announcements, which may affect their perceived informativeness and influence on analyst forecast dispersion.

Furthermore, we include several variables representing the stylistic attributes of the ESG disclosures. The models include 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 (Shrestha, Arslan-Ayaydin, Thewissen, & Torsin, 2021; Strong, 1986; Thewissen, Thewissen, Arslan-Ayaydin, & Herrezuelo, 2025; Xia, Thewissen, Herrezuelo, Arslan-Ayaydin, & Yan, 2024).[7] *FOG* is measured as the average number of words per sentence added to the percentage of complex words in the document. 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 several content analysis-based variables. We control for 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, Berninger, and Kiesel (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. Furthermore, following Hope et al. (2016) and Loughran and McDonald (2014), we 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,

[7] 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 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.

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, Roulstone, and Buskirk (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 ($FWDLOOK$).

In addition, we control for several ESG ratings-related variables from MSCI, LSEG, and Bloomberg. We include the aggregate ESG scores provided by each provider: from MSCI, we use the industry-adjusted ESG score; from LSEG, we include the overall ESG score based on their proprietary weighting of environmental, social, and governance factors; and from Bloomberg, we compute the weighted average of the E, S, and G subcomponent scores. Given MSCI's broader data coverage, we also incorporate its disaggregated scores for the environmental ($EScore_{MSCI}$), social ($SScore_{MSCI}$), and governance ($GScore_{MSCI}$) pillars, as well as its emissions performance score ($EmissionScore_{MSCI}$), to control for the influence of specific ESG dimensions.[88] Finally, we include a measure of ESG ratings disagreement (*RatingsDisagreement*), constructed following Gibson Brandon, Krueger, and Schmidt (2021), defined as the standard deviation of year-specific percentile-ranked ESG scores across the three agencies. This variable captures the divergence in third-party assessments of firms' ESG performance, which prior research has linked to increased capital market uncertainty (Berg et al., 2022; Christensen et al., 2021). In addition to these control variables, we include year and industry fixed-effects.

---

[8] Our results remain robust when we use disaggregated scores from LSEG or Bloomberg instead of MSCI.

3.4. *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 recent years, the number of ESG reports increased from two in 1998 to a maximum of 1,335 in 2021 (followed by a slight dip to 1,228 in 2022). Panel B further presents the distribution of firms' first ESG reports, illustrating the temporal evolution in the adoption of ESG reporting. The number of firms initiating ESG disclosure increased gradually over time but rose sharply in 2017, with a 65.45% increase relative to the previous year. Although the growth continued in 2018, there are indications of a slowdown thereafter. Panel C of Table 1 reports the distribution of the sample by industry based on Barth, Beaver, and Landsman (1998) industry classifications. The durable manufacturers industry contains the largest number of reports and reporting firms, accounting for 18.533% of all the ESG reports and 17.726% of the firms. The computers industry is the second largest reporting industry (11.897% of all ESG reports, 13.644% of all reporting firms). On the other hand, the textile, printing and publishing has the smallest number of ESG reports in our sample (3.08% of all ESG reports), and the food industry had the fewest proportion of firms in the sample (2.507%).

< Insert Table 1 here.>

Table 2 provides the descriptive statistics of our model variables. For our dependent variable, analysts' forecast dispersion ($ADisperse_j$), we find that the average is 0.013 with the median of 0.003. Given the skewed distribution, we take the natural logarithmic value.[9]

< Insert Table 2 here.>

Concerning our control variables, the average firm in our sample has a market capitalization (*SIZE*) of $31,621 million, 67.369% leverage (*LEV*) and 3.482% volatility (*VOLAT*), and is followed by over 15 financial analysts (*NoA*). Moreover, 35.5% of firms observe

9 For observations with zero analyst dispersion, we take the log of the next smallest value.

a decrease in EPS (*EPSdecline*). The average number of business segments (*BUSsegment*) is more than two and geographical segments (*GEOsegment*) is over three. Furthermore, the mean total fraction of shares owned by US institutions is 21.1%. The average financial performance of firms in our sample (*ROA*) equals 4.236%, while the mean MSCI ESG score of firms in our sample equals 5.091 ($ESGscore_{MSCI}$), with the standard deviation of 1.993. In addition, the average ESG report contains 17,177 words and the average FOG index is 20.491, which falls into the category of difficult to read.[1010] The mean tone (*TONE*) is equal to 0.884%, which indicates that the ESG reports tend to be on the positive side. We also find that the ESG emphasis in the report is 6.542%. The disclosure also contains 7.957% of forward-looking sentences, while the number of uncertainty words equals 1.062%, out of the total number of words. Finally, 7.332% of ESG reports are disclosed within five days of the preceding EPS quarterly announcement and the average number of days between the ESG report release date and the most recent preceding quarterly EPS announcement equals 45.511.

Table 3 presents the Pearson correlations of the main variables for the full sample. The correlation between the firm's financial and ESG performance ($ESGscore_{MSCI}$) is equal to 12.4% and significant at a 99% level, which is consistent with prior literature (Dhaliwal et al., 2011, 2012; Hasan et al., 2016; X. Luo & Du, 2015). We also find that analysts' dispersion displays positive correlation with firm attributes, such as leverage, financial activity, and share-price volatility. Furthermore, analysts' dispersion is negatively correlated with firm performance indicators, such as return-on-assets, share-price momentum and ESG ratings.

< Insert Table 3 here.>

4. Topic modeling of ESG reports

Before we analyze the thematic content of ESG reports, we first discuss the

[10] The FOG index is slightly higher than Du and Yu (2020) who find a score of 18.68. 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.

implementation of sentLDA, which details the selection of the number of topics, the topic labeling, and the validation process.

4.1. *The number of topics and validation of the sentLDA output*

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 & Datta, 2014).[11] 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. Figure 1 shows that 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.[12]

< Insert Figure 1 here.>

We manually label the estimated topics to enable semantic interpretation. To ensure that

---

[11] 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).

[12] 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.

the assigned labels effectively capture human comprehension, we undertake a comprehensive procedure in two steps: (i) for each topic, we generate and evaluate a list of highest-weighted mid-length sentences and the 20 most frequent bi-grams (two-word phrases excluding stop words, numbers and symbols) in these sentences, and (ii) graphically examine the similarities between the topics based on the words they emphasize (Figure 2a displays a network graph where each weighted line represents the correlation between adjoining topics based on the weights assigned to each word). Each step is comprehensively detailed in the Appendix. Based on this analysis, Table 4 provides the labels of each topic and briefly describes the 30 topics. 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.

< Insert Figure 2 here.>

< Insert Table 4 here.>

4.2. ESG Disclosure thematic content: Structure, co-occurence and trends

4.2.1. Thematic composition of ESG disclosures

Table 4 reports the summary statistics of the various topics' variables. The topics are assigned to one of E, S, and G pillars. The *Environmental* cluster includes the following topics: *ClimateStrategy*, *EmissionsReduction*, *Emissions*, *EnvironmentalImpact*, *Recycling*, *WaterManagement*, *BiodiversityClimate*, *EnergyEfficiency*, *ProductSustainability*, and *EnvironmentalPerformance*. To form the *Social* cluster, we group the following topics: *EmployeeSafety*, *EmployeeWellbeing*, *EmployeeTraining*, *WorkerCondition*, *Workforce*, *HealthSafety*, *SupplyChainManagement*, *Communities*, *Education*, *V olunteering*, *SocialImpact*, *DataProtection*, *CustomerProduct*. Furthermore, the *Governance* cluster includes *Transparency*, *CorporateGovernance*, *StakeholderManagement*, *CodeOf Conduct*, and *SustainableFinancing*. We also identify two prevalent non-ESG topics in ESG disclosures—specifically, sentences concerning communication of financial performance and sentences referring to other reports or sections within the ESG report, which we label as *OtherInformation*.

Overall, we find that *EmployeeTraining* (32.685 sentences mean) is the most frequent topic in ESG reports, followed by *StakeholderManagement* (31.685 sentences mean). The least discussed ESG topic is *BiodiversityClimate*, with an average of over twelve sentences. Interestingly, despite ESG reports' objective to demonstrate firms' actions and strategies on societal and environmental issues, we find that there is a significant amount of information–almost twelve sentences on average–dedicated to communicating financial concerns and achievements. Hence, contrary to the wider assumptions, ESG reports are not just about ESG activities; they also communicate issues related to corporate financial performance, strategy and management. Moreover, the standard deviations of most topics are substantial in comparison to the mean, indicating much variability in the information composition across ESG reports.

The average report discusses social topics the most, with around 266 sentences, then environmental issues (183.118 sentences on average), followed by governance topics (106.327 sentences on average). Examining the diversity of topics in ESG reports, we find that the contents of ESG reports are generally fairly distributed across the E, S, and G components. We take the number of unique topics in the disclosures as the measure of diversity and observe that a sample average of 24.257 topics, which indicates that the topic content of ESG reports across each ESG pillar is fairly diverse.

4.2.2. Topic co-occurrence

We further examine which topics are more likely to appear together in the same ESG report. Figure 2b reports a network graph with each weighted line indicating the degree of co-occurrence between topics within documents (for ease of interpretation, correlations under 45% are not displayed). Figure 2b shows that, among the topics, *ClimateStrategy*, *Emissions*, *Communities*, and *CodeOfConduct*, occupy a more central position in the network, indicating that these topics generally co-occur with other information. We also find that a variety of less-discussed topics gravitate around these central themes, which highlights that the average report discusses a large array of smaller topics, such as *Recycling*, *CustomerProduct* or *EnergyEfficiency*. We observe, for example, that the topic *WorkerCondition* regularly appears with sentences concerning *Volunteering* (correlation: 50.952%), while sentences discussing the *DataProtection* and *CustomerProduct* are often found in the same reports (correlation: 53.760%). Among all topics, the two that appear most frequently together are the topics *BiodiversityClimate* and *OtherInformation*$_1$, which concerns description of financial performance (correlation: 74.603%). Other topics tend to differentiate themselves. For instance, topics such as *SustainableFinancing* and *ProductSustainability* appear to be less correlated with other ESG topics.

4.2.3. Evolution of ESG reporting themes over time

Examining topic distributions over time, we find that ESG reporting practices have evolved substantially across the sample period. As shown in Figure 3, ESG reports exhibit notable thematic shifts. Figure 3a reveals that, apart from the earliest years, ESG reports have become increasingly dominated by content related to Social (*S*) issues. Environmental (*E*) disclosures grew between 2005 and 2012 but have since declined steadily. Interestingly, this trend coincides with the founding of the Sustainability Accounting Standards Board (SASB) in 2011, which aimed to guide financially material, industry-specific ESG disclosures. Following SASB's establishment, we observe a shift in emphasis: a relative decline in *E* disclosures, a steady rise in *G*, and a marked increase in *S*, suggesting a rebalancing in disclosure practices toward governance and especially social concerns. Notably, while the Paris Agreement (2015) was a landmark international commitment to climate action, it appears to have had limited observable impact on the volume of environmental content in ESG reports. In contrast, the COVID-19 crisis triggered a further acceleration in *S*-related reporting, as firms increasingly addressed issues of workforce safety, equity, and social resilience. This upward trajectory of *S* aligns with the growing importance of corporate social commitment to investors (Breeze, 2013), as well as recent SEC initiatives encouraging disclosures on human capital strategy and risk (Securities and Exchange Commission, 2021). It also mirrors evidence from Pérez, Hunt, Samandari, Nuttall, and Biniek (2022), who document a significant rise in social-themed shareholder proposals and corporate reporting.[13]

< Insert Figure 3 here.>

Furthermore, as shown in Figure 3b, the overall length of ESG reports, measured by sentence count, has declined since 2008. While both *E* and *S* content have contracted at a similar rate in recent years, the volume of governance-related (*G*) sentences has remained stable. This

[13] Pérez et al. (2022) report a 37% increase in social-related shareholder proposals in 2021.

persistent emphasis on *G* aligns with a broad consensus on the financial materiality of governance practices (Bhagat & Bolton, 2009; Paniagua, Rivelles, & Sapena, 2018).

5. The Informational Value of ESG Report Content: Topic Relevance, Diversity, and Report Novelty

5.1. The informational value of ESG reports' thematic content

Building on the identified topic distributions, we now test Hypothesis 1, which examines the informativeness of ESG reports by assessing how the thematic emphasis on E, S and G topics influences information asymmetry, proxied by the dispersion of analysts' earnings forecasts. We estimate a multivariate regression model, as specified in Equation 4, and winsorize all topic proportion variables at the 1^{st} and 99^{th} percentiles to mitigate the influence of outliers and extreme report length. Table 5, Model (1) presents the baseline specification without topic variables (corresponding to Equation 3), while Model (2) extends the analysis by including the E, S, and G topic aggregates (Equation 4). In Model (3) through (8), we progressively add ESG ratings-related variables, such as ESG scores and emissions scores, to maintain sample size in light of differential data coverage. Model (9) reports the full specification, including all control variables. For each model, Panel A presents the estimated regression coefficients, and Panel B reports ANOVA-based model comparisons, estimating the improvement in model fit via reductions in residual sum of squares (RSS) relative to the baseline model (Equation 3).

< Insert Table 5 here.>

In Model (1), we find that several control variables are significantly associated with analysts' forecast dispersion. Firm characteristics such as size (*SIZE*), institutional ownership (*INSTInvestors*), and profitability (*ROA*) are negatively related to dispersion, consistent with the view that larger, more profitable firms with stronger institutional monitoring provide clearer and more predictable information environments. In contrast, leverage (*LEV*), return volatility (*V OLAT*), and R&D intensity (*RD*) show positive associations with dispersion, suggesting that financial complexity and risk introduce interpretive challenges for analysts. These results align with prior evidence that such firms tend to elicit greater disagreement in earnings expectations (e.g., Barron, Kim, Lim, & Stevens, 1998; Lang & Lundholm, 1996). Turning to ESG disclosure characteristics, we observe that report length (WC) is positively associated with dispersion, consistent with the notion that longer reports may introduce informational overload rather than clarity (Christensen et al., 2021). Similarly, higher complexity, reflected in lower readability (FOG) and greater uncertainty tone ($UNCERTAIN$), is also associated with increased dispersion, highlighting the role of narrative opacity in reducing transparency. In contrast, linguistic features that convey intention and clarity, including positive sentiment ($POSITIVE$), forward-looking language (*FWDlooking*), and ESG-specific terminology (*ESGemph*), are negatively associated with dispersion. These findings suggest that certain stylistic and rhetorical strategies can enhance the informativeness of ESG disclosures by providing more directionally useful cues to analysts.

In Model (2) of Panel A of Table 5, we incorporate the aggregate topic proportions for E, S and G content to assess their individual relationship with analysts' forecast dispersion. Panel B reports the results of ANOVA-based model comparisons, evaluating the explanatory gain from including thematic content. Specifically, we compare the residual sum of squares (RSS) from the baseline specification (Equation 3) to the extended model that includes both topic variables and controls (Equation 4). The RSS declines significantly with the inclusion of E, S, and G topics, and the improvement is statistically significant at the 99% confidence level.

This result remains robust after the reduction in sample size when incorporating additional controls, including ESG ratings from MSCI, Bloomberg, and LSEG, as well as CO_2 emissions scores and ESG ratings disagreement (Model (3) to (8)). In Model (3) and (4), we incorporate MSCI's CO_2 emissions score and the aggregate ESG score, respectively. Across these models, we find in Panel A of Table 5 that ESG ratings are generally negatively associated with analysts' forecast dispersion, indicating that higher external ESG performance assessments are associated with greater informational clarity. In Model (5) and (6), we substitute to ESG ratings from LSEG and Bloomberg, respectively. While the Bloomberg ESG score shows a significant negative relationship with dispersion, the LSEG score, however, does not. In Model (7), we introduce a measure of ratings disagreement – the standard deviation of normalized ESG scores across the three agencies – and find it is positively associated with dispersion, consistent with prior evidence that diverging ESG assessments signal increased market uncertainty (e.g., Christensen et al., 2021; Liu, Dai, Dong, & Liu, 2024). As shown in Panel B, the inclusion of ESG topic variables results in a statistically significant reduction in the residual sum of squares (RSS) for Models (3) to (8), indicating a robust improvement in model fit.

These findings indicate that the thematic composition of ESG reports provides, on average, incremental explanatory power beyond standard ESG-report characteristics, firm-level controls, and fixed effects identified in prior literature. Importantly, across all model specifications, the estimated coefficients for the E, S, and G topic variables remain consistent in sign and significance, despite variation in sample size due to differential ratings coverage.

5.2. *Differential informativeness of E, S, and G disclosures*

In line with Hypothesis 2, Model (2) shows that the relationship between ESG disclosures and analysts' forecast dispersion varies systematically across environmental, social, and governance themes. Specifically, environmental content (E) is significantly negatively associated with dispersion ($\beta = -0.048$, $p < 0.01$), while social content (S) is significantly positively associated ($\beta = 0.046$, $p < 0.01$). Governance content (G), by contrast, is not significantly related to dispersion ($\beta = -0.009$, $p > 0.1$). From the estimated coefficients, we derive that one standard deviation increase in E sentences corresponds to an 8.79% decrease in the dispersion of analysts' earnings forecasts. Similarly, a one standard-deviation increase in S sentences is associated with a 12.26% increase in the dispersion of analysts' earnings forecast, indicating an economically meaningful relationship between managerial choices in ESG disclosure topics and information asymmetry.

The negative coefficient for environmental disclosures reflects the relatively standardized, quantifiable, and financially material nature of topics such as carbon emissions, energy efficiency, and climate-related risks. These topics are increasingly governed by formal global reporting frameworks, including the GHG Protocol and the Task Force on Climate-related Financial Disclosures (TCFD), which enhance comparability, reduce ambiguity, and support convergence in analysts' interpretations (He, Luo, Shamsuddin, & Tang, 2022). In contrast, the positive association between social disclosures and forecast dispersion suggests that social themes, such as labor practices, diversity, or community engagement, introduce greater interpretive uncertainty. These topics tend to be more qualitative, less consistently defined across firms, and more susceptible to symbolic reporting (e.g., bluewashing), thereby increasing the scope for divergent analyst interpretations and forecast disagreement. Governance disclosures, while important, may not significantly relate to dispersion because most material governance information is already mandated through regulatory filings (e.g., proxy statements or 10-Ks), and governance structures often exhibit less cross-firm variability.

This pattern aligns with prior research emphasizing that the informativeness of disclosure is conditional on its clarity and verifiability. As Cookson and Niessner (2020) argue, greater disclosure can paradoxically increase information asymmetry when it introduces complexity or interpretive ambiguity. Similarly, Lamont (2012) highlights that nascent or loosely institutionalized domains, such as many social disclosure categories, lack common evaluative frameworks, resulting in heterogeneous market interpretations. Together, these findings show that not all topics in ESG reports are equally decision-useful.

Further supporting this interpretation, Table 4 reports the average number of named and numeric entities per sentence, used as a proxy for sentence-level specificity, across E, S, and

G content. We find that environmental disclosures are the most specific, averaging 0.947 entities per sentence, followed by social (0.844), and governance (0.617). This higher specificity of environmental topics likely contributes to their greater informativeness, as they offer clearer, more verifiable signals to capital market participants, which thus helps promote consensus among analysts.

5.2.1. Granular analysis at ESG sub-category level

We extend our main results in Table 6, reporting the relationships between the granular topic variables ($Topic_{k,t}$) and analyst forecast dispersion. To enhance interpretability and reduce semantic overlap among narrowly defined topics, we consolidate them into broader sub-groups informed by leading ESG disclosure frameworks, such as the European Sustainability Reporting Standards (ESRS). This classification approach allows us to align topic model outputs with widely recognized disclosure themes. For instance, as shown in 4, we combine topics such as *EmployeeSafety*, *EmployeeWellbeing*, *EmployeeTraining*, *Workforce*, and *HealthSafety* into the sub-category Workforce (S1), capturing firm-level human capital practices. Similarly, in the social dimension, in the governance dimension, topics such as *Trasparency*, *CorporateGovernance*, and *StakeholderManagement* are grouped under Corporate Governance (G1).

< Insert Table 6 here.>

In Model (1), (2) and (3) of Table 6, we first present the regression results examining the relationships between analysts' forecast dispersion and the E, S, and G sub-categories separately. Model (4) includes all the E, S, and G sub-categories.

Consistent with Table 5, we find that not all types of information exhibit the same level of interpretability. Within the environmental subtopics, disclosures related to climate and biodiversity risks (Biodiversity (E4)), and circular economy (ResourceUse (E5)) are consistently significant, exhibiting negative associations with analysts' dispersion. These themes often reflect firm-specific environmental risk exposure and operational efficiency, respectively. As both dimensions are increasingly standardised through frameworks such as the GHG Protocol, they are likely to be perceived as material (see Kasperzak, Kureljusic, Reisch, & Thies, 2023). In contrast, subtopics such as climate strategy, pollution, and water resources do not show a significant relationship, suggesting these disclosures may lack specificity or are perceived as less decision-relevant by analysts.

In the social dimension, we find that disclosures related to value chain management (e.g., supplier practices, responsible sourcing) and customer-related topics (e.g., product safety, consumer protection) are positively associated with dispersion. This suggests that such disclosures introduce interpretive ambiguity, possibly due to the complex trade-offs between reputational benefits and implementation costs. For example, while investments in sustainable supply chains can lead to financial gains and reduced compliance risks, they may also involve substantial upfront costs and operational complexities (see Carter & Rogers, 2008). Similarly, product safety and quality measures can enhance consumer loyalty and company reputation but may increase product costs and the company's susceptibility to market fluctuations (see Rammer, 2023).

Lastly, for the governance subcategories, disclosures around transparency, stakeholder governance, and standard governance structures (CorporateGovernance, G1) are negatively associated with dispersion (e.g., Bhagat & Bolton, 2009; Gompers, Ishii, & Metrick, 2003). Conversely, disclosures under Business Practices (G2), including ethics and integrity statements, are not significantly associated with financial analysts' dispersion. This may be because such disclosures are often aspirational, unverifiable, or perceived as symbolic, leading analysts to question their credibility (see, e.g., Kaptein & Schwartz, 2008).

5.2.2. Period analysis

In line with the theoretical model proposed by Pastor, Stambaugh, and Taylor (2021), which demonstrates that changes in investor sustainability preferences can affect asset prices, this section explores how the informativeness of ESG topics has evolved over the sample period.[14] To capture this dynamic, the sample period is divided into four subperiods, and the relationship between E, S, and G topic variables and analyst forecast dispersion is examined within each.

The first period (1998–2011) reflects the early phase of voluntary, unstructured sustainability reporting, preceding the rise of global standard-setters such as the GRI and CDP. The second period, 2011–2014, begins with the founding of the Sustainability Accounting Standards Board (SASB) and marks the initial regulatory momentum globally, including India's National Voluntary Guidelines (2011), the German Sustainability Code (2011), and the EU's Non-Financial Reporting Directive (2014). The third period, 2015–2019, captures the widespread adoption of ESG practices following the introduction of the UN Sustainable Development Goals (SDGs) and the Paris Agreement (both in 2015), as well as the release of the GRI Standards (2016) and TCFD recommendations (2017). The final period, 2020 and after, represents the post-COVID era, marked by heightened scrutiny of corporate purpose and social impact.

The results presented in Table 6, Part B, indicate that the informativeness of ESG content varies substantially across time periods. We find the strongest relationship in 2015–2019, a period marked by sharp increases in policy alignment and public awareness of ESG issues. Panel B reports a significant rise in aggregate informativeness during this period, while Panel A shows that environmental content is most strongly associated with reduced analyst dispersion, likely reflecting greater standardization and relevance of climate-related disclosures. However, in the most recent period (2020 and after), we observe a decline in overall topic informativeness, as evidenced by a reduction in explanatory power across models. The negative relationship between

[14] In the framework of Pastor et al. (2021), a "greenium" arises as investors become increasingly willing to pay a premium for assets perceived as environmentally responsible.

environmental disclosures and dispersion weakens, while the positive association between social disclosures and dispersion becomes more pronounced. This shift may reflect increased ambiguity or symbolic signaling in ESG narratives following the COVID-19 crisis, raising increasing concerns about ESG reports' information value (Kaddouri, 2024; Kathan, Utz, Dorfleitner, Eckberg, & Chmel, 2025).

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

We next test the curvilinear relationship between topic diversity and analyst forecast dispersion following the release of ESG reports. To do so, we estimate Equation 5 and report the results in Table 7 Part A.

< Insert Table 7 here.>

In Model (1), which includes only a linear term for topic diversity, we observe a statistically significant positive coefficient, suggesting that increased thematic breadth in ESG reports is associated with higher forecast dispersion. That is, more diverse ESG content is likely to be detrimental to transparency and is therefore associated with greater information asymmetry. However, in Model (2), we include both a linear and a quadratic term for topic diversity. The results show a significant negative coefficient for the linear term (β = –0.042, $p < 0.01$) and a significant positive coefficient for the quadratic term (β = 0.001, $p < 0.05$), indicating a rather convex relationship.

The latter finding suggests that increasing topic diversity initially reduces information asymmetry by providing a more comprehensive view of the firm. However, once a certain threshold is crossed, excessive thematic breadth may introduce informational overload or interpretive ambiguity, ultimately leading to greater analyst disagreement. Supporting this, Panel B shows a significant improvement in model fit when the quadratic term is introduced. Based on the estimated coefficients, we calculate that the optimal number of unique ESG topics is 21, which corresponds to approximately 70% of the total topic set derived from our unsupervised classification model.

5.4. *First versus repeated ESG reports*

Hypothesis 4 examines the difference in the information value of the thematic content of repeated versus first-time reporters. We create two sub-samples based on whether the ESG report is the firm's first or a subsequent report. For each sub-sample, we then run Equation 3 and 4 and compare the goodness of fit based on ANOVA tests. Our sample of first-time reporters includes 1,114 firm-year observations, whereas the repeated ESG reporter sample includes 5,056 firm-year observations. Panel B of Table 7 shows that, despite the lower number of observations in the former sample, the informational value of ESG content is greater with larger effect magnitudes, particularly for S and G topic variables (First Report (S): $\beta = 0.060$, $p < 0.1$; Subsequent Report (S): $\beta = 0.037$, $p < 0.01$). Crucially, the difference in estimated coefficients for E is rather marginal, suggesting a more robust informativeness of environmental information, in support of the argument that such topics are more standardized and easier to interpret. Moreover, the model fit improves substantially, as indicated by the significant decrease in residual sum of squares ($\Delta RSS = -24.493$, $p = 0.01$). Overall, our results are in line with Hypothesis 4.

These findings support the interpretation that first-time ESG disclosures carry distinct informational value and suggest that the content of first-time reports disrupts the firm's information environment significantly more relative to subsequent reports. They also reinforce the view that repeated ESG reporting, while potentially valuable for long-term transparency, may offer diminishing informational returns in the absence of substantive updates or changes in content (Dhaliwal et al., 2012; Guidry & Patten, 2010).

Overall, these results reveal distinct patterns in the informativeness of ESG disclosures. Environmental content is interpreted more consistently by analysts and is associated with reduced forecast dispersion. In contrast, social disclosures, often more subjective and less standardized (Jason, 2020), are linked to greater information asymmetry. Governance-related content shows no robust association with forecast dispersion, likely because key governance attributes are already available through mandatory financial filings and thus provide little incremental value in ESG reports. Thematic diversity also matters: while broader coverage initially enhances informativeness by reducing analysts' disagreement, the relationship is non-linear. Beyond an optimal level of topic breadth, additional disclosure introduces complexity and ambiguity, ultimately increasing information asymmetry. Finally, the informational value of ESG reports diminishes with repeated issuance, suggesting that first-time disclosures carry the strongest signaling effect.

6. Cross-sectional Analyses – Reporting Period, Institutional Investors and Concurrent Disclosures

In this section, we extend the analysis with a series of cross-sectional tests. While our baseline results establish a robust link between ESG disclosure content and analysts' forecast dispersion, the strength and nature of this relationship may vary with firm characteristics and disclosure context. To explore these nuances, we proceed in three steps. First, we analyze whether the informativeness of ESG content depends on validation from third-party ESG rating agencies, by testing whether the relationship with forecast dispersion is moderated by firms' ESG performance scores and the consistency of ratings across agencies. Second, we examine the role of the informational environment by comparing firms with differing levels of analyst coverage and institutional ownership. Third, we investigate the effect of concurrent disclosures by assessing whether ESG reports released in close proximity to earnings announcements exhibit different levels of informativeness.

6.1. ESG performance and ratings disagreement

The informativeness of ESG disclosures is likely to vary systematically with the firm's underlying ESG profile. We therefore examine whether the effects of ESG content depend on

(i) the level of ESG performance, as captured by the MSCI ESG score, and (ii) the degree of disagreement among ESG rating agencies.

6.1.1. ESG performance.

From a signaling perspective, ESG ratings provide a measure of the credibility of a firm's sustainability claims. Firms with strong ESG scores have accumulated reputational capital and face higher costs of opportunistic disclosure, which enhances the perceived reliability of their reports (see Choi & Storr, 2022; Connelly, Certo, Ireland, & Reutzel, 2011). Analysts are therefore more likely to interpret environmental content from high-ESG firms as credible and incorporate it into their forecasts, leading to greater convergence. On the contrary, disclosures from firms with weaker ESG profiles are more easily interpreted as symbolic or even as greenwashing, which limits their effectiveness in reducing information asymmetry (see Hawn & Ioannou, 2016; Y. Kim et al., 2012). In line with our expectations, results in Part A of Table 8 show that the informativeness of ESG disclosures, particularly the environmental topics, varies systematically with firms' ESG profiles. In the sub-sample of firms with higher ESG ratings, the inclusion of ESG topic variables improves model fit, consistent with the view that reports from credible firms are perceived as more reliable and reduce information asymmetry. In contrast, when firms receive lower ESG scores, ESG disclosures contribute with less incremental value. Additionally, we observe that social topics are consistently linked to greater forecast dispersion, reflecting the difficulty in assessing the implications of such information for analysts.

< Insert Table 8 here.>

6.1.2. Ratings disagreement.

Similarly, we find that ESG disclosures are generally more informative when ratings disagreement is low. Divergence among ESG rating agencies reflects the degree of uncertainty surrounding a firm's ESG profile. Prior research documents that such disagreement is substantial and stems from differences in scope, measurement, and weighting of ESG criteria (Berg et al., 2022). In addition, high levels of disagreement signal greater information asymmetry, as external evaluators provide conflicting assessments to the market (Liu et al., 2024). In this context, firm-level ESG disclosures tend to be viewed with skepticism: analysts interpret the information inconsistently, which amplifies ambiguity and weakens the disclosures' association with analysts' forecast. In contrast, when rating agencies reach broad consensus and external signals are more coherent, ESG disclosures gain credibility and show a stronger association with analysts' forecasts.

Consistent with this reasoning, Part A of Table 8 demonstrates that the informativeness of ESG disclosures depends on the degree of consensus among rating agencies. When disagreement is high, the information conveyed in ESG reports appears less credible, while convergence among agencies enhances their credibility. Nonetheless, important differences emerge across environmental, social, and governance topics. Specifically, when ratings disagreement is low, E and G disclosures reduce forecast dispersion, and there is notably stronger negative relationship with the governance topics. However, regarding social topics, the associations are positive in both sub-samples, underscoring the persistent inconsistencies in how analysts interpret the implications of firms' social actions.

6.2. *Informational environment*

Building on the role of external ESG ratings, we next examine whether the value of ESG reports is conditioned by the broader information environment, focusing on analyst coverage and institutional ownership.

6.2.1. Financial analysts

Two competing mechanisms can be envisaged regarding the role of analyst coverage in shaping the informativeness of ESG disclosures. On the one hand, ESG reports could be most valuable when analyst following is limited. With fewer intermediaries producing and disseminating firm-specific information, information asymmetry is greater, and investors may rely more heavily on ESG reports to form expectations. This logic emphasizes a substitution effect, consistent with prior research on voluntary disclosure in settings of weak external monitoring (see, e.g., Healy & Palepu, 2001). On the other hand, ESG disclosures may become more informative as analyst coverage increases. Analysts do not merely generate independent forecasts. They also act as information intermediaries who interpret, contextualize, and transmit corporate disclosures to a broader investor base. For example, Yu (2008) finds that when firms are followed by more analysts, managers engage less in earnings management. The mechanism is that analysts provide intense scrutiny and monitoring, which disciplines managers and improves the credibility of reported information. Therefore, instead of disclosure and analysts being substitutes, non-financial disclosures gain credibility and market impact when processed and certified by sophisticated information users, including analysts (Dhaliwal et al., 2012; Hong, Huseynov, Sardarli, & Zhang, 2020; Yu, 2008).

Our findings support this second perspective. We observe that ESG disclosures reduce forecast dispersion more strongly in environments with broader analyst coverage. This suggests that ESG reports and analyst coverage function as complements rather than substitutes: analysts draw attention to ESG information, highlight its salient elements, and ensure its integration into earnings forecasts. In contrast, when analyst coverage is sparse, ESG disclosures may be discounted as less credible or underutilized by investors with limited attention. This interpretation is consistent with attention-based theories (Hirshleifer & Teoh, 2003), which emphasize that the incorporation of information into market expectations depends not only on its availability but also on the institutional channels that direct and amplify it.

6.2.2. Institutional investors

Additionally, the presence of institutional investors may play a central role in shaping firms' information environment through active monitoring and engagement with management (Boone & White, 2015; Drobetz, El Ghoul, Guedhami, & Yu, 2024). As such, we expect institutional investors to influence the informativeness of ESG disclosures. Specifically, institutional investors, through their privileged access to firm-specific insights and the potential for private communication (Admati & Pfleiderer, 2009; Edmans & Manso, 2011), may reduce the marginal value of public disclosures, including ESG reports. Furthermore, their influence on corporate disclosure practices may lead to ESG-related information being conveyed through informal or pre-emptive channels, reducing the novelty and decision-usefulness of formal ESG reporting. In line with this, Krueger, Sautner, Tang, and Zhong (2024) suggest that disclosure mandates are particularly effective in contexts with limited institutional oversight, implying that the informativeness of ESG reports is shaped by the presence of institutional investors. Thus, examining how institutional ownership moderates the value of ESG disclosures is crucial.

To test this hypothesis, we partition the sample at the median level of institutional ownership and estimate our models separately for the high- and low-ownership subgroups. As shown in Panel A of Table 8 Part B, we find that the reduction in RSS when moving from the baseline model (Equation 3) to the extended model including ESG topic variables (Equation 4) is statistically significant for firms with below-median institutional ownership ($\Delta RSS = -33.651$, $p = 0.01$). In contrast, for firms with higher institutional ownership, the change in RSS is insignificant ($\Delta RSS = -9.701$, $p = 0.23$). This indicates that, in environments with strong institutional oversight, ESG reports appear to contain less incremental value, consistent with the notion that institutional investors facilitate earlier or alternative diffusion of relevant ESG information.

6.3. Concurrent disclosures

As a final cross-sectional test, we assess whether the timing of ESG reports relative to earnings announcements moderates their informativeness. Earnings announcements are essential financial events that substantially reduce information asymmetry (Arslan-Ayaydin, Boudt, & Thewissen, 2016; Arslan-Ayaydin, Thewissen, & Torsin, 2021; Bhattacharya, Desai, & Venkataraman, 2013; Stoumbos, 2023). When ESG reports are released in close proximity to earnings news, their informational impact may be diminished, either because they are overshadowed by the earnings announcement or because relevant financial details are already incorporated into the earnings release.

To test this prediction, we classify firm-years according to whether the ESG report was issued within five days of the most recent earnings announcement. Approximately 7.3% of reports fall within this five-day window. As shown in Table 8 Part C, ESG reports released outside this window exhibit a stronger negative association with analysts' forecast dispersion, suggesting that analysts extract greater incremental value when ESG disclosures are temporally separated from earnings announcements. In contrast, when ESG and earnings reports are released in close succession, the marginal impact of ESG content is weaker, consistent with reduced analyst attention.

These findings highlight that the informativeness of ESG reports is context-dependent: when released alongside major financial disclosures, analysts appear to allocate less attention to ESG information, limiting its integration into forecasts. This interpretation is consistent with attention-based theories, which posit that information processing is constrained by limited cognitive resources, leading market participants to prioritize salient earnings news over ESG disclosures when both are presented simultaneously (see, e.g., DellaVigna & Pollet, 2009; Hirshleifer & Teoh, 2003; Kahneman, 1973).

7. Additional Analyses

In this section, we conduct three additional analyses to further investigate the mechanisms underlying the relationship between ESG topic content and analysts' forecast dispersion. First, we explore an alternative measure of ESG disclosures' informativeness by examining the relationship between the thematic content of disclosures and an external measure of ESG performance, the MSCI ESG score. Second, we investigate whether the thematic content of ESG reports is associated with ESG ratings disagreement across major rating agencies, which we use as a proxy for interpretability. Third, we examine how the specificity of ESG disclosures, in terms of named entities and numerical details, correlates with the level of information asymmetry, particularly in reducing forecast dispersion. Each of these tests help shed light on different aspects of how the content and clarity of ESG disclosures impact market interpretation and analyst behavior.

7.1. ESG performance

Model (1) in Table 9 examines the relationship between the E, S, and G topic variables and ESG performance, measured by the MSCI industry-adjusted ESG score. The results indicate that environmental topics are positively and significantly associated with ESG scores. This finding aligns with our main results in Table 5, which show that environmental disclosures reduce information asymmetry. The positive association further suggests that environmental topics are more readily interpretable, as they are often standardized, quantifiable, and comparable across firms, making them more likely to be incorporated into third-party ESG assessments. Overall, Panel B shows that including the E, S, and G topic variables substantially enhances the model's explanatory power for ESG scores.

< Insert Table 9 here.>

7.2. ESG ratings disagreement

Model (2) explores how the thematic content of ESG reports relates to ratings disagreement across three major agencies. Consistent with Gibson Brandon et al. (2021), we find that environmental topics are negatively associated with divergence in external assessments. In other words, raters tend to agree more on environmental issues, reflecting the fact that environmental disclosures are generally more standardized, quantifiable, and verifiable. This supports the view that clearer, more comparable ESG information fosters greater consistency in ratings across agencies.

Beyond their impact on analysts' forecast dispersion, our results indicate that the thematic content of ESG reports also shapes the broader information environment in which investors operate. Ratings disagreement often reflects divergent interpretations of the same information, and our evidence suggests that emphasizing the environmental pillar, which are rather quantifiable and easier to interpret, can reduce such disagreements and improve the reliability of external assessments. This has important implications for investors, who depend on ESG ratings to guide decisions and manage uncertainty in the market.

Additionally, Panel B shows that including the E, S, and G topic variables significantly improves model performance, highlighting that ESG report content helps explain discrepancies in ratings across agencies.

7.3. *Specificity of ESG disclosures*

Finally, in Model (3), we examine the average level of specificity in sentences discussing E, S, and G topics as explanatory variables. Specifically, we analyze how increased specificity, measured through the inclusion of named entities and numerical details, impacts analysts' dispersion. Among the three ESG pillars, the specificity of social topics is significant (albeit marginally) in relation to analysts' forecast dispersion. This suggests that providing concrete details, such as names of people and places, is particularly useful for improving the interpretability of social disclosures, which tend to be less standardized than environmental and governance topics.

For example, stating "Employee turnover fell from 12% to 8% in Q1 at our Chicago office" is much clearer than a vaguer statement like "turnover improved," which leaves much more room for interpretation. Without specific names, dates, locations, or quantities, audiences are left to fill in the gaps with their own assumptions, thereby increasing dispersion and fueling information asymmetry. Named entities provide tangible, observable reference points that are often deemed more reliable and harder to challenge (Petty & Cacioppo, 1986). Our findings support the argument that specificity enhances the interpretability of social disclosures, which are inherently more difficult to verify. This aligns with our central thesis that social issues, due to their inherent vagueness, are more prone to greater information asymmetry.

8. Conclusion

ESG reports are now a prominent part of corporate disclosure, as ESG criteria become ever more important in investors' capital allocation considerations (Eccles et al., 2012). However, disclosures vary widely in structure and focus, often without standardized information (Berg et al., 2022); and moreover, market participants also rely on concurrent sources such as ESG ratings and emissions databases (Bofinger et al., 2022; Christensen et al., 2021; Kimbrough et al., 2024). While some studies show ESG reporting improves transparency, reduces ratings disagreement, and lowers information asymmetry and cost of capital (e.g., Cui, Jo, & Na, 2016; Dhaliwal et al., 2011, 2014), others warn of greenwashing, investor confusion, and negligible financial impact (e.g., Berg et al., 2022; Gibson Brandon et al., 2021; E.-H. Kim & Lyon, 2015). In this context, we provide new evidence on how the thematic content of ESG reports relates to information asymmetry, proxied by analysts' earnings-forecast dispersion.

Using a hand-collected dataset of ESG reports form 1998 and 2023, we first develop a taxonomy of prominent ESG topics and document substantial thematic heterogeneity across firms' disclosures. Departing from prior work that focuses on the format or volume of disclosure (Cormier & Magnan, 2013; Du & Yu, 2020), we apply machine-learning methods (Bao & Datta, 2014) to examine *what* is disclosed. Our results show that: (1) the three E, S, and G dimensions vary markedly in their association with forecast dispersion, and there are further differences within the same ESG pillar; (2) greater topical diversity generally reduces information asymmetry, though the marginal benefit declines beyond a certain threshold; and (3) these patterns hold in sub-samples defined by first versus subsequent reports, different time periods, levels of institutional ownership, and timing around earnings announcements. Collectively, our findings underscore that audiences respond not only to whether ESG information is provided, but critically to which themes are addressed.

Consistent with topic modeling research in finance (Bellstam et al., 2020; Hoberg & Lewis, 2017; Shrestha, Thewissen, Arslan-Ayaydin, & Parhankangas, 2023; Thewissen, Shrestha, Torsin, & Pastwa, 2022), our analysis is descriptive and establishes associations rather than causal effects. Nevertheless, as ESG reporting proliferates and regulatory scrutiny intensifies, our insights have practical implications for investors, academics, and policymakers. The replicable, automated classification of ESG topics can help regulators monitor the substance of corporate disclosures and identify the themes most valued by investors. Future research could extend this approach to an international setting (Aerts et al., 2008; Hummel, Mittelbach-Hörmanseder, Cho, & Matten, 2022). It could also employ sentence-level topic models (e.g., sentLDA) to develop more fine-grained textual measures of specific ESG initiatives, or investigate the firm- and manager-level determinants of topic selection through the lens of obfuscation theory

(Schrand & Walther, 2000).

References

Admati, A. R., & Pfleiderer, P. (2009). The "wall street walk" and shareholder activism: Exit as a form of voice. *The Review of Financial Studies*, *22*(7), 2645–2685.

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

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

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

Arslan-Ayaydin, Ö, Chen, S., Ni, S. X., & Thewissen, J. (2022). Is cross-listing a panacea for improving earnings quality? the case of h- and b-share firms in China. *International Review of Financial Analysis*, *81* , 102113.

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

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

Azarbonyad, H., Dehghani, M., Kenter, T., Marx, M., Kamps, J., & de Rijke, M. (2017). Hierarchical re-estimation of topic models for measuring topical diversity. In Advances In Information Retrieval: 39th European Conference on IR Research, ECIR 2017, Aberdeen, UK, April 8-13, 2017, Proceedings 39 (pp. 68–81).

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

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

Barron, O. E., Kim, O., Lim, S. C., & Stevens, D. E. (1998). Using analysts' forecasts to measure properties of analysts' information environment. *The Accounting Review*, *73*(4), 421–433.

Barth, M. E., Beaver, W. H., & Landsman, W. R. (1998). 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., & McNichols, M. F. (1994). Estimation and market valuation of environmental liabilities relating to superfund sites. *Journal of Accounting Research*, *32*, 177.

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

Berg, F., Kölbel, J., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, *26*(6), 1315–1344.

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

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

Bhattacharya, N., Desai, H., & Venkataraman, K. (2013). Does earnings quality affect information asymmetry? Evidence from trading costs. *Contemporary Accounting Research*, *30* (2), 482–516.

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

BNP Paribas Securities Services. (2021). The ESG Global Survey 2021: The path to ESG no turning back for asset owners and managers. Paris. Retrieved from https://securities.cib.bnpparibas/app/ uploads/sites/3/2021/09/the-esg-global-survey-2021.pdf (Accessed 9 May 2025)

Bofinger, Y., Heyden, K. J., & Rock, B. (2022). Corporate social responsibility and market efficiency: Evidence from ESG and misvaluation measures. *Journal of Banking & Finance*, *134*, 106322.

Boone, A. L., & White, J. T. (2015). The effect of institutional ownership on firm transparency and information production. *Journal of Financial Economics*, *117*(3), 508-533.

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

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

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

Breeze, R. (2013). Corporate Discourse (Vol. 1). A&C Black.

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

Carter, C. R., & Rogers, D. S. (2008). A framework of sustainable supply chain management: moving toward new theory. *International Journal of Physical Distribution & Logistics Management*, *38*(5), 360–387.

Chapman, K. L., Reiter, N., White, H. D., & Williams, C. D. (2019). Information overload and disclosure smoothing. *Review of Accounting Studies*, *24*, 1486–1522.

Chen, Y.-C., Hung, M., & Wang, Y. (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.

Cho, C. H., Guidry, R. P., Hageman, A. M., & Patten, D. M. (2012). Do actions speak louder than words? An empirical investigation of corporate environmental reputation. *Accounting, Organizations and Society*, *37*(1), 14–25.

Choi, G. S., & Storr, V. H. (2022). The market as a process for the discovery of whom not to trust. *Journal of Institutional Economics*, *18*(3), 467–482.

Christensen, D. M., Serafeim, G., & Sikochi, A. (2021). 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., Fang, X., Li, Y., & Richardson, G. (2013). The relevance of environmental disclosures: Are such disclosures incrementally informative? *Journal of Accounting and Public Policy*, *32*(5), 410–431.

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

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, *37*(1), 39–67.

Cookson, J. A., & Niessner, M. (2020). Why don't we agree? Evidence from a social network of investors. *The Journal of Finance*, *75*(1), 173-228.

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

Cormier, D., & Magnan, M. (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., Jo, H., & Na, H. (2016). Does corporate social responsibility affect information asymmetry? *Journal of Business Ethics*, *148* (3), 549–572.

DellaVigna, S., & Pollet, J. M. (2009). Investor inattention and Friday earnings announcements. *The Journal of Finance*, *64* (2), 709–749.

Dhaliwal, D., Li, O., Tsang, A., & Yang, Y. (2011). Voluntary nonfinancial disclosure and the cost of equity capital: The initiation of corporate social responsibility reporting. *The Accounting Review*, *86* (1), 59– 100.

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

Diether, K. B., Malloy, C. J., & Scherbina, A. (2002). Differences of opinion and the cross section of stock returns. *The Journal of Finance*, *57* (5), 2113–2141.

Drobetz, W., El Ghoul, S., Guedhami, O., & Yu, X. (2024). Beyond ownership: The role of institutional investors in international corporate governance. *Corporate Governance: An International Review*.

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

Dye, R. A. (1985). Disclosure of nonproprietary information. *Journal of Accounting Research*, *23* (1), 123–145.

Dyer, T., Lang, M., & Stice-Lawrence, L. (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., Ioannou, I., & Serafeim, G. (2012). The impact of corporate culture of sustainability on corporate behavior and performance. (Working paper, Harvard Business School, Harvard University, Boston, MA)

Edmans, A., & Manso, G. (2011). Governance through trading and intervention: A theory of multiple blockholders. *The Review of Financial Studies*, *24* (7), 2395–2428.

Edmunds, A., & Morris, A. (2000). The problem of information overload in business organisations: a review of the literature. *International Journal of Information Management*, *20* (1), 17–28.

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

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

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

Gompers, P., Ishii, J., & Metrick, A. (2003). Corporate governance and equity prices. *The Quarterly Journal of Economics*, *118* (1), 107–156.

Governance Accountability Institute, Inc. (2024). Navigating the way to sustainability. Retrieved from https://www.ga-institute.com/index.php?id=9127

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

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

Guidry, R. P., & Patten, D. M. (2010). Market reactions to the first-time issuance of corporate sustainability reports: Evidence that quality matters. *Sustainability Accounting, Management and Policy Journal*.

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

Hawn, O., & Ioannou, I. (2016). Mind the gap: The interplay between external and internal actions in the case of corporate social responsibility. *Strategic Management Journal*, *37* (13).

He, R., Luo, L., Shamsuddin, A., & Tang, Q. (2022). Corporate carbon accounting: a literature review of carbon accounting research from the Kyoto protocol to the Paris agreement. *Accounting & Finance*, *62* (1), 261–298.

Healy, P. M., & Palepu, K. G. (2001, September). 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.

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

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

Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, *36* (1-3), 337–386.

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

Holder-Webb, C. J. N. L., L., & Wood, D. (2008). A survey of governance disclosures among U.S. firms. *Journal of Business Ethics*.

Hong, Y., Huseynov, F., Sardarli, S., & Zhang, W. (2020). Bank earnings management and analyst coverage: evidence from loan loss provisions. *Review of Quantitative Finance and Accounting*, *55*(1), 29–54.

Hope, O.-K. (2003). Disclosure practices, enforcement of accounting standards, and analysts' forecast accuracy: An international study. *Journal of Accounting Research*, *41*(2), 235–272.

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

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

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

Ivers, M. (2010). Random house guide to good writing. Ballantine Books.

Jamie Yixing Tong, S. N. S. C., James Thewissen. (2025). Ceo retirement and cost stickiness. *Journal of International Accounting, Auditing and Taxation*.

Jason, S. (2020). Fixing the s in ESG. *Stanford Social Innovation Review*. Retrieved from https://doi.org/ 10.48558/D4WJ-QF36

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

Kaddouri, O. (2024). The new normal of corporate sustainability: A post-pandemic perspective on greenwashing discourse and manager perceptions. *European Management Journal*.

Kahneman, D. (1973). Attention And Effort (Vol. 1063). Citeseer.

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

Kaptein, M., & Schwartz, M. S. (2008). The effectiveness of business codes: A critical examination of existing studies and the development of an integrated research model. *Journal of Business Ethics*, *77*, 111–127.

Kasperzak, R., Kureljusic, M., Reisch, L., & Thies, S. (2023). Accounting for carbon emissions—current state of sustainability reporting practice under the GHG protocol. *Sustainability*, *15*(2), 994.

Kathan, M. C., Utz, S., Dorfleitner, G., Eckberg, J., & Chmel, L. (2025). What you see is not what you get: Esg scores and greenwashing risk. *Finance Research Letters*, *74*, 106710.

Kim, E.-H., & Lyon, T. P. (2015). Greenwash vs. brownwash: Exaggeration and undue modesty in corporate sustainability disclosure. *Organization Science*, *26*(3), 705–723.

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

Kim, Y., & Son, C. (2022). Evaluation of online communities for technology foresight: Data-driven approach based on expertise and diversity. *Sustainability*, *14*(20), 13040.

Kimbrough, M. D., Wang, X., Wei, S., & Zhang, J. (2024). Does voluntary ESG reporting resolve disagreement among ESG rating agencies? *European Accounting Review*, *33*(1), 15–47.

Krueger, P., Sautner, Z., Tang, D. Y., & Zhong, R. (2024). The effects of mandatory ESG disclosure around the world. *Journal of Accounting Research*, *62*(5), 1795-1847.

Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, *38*(Volume 38, 2012), 201-221.

Lang, M. H., & Lundholm, R. J. (1996). Corporate disclosure policy and analyst behavior. *The Accounting Review*, *71*(4), 467–492.

Liu, X., Dai, J., Dong, X., & Liu, J. (2024). Esg rating disagreement and analyst forecast quality. *International Review of Financial Analysis*, *95*, 103446.

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

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

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

Luo, Y., Ni, C., & Thewissen, J. (2024). Currency flotation and dividend policies: Evidence from china's central parity reform. *Financial Management*, *53*(1), 145-174.

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

McGuire, J. B., Sundgren, A., & Schneeweis, T. (1988). Corporate social responsibility and firm financial performance. *Academy of Management Journal*, *31*(4), 854–872.

Miihkinen, A. (2012). What drives quality of firm risk disclosure? The impact of a national disclosure standard and reporting incentives under ifrs. *The International Journal of Accounting*, *47*(4), 437–468.

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

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

Pastor, L., Stambaugh, R. F., & Taylor, L. A. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics*, *142*(2), 550-571.

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

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model Of Persuasion. Springer.

Pérez, L., Hunt, V., Samandari, H., Nuttall, R., & Biniek, K. (2022). Does ESG really matter—and why? *The McKinsey Quarterly*.

Rammer, C. (2023). Measuring process innovation output in firms: Cost reduction versus quality improvement. *Technovation*, *124*, 102753.

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.

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

Schrand, C. M., & Walther, B. R. (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. (2021). Annual regulatory agenda. *Press Release*. Retrieved from https://www.sec.gov/news/press-release/2021-99

Shrestha, P., Arslan-Ayaydin, Ö., Thewissen, J., & Torsin, W. (2021). Institutions, regulations and initial coin offerings: An international perspective. *International Review of Economics & Finance*, *72*, 102–120.

Shrestha, P., Thewissen, J., Arslan-Ayaydin, Ö., & Parhankangas, A. (2023). A sense of risk: Responses to crowdfunding risk disclosures. *Strategic Entrepreneurship Journal*, *17* (4), 925–970.

Stoumbos, R. (2023). The growth of information asymmetry between earnings announcements and its implications for reporting frequency. *Management Science*, *69* (3), 1901–1928.

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

Thewissen, J., Shrestha, P., Torsin, W., & Pastwa, A. M. (2022). Unpacking the black box of ICO white papers: A topic modeling approach. *Journal of Corporate Finance*, *75*, 102225.

Thewissen, J., Thewissen, J., Arslan-Ayaydin, Ö., & Herrezuelo, D. B. (2025). Hitting the right note: The impact of lexical complexity on initial coin offering success. *Financial Review*.

Thewissen, J., Thewissen, J., Torsin, W., & Arslan-Ayaydin, Ö (2022). Linguistic errors and investment decisions: the case of ICO white papers. *The European Journal of Finance*, 1–43.

Tong, J. Y., Thewissen, J., Ni, S. X., & Chen, S. (2025). CEO retirement and cost stickiness. *Journal of International Accounting, Auditing and Taxation*, 100738.

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

Unerman, J., & O'Dwyer, B. (2007). The business case for regulation of corporate social responsibility and accountability. *Accounting Forum*, *31*(4), 332–353.

Verrecchia, R. E. (2001). Essays on disclosure. *Journal of Accounting and Economics*, *32*(1), 97-180.

Wang, K. T., & Li, D. (2016). Market reactions to the first-time disclosure of corporate social responsibility reports: Evidence from China. *Journal of Business Ethics*, *138*, 661–682.

Xia, F., Thewissen, J., Herrezuelo, D. B., Arslan-Ayaydin, Ö., & Yan, S. (2024). Picturing success: The informational role of visual cues in initial coin offerings. *The British Accounting Review*, 101503.

Xin, Y., Thewissen, J., Tsang, A., & Yan, S. (2025). Board gender diversity reforms and corporate social responsibility: International evidence. *Journal of Behavioral and Experimental*

Finance, *45*, 101025.

Yu, F. F. (2008). Analyst coverage and earnings management. *Journal of Financial Economics*, *88* (2), 245–271.

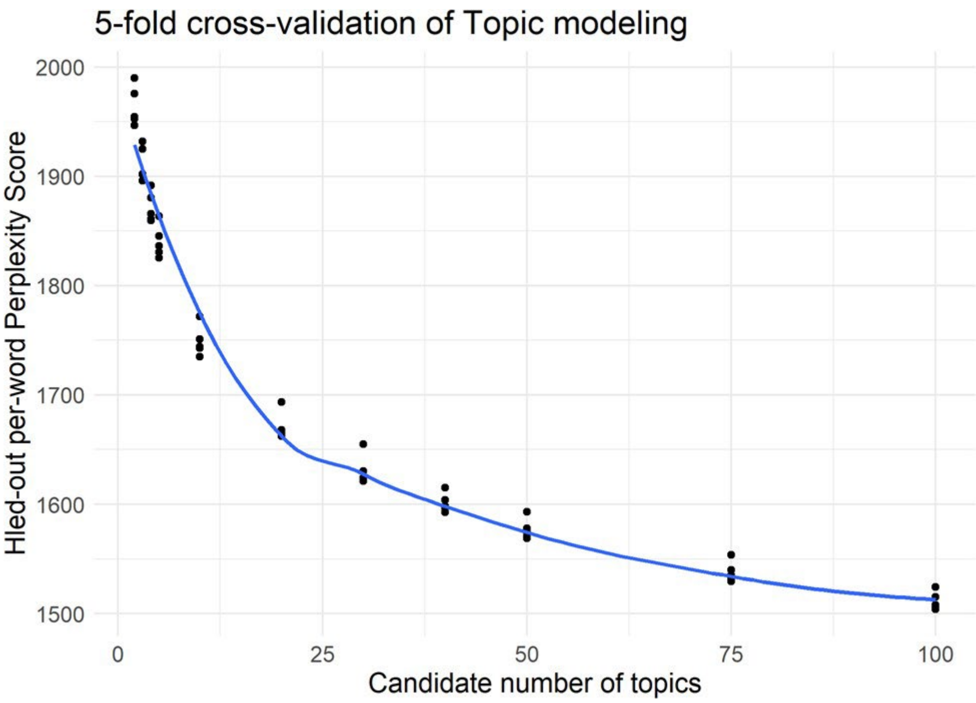
Zhang, Y., Jin, R., & Zhou, Z.-H. (2010). Understanding bag-of-words model: A statistical framework. *International Journal of Machine Learning and Cybernetics*, *1* (1-4), 43–52.

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

Zulu-Chisanga, S. (2019). 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.

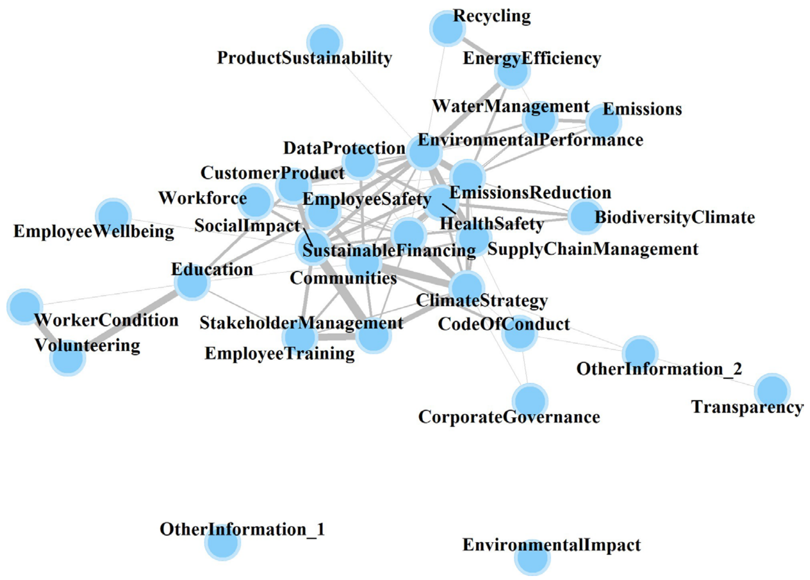
9. Tables and Figures

Figure 1.: **Held-Out per-Word Perplexity as a Function of the Number of Topics**

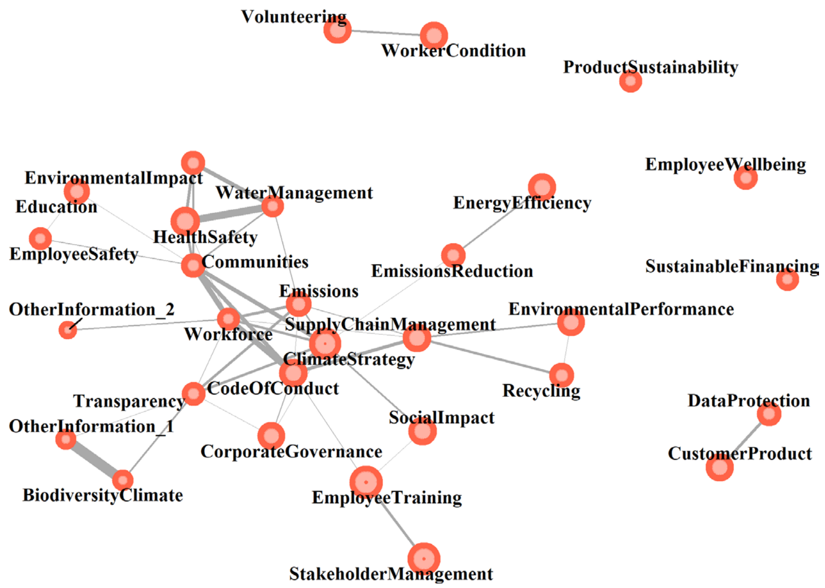


Note: The figure plots the perplexity scores of the estimated sentLDA model at different numbers of topics from 1 to 100. The procedure is detailed in Section 4.1 of the manuscript.

Figure 2.: **Network graphs on topic interlinkages**



(a) Word correlation among topics

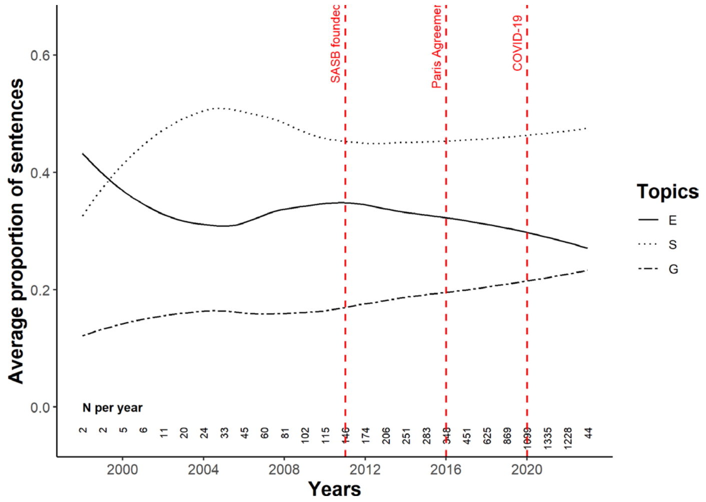


(a) Co-occurrence of topics across documents

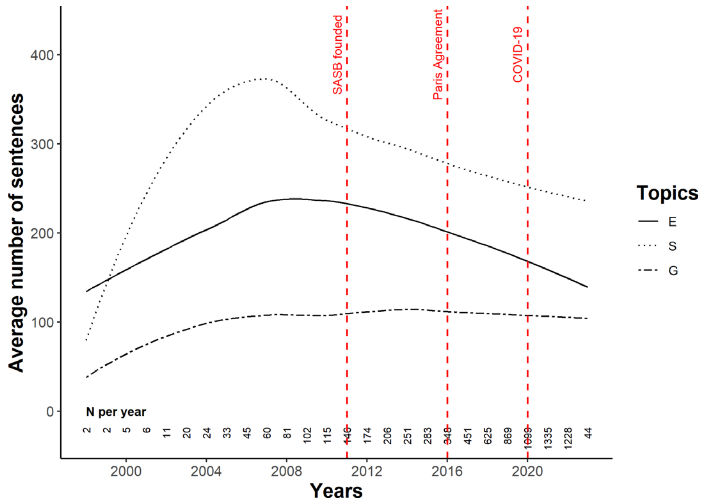
Note: Figure 2a presents a network graph illustrating the correlation between the ESG topics with respect to the weights assigned to the constituting words. Figure 2b presents a network graph illustrating the correlation between the ESG topics with respect to their co-occurrences across documents. The size of the nodes is proportioned to represent the number of sentences assigned to the given topic in the total sample.

Figure 3.: **Evolution of ESG thematic content between 2007 and 2020**

(a) **Panel A: Average E, S, and G proportions over time**



(b) **Panel B: Average E, S, and G sentences over time**



Note: The figures depict the temporal evolution of the E, S, and G components in the sample ESG disclosures. Figure 3a represents the E, S, and G composition in average yearly proportions of the total number of sentences, and figure 3b presents it in average yearly total number of sentences dedicated to each category.

Table 1.: **Sample distribution**

Panel A: Sample distribution by year

Year	Frequency	Percent	Cumulative frequency	Cumulative percent
1998	2.000	0.026	2.000	0.026
1999	2.000	0.026	4.000	0.053
2000	5.000	0.066	9.000	0.119
2001	6.000	0.079	15.000	0.198
2002	11.000	0.145	26.000	0.344
2003	20.000	0.264	46.000	0.608
2004	24.000	0.317	70.000	0.925
2005	33.000	0.436	103.000	1.362
2006	45.000	0.595	148.000	1.956
2007	60.000	0.793	208.000	2.750
2008	81.000	1.071	289.000	3.820
2009	102.000	1.348	391.000	5.169
2010	115.000	1.520	506.000	6.689
2011	146.000	1.930	652.000	8.619
2012	174.000	2.300	826.000	10.919
2013	206.000	2.723	1032.000	13.642
2014	251.000	3.318	1283.000	16.960
2015	283.000	3.741	1566.000	20.701
2016	348.000	4.600	1914.000	25.301
2017	451.000	5.962	2365.000	31.262
2018	625.000	8.262	2990.000	39.524
2019	869.000	11.487	3859.000	51.011
2020	1099.000	14.527	4958.000	65.539
2021	1335.000	17.647	6293.000	83.186
2022	1228.000	16.233	7521.000	99.418
2023	44.000	0.582	7565.000	100.000

Panel B: Sample distribution by year for first ESG reports

Year	Frequency	Percent	Cumulative frequency	Cumulative percent
1998	2.000	0.117	2.000	0.117
1999	1.000	0.058	3.000	0.175
2000	2.000	0.117	5.000	0.292
2001	3.000	0.175	8.000	0.466
2002	5.000	0.292	13.000	0.758
2003	11.000	0.641	24.000	1.399
2004	3.000	0.175	27.000	1.574
2005	9.000	0.525	36.000	2.099
2006	14.000	0.816	50.000	2.915
2007	19.000	1.108	69.000	4.023
2008	25.000	1.458	94.000	5.481
2009	20.000	1.166	114.000	6.647
2010	24.000	1.399	138.000	8.047
2011	31.000	1.808	169.000	9.854
2012	30.000	1.749	199.000	11.603
2013	37.000	2.157	236.000	13.761
2014	51.000	2.974	287.000	16.735
2015	50.000	2.915	337.000	19.650
2016	72.000	4.198	409.000	23.848
2017	110.000	6.414	519.000	30.262
2018	182.000	10.612	701.000	40.875
2019	271.000	15.802	972.000	56.676
2020	281.000	16.385	1253.000	73.061
2021	281.000	16.385	1534.000	89.446
2022	175.000	10.204	1709.000	99.650
2023	6.000	0.350	1715.000	100.000

Panel C: Sample distribution by industry

Industry	No. of ESG reports	(%)	No. of firms	(%)
Chemicals	403.000	5.327	61.000	3.557
Computers	900.000	11.897	234.000	13.644
Durable manufacturers	1402.000	18.533	304.000	17.726
Extractive industries	318.000	4.204	62.000	3.615
Financial institutions	879.000	11.619	221.000	12.886
Food	270.000	3.569	43.000	2.507
Insurance and real estate	524.000	6.927	134.000	7.813
Mining and construction	296.000	3.913	55.000	3.207
Pharmaceuticals	261.000	3.450	78.000	4.548
Retail	554.000	7.323	136.000	7.930
Services	505.000	6.675	122.000	7.114
Textiles, printing and publishing	233.000	3.080	55.000	3.207
Transportation	560.000	7.403	119.000	6.939
Utilities	440.000	5.816	87.000	5.073
Others	20.000	0.264	4.000	0.233
Total	7,565	100	1,715	100

Note: Panel A reports the distribution of all ESG reports by year. Panel B shows the distribution of ESG reports that are the first report of their respective companies by year. Panel C presents the sample distribution by industry. Industry classifications follow Barth et al. (1998).

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

	Mean	Median	St. Dev.	Q1	Q3
Dependent Variables					
(1) ADisperse$_{t1}$	0.013	0.003	0.030	0.001	0.010
(2) SIZE (in mil.)#	31620.732	9652.527	83070.649	3347.976	27773.592
(3) ROA%	4.236	3.983	10.152	1.063	8.295
(4) LEV%	67.369	65.854	24.874	51.679	82.056
(5) VOLAT%	3.482	1.803	6.323	0.713	3.966
(6) FIN%	-0.376	-0.592	10.265	−4.395	1.753
(7) LIQUID	2.054	1.655	2.045	1.054	2.532
(8) RD%	3.610	0.000	7.859	0.000	2.783
(9) NoA#	15.687	14.000	9.890	8.000	22.000
(10) MOMENTUM	0.011	0.011	0.401	−0.256	0.276
(11) EPSdecline	0.355	0.000	0.479	0.000	1.000
(12) SURPRISE	0.000	0.000	0.020	−0.000	0.002
(13) BUSsegment	2.621	2.000	1.820	1.000	4.000
(14) GEOsegment	3.329	2.000	2.897	1.000	4.000
(15) INSTInvestors%	21.137	15.957	19.167	7.255	30.954
(16) 1stESGReport	0.227	0.000	0.419	0.000	0.000
(17) 5DaysAfterEPS$_d$	0.073	0.000	0.261	0.000	0.000
(18) DaysAfterEPS	45.511	45.000	36.627	22.000	66.000
(19) WC(in ’000)#	17.177	13.060	15.751	7.375	22.122
(20) FOG	20.491	20.495	1.894	19.449	21.464
(21) TONE%	0.884	0.890	0.635	0.505	1.272
(22) UNCERTAIN%	1.062	0.569	2.257	0.346	0.959
(23) SPECIFIC%	1.857	1.762	0.559	1.514	2.081
(24) FWDLOOK%	7.957	7.556	3.108	5.921	9.524
(25) ESGemph%	6.542	6.597	1.170	5.831	7.284
(26) EmissionScore$_{MSCI}$	0.599	0.638	0.276	0.398	0.838
(27) ESGscore$_{MSCI}$	5.091	5.000	1.993	3.775	6.542
(28) ESGscore$_{Refinitiv}$	0.597	0.610	0.159	0.483	0.715
(29) ESGscore$_{Bloomberg}$	4.526	4.543	1.106	3.710	5.313
(30) RatingsDisagreement	14.133	13.555	6.966	8.858	18.425

Note: This table presents the summary statistics (mean, median, standard deviation, 1st quartile and 3rd quartile) of the dependent variable, as well as for the model covariates. # denotes that we use the natural logarithmic values in the regression models; % denotes that we percentage figures; d indicates that the variable is a dummy variable.

Table 3.: **Correlation matrix**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1) ADisperse*t1*																	
(2) SIZE	-0.261***																
(3) ROA%	-0.216***	0.279***															
(4) LEV%	0.055***	0.028**	-0.117***														
(5) VOLAT%	0.164***	-0.155***	-0.293***	-0.077***													
(6) FIN%	0.061***	-0.099***	-0.337***	-0.024**	0.024**												
(7) LIQUID	0.169***	-0.209***	-0.169***	0.040***	0.274***	0.050***											
(8) RD%	-0.013	0.103***	-0.178***	-0.207***	0.155***	0.120***	0.074***										
(9) NoA#	-0.144***	0.559***	0.074***	0.049***	0.004	-0.096***	0.181***	0.155***									
(10) MOMENTUM	-0.031**	-0.059***	0.028**	0.042***	0.023*	-0.033***	0.030**	-0.057***	-0.066***								
(11) EPSdecline	0.092***	-0.075***	-0.248***	-0.039***	0.091***	0.113***	0.129***	0.109***	0.026**	-0.114***							
(12) SURPRISE	-0.035***	0.039***	0.140***	-0.040***	-0.084***	-0.030***	-0.002	0.009	0.021*	0.058***	-0.084***						
(13) BUSsegment#	-0.048***	0.113***	0.046***	0.039***	-0.074***	-0.025**	-0.114***	-0.103***	0.016	0.016	-0.003	0.022*					
(14) GEOsegment#	-0.046***	0.148***	0.069***	-0.093***	0.028**	-0.066***	0.009	0.213***	0.107***	-0.022*	0.081***	-0.021*	0.111***				
(15) INSTInvestors%	0.122***	-0.719***	-0.212***	-0.036***	0.093***	0.091***	0.099***	-0.037***	-0.441***	0.043***	0.051***	-0.044***	-0.119***	-0.113***			
(16) 1stESGReport	0.034***	-0.205***	-0.062***	-0.058***	0.052***	0.114***	0.024**	0.074***	-0.132***	-0.016	-0.004	-0.021*	-0.064***	-0.083***	0.196***		
(17) 5DaysAfterEPS.	0.001	-0.002	0.01	-0.015	0.016	0.012	0.003	0.026**	0.001	0.003	-0.012	-0.005	-0.053***	-0.014	0.021*	0.013	
(18) DaysAfterEPS	-0.012	-0.021*	-0.012	0.021*	-0.003	-0.012	-0.005	-0.013	-0.022*	0	0.001	-0.006	0.001	-0.014	0.012	0.005	-0.336***
(19) WC#	0.041***	0.323***	0.056***	0.034***	-0.036***	-0.031***	-0.070***	-0.048***	0.134***	0.006	0.019	-0.009	0.075***	0.156***	-0.317***	-0.281***	-0.017
(20) FOG	0.101***	-0.104***	-0.067***	-0.035***	0.048***	0.055***	-0.061***	0.045***	-0.189***	0.030**	-0.035***	-0.012	0.002	-0.044***	0.065***	0.022*	-0.005
(21) TONE%	-0.118***	0.039***	0.051***	0.044***	-0.090***	-0.034***	-0.020*	0.002	0.082***	-0.043***	-0.047***	0.015	0.042***	-0.027**	0.018	0.064***	-0.002
(22) UNCERTAIN%	0.019	-0.086***	-0.032***	0.021*	0.013	-0.003	0.008	-0.047***	-0.061***	0.01	-0.012	0.013	-0.002	-0.048***	0.063***	0.081***	0.008
(23) SPECIFIC%	-0.049***	0.013	0.032***	0.081***	-0.051***	-0.030**	0.005	-0.093***	0.060***	-0.008	-0.026**	0.009	0.017	-0.093***	-0.004	0.005	-0.004
(24) FWDLOOK%	0.024**	-0.192***	-0.077***	-0.077***	0.070***	0.091***	0.092***	0.075***	-0.121***	0.016	0.004	-0.018	-0.073***	-0.050***	0.188***	0.155***	-0.009
(25) ESGemph%	-0.041***	-0.032***	0.030**	-0.080***	0.055***	-0.054***	0.029**	0.063***	0.025**	0.034***	0.006	0	-0.006	0.155***	0.017	-0.058***	0.005

(26) EmissionScore*MSCI*	-0.062***	0.457***	0.138***	0.078***	-0.075***	-0.120***	-0.066***	-0.042***	0.376***	-0.01	0.005	0.011	0.060***	0.111***	-0.419***	-0.363***	-0.002
(27) ESGscore*MSCI*	-0.104***	0.184***	0.125***	-0.019	-0.067***	-0.100***	-0.078***	0.078***	0.110***	0.005	-0.01	0.016	0.046***	0.105***	-0.163***	-0.176***	0.003
(28) ESGscore*LSEG*	-0.105***	0.466***	0.124***	0.083***	-0.081***	-0.119***	-0.080***	-0.017	0.357***	0.004	0.007	0.013	0.081***	0.178***	-0.424***	-0.368***	-0.030**
(29) ESGscore*Bloomberg*	-0.101***	0.253***	0.094***	0.012	-0.023	-0.082***	0	0.029*	0.136***	0.040**	-0.014	0.007	-0.019	0.077***	-0.240***	-0.266***	-0.022
(30) RatingsDisagreement	-0.02	0.159***	0.016	0.045***	-0.030*	-0.031*	-0.042**	0.031*	0.119***	-0.036*	0.036**	0.004	0.093***	0.115***	-0.148***	-0.079***	-0.019

	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)
(19) WC#	-0.002											
(20) FOG	-0.003	0.061***										
(21) TONE%	0.011	-0.264***	-0.066***									
(22) UNCERTAIN%	-0.009	-0.296***	-0.001	0.080***								
(23) SPECIFIC%	0.015	-0.479***	-0.198***	0.101***	0.148***							
(24) FWDLOOK%	0.001	-0.114***	0.276***	0.082***	0.055***	-0.204***						
(25) ESGemph%	0.002	-0.085***	0.280***	0.009	0.058***	-0.043***	0.086***					
(26) EmissionScore*MSCI*	-0.017	0.424***	-0.071***	-0.073***	-0.093***	0.001	-0.195***	0.156***				
(27) ESGscore*MSCI*	-0.022*	0.178***	-0.02	-0.003	-0.063***	-0.044***	-0.050***	0.107***	0.264***			
(28) ESGscore*LSEG*	-0.026*	0.434***	-0.067***	-0.049***	-0.104***	-0.047***	-0.162***	0.173***	0.719***	0.311***		
(29) ESGscore*Bloomberg*	0.013	0.316***	0.075***	-0.105***	-0.038**	-0.044***	-0.063***	0.235***	0.404***	0.268***	0.416***	
(30) RatingsDisagreement	-0.028	0.117***	-0.111***	0.027	-0.063***	-0.044**	-0.073***	-0.017	0.220***	-0.069***	0.437***	-0.191***

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: **Summary statistics of ESG topics**

ESG Topic	**ESG sub-group**	**Description**	**Mean**	**Median**	**St. Dev.**	**Q1**	**Q3**
E (Envrionmental)			183.118	124	191.38	59	241
(1) ClimateStrategy	ESRS E1	Covers climate strategy and target-setting	29.841	18	38.763	6	39
(2) EmissionsReduction	ESRS E1	Covers greenhouse gas emissions reduction efforts	16.626	5	32.252	1	18
(3) Emissions	ESRS E1	Involves greenhouse gas emissions reporting	18.361	11	24.509	3	24
(4) EnvironmentalImpact	ESRS E2	Discusses limiting environmental impacts and sustainable resource use	16.136	2	47.339	0	9
(5) Recycling	ESRS E2	Describes managing materials and waste	16.872	6	30.822	1	19
(6) WaterManagement	ESRS E3	Reports on water usage and discharge	14.265	4	26.075	0	16
(7) BioDiversity	ESRS E4	Addresses environmental risks like climate change and biodiversity loss	12.378	3	37.985	1	11
(8) EnergyEfficiency	ESRS E5	Relates to energy efficiency and operational technology	22.812	8	48.192	2	21
(9) ProductSustainability	ESRS E5	Highlights product sustainability and packaging impact	14.28	0	52.074	0	3
(10) EnvironmentalPerformance	ESRS E5	Describes environmental performance targets	21.546	13	25.823	5	28
S (Social)			266.52	202	233.558	104	360
(11) EmployeeSafety	ESRS S1	Covers internal employee initiatives like safety or recognition programs	14.5	8	18.594	3	19
(12) EmployeeWellbeing	ESRS S1	Emphasizes employee health and well-being	17.149	4	47.265	1	12
(13) EmployeeTraining	ESRS S1	Centers on employee development and retention	32.685	23	34.85	8	46
(14) WorkerCondition	ESRS S1	Describes measures affecting the workforce	22.246	10	39.293	2	26
(15) Workforce	ESRS S1	Provides workforce statistics and engagement	14.58	8	20.128	3	19
(16) HealthSafety	ESRS S1	Covers occupational health and safety measures	24.438	11	37.954	2	30
(17) SupplyChainManagement	ESRS S2	Addresses responsible supply chain management	22.458	7	38.81	1	26
(18) Communities	ESRS S3	Involves engagement with external stakeholders and communities	16.151	8	23.524	2	21
(19) Education	ESRS S3	Highlights community educational programs and youth development	19.273	9	29.534	2	24
(20) Volunteering	ESRS S3	Describes community involvement and philanthropy	22.73	15	29.175	6	31
(21) SocialImpact	ESRS S3	Discusses striving for positive social impact	23.44	11	33.305	4	29
(22) DataProtection	ESRS S4	Focuses on customer data protection and cybersecurity	16.038	5	29.719	1	19
(23) CustomerProduct	ESRS S4	Relates to sustainable product innovation	20.833	6	41.897	1	22
G (Governance)			106.327	85	98.161	47	134
(24) Transparency	ESRS G1	Contains references to sustainability reporting in annual reports	15.722	10	19.42	3	21
(25) CorporateGovernance	ESRS G1	Reflects corporate governance practices	21.706	14	30.461	5	29

(26) StakeholderManagement	ESRS G1	Expresses a broad commitment to various stakeholders	31.685	27	24.308	14	44
(27) CodeOfConduct	ESRS G2	Focuses on corporate ethics and conduct	22.699	16	24.658	6	32
(28) SustainableFinancing	ESRS G2	Details integrating sustainability into financial services	14.515	1	50.242	0	5
(29) OtherInformation_1		Highlights financial performance metrics	11.98	1	89.503	0	4
(30) OtherInformation_2		Includes references to sourcing information	9.236	3	17.598	0	10
TotalTopics			577.181	449	477.323	256	753
TopicDiversity		Number of topics in a document	24.257	26	4.727	22	28
E *specificity*		Average number of named entities in E sentences	0.957	0.902	0.59	0.72	1.11
S *specificity*		Average number of named entities in S sentences	0.844	0.786	0.451	0.635	0.975
G *specificity*		Average number of named entities in G sentences	0.617	0.577	0.321	0.415	0.77

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, the diversity in topics and sentence-based specificity per ESG category. Note that the topic variables are scaled by 100 sentences in the regression analyses.

Table 5.: **The impact of ESG reports' thematic content (E, S and G) on information asymmetry**

	ADisperse$_{t1}$							
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Topic variables								
E		-0.048***	-0.045***	-0.038***	-0.031**	-0.051***	-0.048***	-0.048***
		(0.012)	(0.015)	(0.013)	(0.013)	(0.015)	(0.016)	(0.018)
S		0.046***	0.055***	0.041***	0.064***	0.069***	0.071***	0.072***
		(0.011)	(0.013)	(0.012)	(0.012)	(0.014)	(0.016)	(0.017)
G		-0.009	-0.016	0.01	-0.061**	-0.053*	-0.05	-0.049
		(0.025)	(0.031)	(0.026)	(0.028)	(0.031)	(0.035)	(0.037)
ESG Ratings, CO_2 emissions and ESG disagreement								
EmissionsScore$_{MSCI}$			-0.060***					-0.059***
			(0.011)					(0.014)
ESGscore$_{MSCI}$				-0.093***				-0.085***
				(0.009)				(0.014)
ESGscore$_{LSEG}$					0.132			0.325
					(0.150)			(0.259)
ESGscore$_{Bloomberg}$						-0.056**		-0.017
						(0.022)		(0.029)
RatingsDisagreement							0.008**	0.005
							(0.003)	(0.004)
Control Variables								
SIZE#	-0.243***	-0.245***	-0.210***	-0.184***	-0.265***	-0.335***	-0.261***	-0.265***
	(0.019)	(0.019)	(0.026)	(0.022)	(0.024)	(0.028)	(0.033)	(0.035)
ROA%	-0.020***	-0.021***	-0.022***	-0.022***	-0.016***	-0.018***	-0.019***	-0.018***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
LEV%	0.003***	0.003***	0.002*	0.002***	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
VOLAT%	0.028***	0.027***	0.022***	0.024***	0.025***	0.017***	0.013***	0.008**
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
FIN%	-0.002	-0.001	-0.001	-0.001	0	0	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
LIQUID	0.099***	0.101***	0.109***	0.117***	0.127***	0.131***	0.153***	0.128***
	(0.009)	(0.009)	(0.012)	(0.011)	(0.011)	(0.011)	(0.013)	(0.014)
RD%	0.011***	0.010***	0.013***	0.014***	0.006**	0.005	0.008*	0.011***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
NoA#	-0.069**	-0.074**	-0.079**	-0.139***	0.359***	0.328***	0.319***	0.369***
	(0.031)	(0.031)	(0.039)	(0.033)	(0.043)	(0.052)	(0.058)	(0.061)
EPSdecline	0.152***	0.147***	0.135***	0.144***	0.133***	0.137***	0.120**	0.107*
	(0.038)	(0.038)	(0.046)	(0.039)	(0.041)	(0.048)	(0.052)	(0.055)
SURPRISE	1.481*	1.456*	1.539	1.044	0.089	0.516	-0.289	0.267
	(0.861)	(0.860)	(1.134)	(0.986)	(1.196)	(1.285)	(1.579)	(1.595)
BUSsegment	0.01	0.007	0.042	0.028	0.024	0.014	0.012	0.015
	(0.024)	(0.024)	(0.030)	(0.025)	(0.027)	(0.031)	(0.034)	(0.036)
GEOsegment	0.127***	0.132***	0.077**	0.125***	0.048*	0.054	0.026	0.043

	(0.026)	(0.026)	(0.032)	(0.028)	(0.029)	(0.035)	(0.039)	(0.041)
INSTInvestors%	-0.002	-0.002*	-0.006***	-0.005***	0.006***	-0.001	0.003	0.002
	(0.130)	(0.130)	(0.206)	(0.164)	(0.181)	(0.237)	(0.297)	(0.309)
1stESGReport	0.012	0.007	-0.052	-0.053	-0.026	-0.028	-0.038	-0.114
	(0.043)	(0.043)	(0.053)	(0.046)	(0.048)	(0.059)	(0.064)	(0.070)
5DaysAfterEPS$_d$	0.193***	0.190***	0.092	0.140*	0.125*	0.064	0.005	0.01
	(0.067)	(0.067)	(0.082)	(0.072)	(0.073)	(0.089)	(0.097)	(0.100)
DaysAfterEPS	0	0	-0.001	-0.001	0	-0.001	-0.002*	-0.002*
	0.000	0.000	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
WC#	0.293***	0.281***	0.300***	0.302***	0.147***	0.124***	0.111***	0.166***
	(0.027)	(0.029)	(0.035)	(0.030)	(0.033)	(0.039)	(0.041)	(0.047)
FOG	0.060***	0.059***	0.058***	0.063***	0.020*	0.022	0.027*	0.026
	(0.010)	(0.010)	(0.012)	(0.011)	(0.011)	(0.014)	(0.015)	(0.016)
TONE%	-0.160***	-0.168***	-0.101***	-0.106***	-0.04	-0.116***	-0.069*	-0.041
	(0.028)	(0.028)	(0.035)	(0.030)	(0.031)	(0.037)	(0.041)	(0.044)
UNCERTAIN%	0.034***	0.035***	0.028***	0.025***	0.025***	0.024***	0.014	0.018*
	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)	(0.010)
SPECIFIC%	0.079**	0.081**	0.082*	0.088**	0.037	0.003	0.032	0.071
	(0.037)	(0.037)	(0.043)	(0.039)	(0.039)	(0.047)	(0.048)	(0.052)
FWDLOOK%	-0.017***	-0.016***	-0.027***	-0.022***	-0.008	-0.013	-0.018**	-0.022**
	(0.006)	(0.006)	(0.007)	(0.006)	(0.007)	(0.008)	(0.009)	(0.009)
ESGemph%	-0.073***	-0.071***	-0.062***	-0.059***	0.031*	0.016	0.015	0.025
	(0.016)	(0.016)	(0.020)	(0.017)	(0.018)	(0.021)	(0.023)	(0.025)
(Intercept)	-8.839***	-8.637***	-6.253***	-7.370***	-7.012***	-4.823***	-5.847***	-5.807***
	(1.070)	(1.071)	(0.574)	(1.428)	(0.705)	(0.581)	(0.646)	(0.697)
Fixed industry and year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. obs.	7,565	7,565	4,807	6,170	5,177	4,046	3,227	2,902
Adj. R^2	0.253	0.255	0.234	0.261	0.291	0.291	0.305	0.305

Panel B: Model Comparison

		RSS	**RSS**	**RSS**	**RSS**	**RSS**	**RSS**	**RSS**
Model with Topic Variables		13,722.53	8,867.87	11,152.02	8,317.21	6,910.14	5,125.91	4,605.69
Model without Topic Variables		13,769.11	8,904.61	11,185.08	8,360.21	6,957.16	5,162.24	4,638.11
Difference		-46.584***	-36.741***	-33.058***	-42.994***	-47.016***	-36.333***	-32.425***

Note: The table presents results concerning Hypothesis 1 and 2, which examines the relationship between E, S, and G aggregate topic variables and analysts dispersion. Panel A presents the estimated coefficients for the model variables, including E, S, and G aggregate topic variables. As a goodness-of-fit measure, adjusted R2 are provided. Panel B reports the results from ANOVA tests comparing the sum of squared residuals (RSS) of models with topic category variables (Eq. 4) and without (Eq. 3). *, ** and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-sided t-tests.

Table 6.: **ESG Report Content and Analyst Dispersion: Topic Groups and Period Effects**

PART A: Topic Groups

	ADisperse$_{t1}$			
	Model 1	**Model 2**	**Model 3**	**Model 4**
Panel A: Regression results				
Environmental Topic Groups				
E1: ClimateChange	0.071**			0.044
	-0.03			-0.034
E2: Pollution	-0.022			-0.03
	-0.036			-0.037
E3: WaterResources	0.077			0.133
	-0.084			-0.086
E4: Biodiversity	-0.190***			-0.204***
	-0.072			-0.073
E5: ResourceUse	-0.108***			-0.128***
	-0.024			-0.026
Social Topic Groups				
S1: Workforce		-0.037*		-0.014
		-0.021		-0.023
S2: Valuechain		0.099*		0.122**
		-0.055		-0.056
S3: Communities		0.028		0.039
		-0.03		-0.032
S4: Customer		0.125***		0.140***
		-0.033		-0.034
Governance Topic Groups				
G1: CorporateGovernance			-0.067	-0.117**
			-0.042	-0.05
G2: BusinessPractices			0.107***	0.058
			-0.036	-0.039
Controls	Yes	Yes	Yes	Yes
Fixed industry and year effects	Yes	Yes	Yes	Yes
Num. obs.	6,170	6,170	6,170	6,170
Adj. R^2	0.262	0.263	0.26	0.267
Panel B: Model Comparison				
	RSS	**RSS**	**RSS**	**RSS**
Model with Topic Variables	11,142.07	11,133.04	11,168.58	11,051.57
Model without Topic Variables	11,185.08	11,185.08	11,185.08	11,185.08
Difference	-43.01***	-52.039***	-16.496**	-133.514***

PART B: Period Comparison

	Before 2011	**2011-2014**	**2015-2019**	**2020 and After**
	Early adoption	**SASB founded**	**SDG \ Paris Agreement**	**COVID-19 pandemic**
	ADisperset1	**ADisperset1**	**ADisperset1**	**ADisperset1**
Panel A: Aggregate-level E, S and G				
E	-0.009	-0.035	-0.046**	-0.018*
	-0.037	-0.028	-0.02	-0.018
S	0.059*	0.006	0.034*	0.037**

	-0.031	-0.028	-0.018	-0.016
G	-0.101	0.071	0.059	-0.022
	-0.091	-0.056	-0.036	-0.037
Controls	Yes	Yes	Yes	Yes
Fixed industry and year effects	Yes	Yes	Yes	Yes
Num. obs.	471	723	2,128	2,848
Adj. R^2	0.409	0.389	0.281	0.23
Panel B: Model Comparison				
	RSS	**RSS**	**RSS**	**RSS**
Model with topic variables	568.654	981.773	3692.994	5340.398
Model without topic variables	573.76	985.49	3711.1	5350.993
Difference	-5.106	-3.717	-18.106**	-10.595

Note: The table presents results on the relationship between ESG disclosures and analyst forecast dispersion. Part A reports estimates for ESG topic groups, and Part B presents results by period. Within each part, Panel A provides regression coefficients for ESG categories, with adjusted R2 as a measure of model fit, and Panel B compares models with and without ESG topic variables (Eq. 4 vs. Eq. 3) using ANOVA tests based on the residual sum of squares (RSS). $*$, $**$ and $***$ denote statistical significance at the 10, 5, and 1 percent levels, respectively, based on two-sided t-tests.

Table 7: **Impact of diversity in thematic content of ESG reports, and impact of issuing the 1st ESG report**

	Part A Topic Diversity		**Part B Nth Report**	
	Linear Model Model 1	**Quadratic Model Model 2**	**First Report Model 3**	**Subsequent Report Model 4**
	$ADisperse_{t1}$	**$ADisperse_{t1}$**	**$ADisperse_{t1}$**	**$ADisperse_{t1}$**
Panel A: Regression results				
TopicDiversity	0.014***	-0.042*		
	(0.004)	(0.023)		
Topic $Diversity^2$		0.001**		
		(0.001)		
E			-0.035*	-0.038***
			(0.018)	(0.014)
S			0.060*	0.037***
			(0.031)	(0.012)
G			0.111	-0.004
			(0.074)	(0.028)
Control variables				
SIZE#	-0.182***	-0.184***	-0.245***	-0.171***
	(0.022)	(0.022)	(0.055)	(0.024)
ROA%	-0.022***	-0.022***	-0.016***	-0.025***
	(0.002)	(0.002)	(0.005)	(0.003)
LEV%	0.002***	0.002***	0.002	0.002**
	(0.001)	(0.001)	(0.002)	(0.001)
VOLAT%	0.024***	0.024***	0.037***	0.016***
	(0.003)	(0.003)	(0.007)	(0.004)
FIN%	-0.001	-0.001	0.003	-0.002
	(0.002)	(0.002)	(0.004)	(0.002)
LIQUID	0.115***	0.114***	0.084***	0.128***
	(0.011)	(0.011)	(0.026)	(0.012)
RD%	0.016***	0.016***	0.015**	0.017***
	(0.003)	(0.003)	(0.006)	(0.003)
NoA#	-0.136***	-0.134***	0.013	-0.190***
	(0.033)	(0.033)	(0.079)	(0.037)
EPSdecline	0.147***	0.145***	0.131	0.148***
	(0.039)	(0.039)	(0.099)	(0.044)
SURPRISE	0.976	0.966	-0.84	1.877*
	(0.986)	(0.986)	(2.327)	(1.131)
BUSsegment	0.032	0.032	0	0.028
	(0.025)	(0.025)	(0.064)	(0.028)
GEOsegment	0.121***	0.121***	0.113	0.133***
	(0.028)	(0.028)	(0.070)	(0.031)
INSTInvestors%	-0.446***	-0.453***	-0.001	-0.007***
	(0.164)	(0.164)	(0.003)	(0.002)
1stESGReport	-0.048	-0.043		
	(0.046)	(0.046)		
$5DaysAfterEPS_d$	0.136*	0.140*	0.279	0.075
	(0.072)	(0.072)	(0.177)	(0.081)
DaysAfterEPS	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.002)	(0.001)
WC#	0.285***	0.288***	0.227***	0.299***
	(0.031)	(0.031)	(0.071)	(0.035)
FOG	0.064***	0.061***	0.026	0.081***
	(0.011)	(0.011)	(0.026)	(0.013)
TONE%	-0.101***	-0.093***	-0.082	-0.118***
	(0.030)	(0.030)	(0.068)	(0.035)
UNCERTAIN%	0.021***	0.021***	0.017	0.030**
	(0.008)	(0.008)	(0.012)	(0.012)
SPECIFIC%	0.091**	0.083**	0.077	0.080*
	(0.039)	(0.039)	(0.075)	(0.046)
FWDLOOK%	-0.023***	-0.022***	-0.002	-0.028***
	(0.006)	(0.006)	(0.013)	(0.008)
ESGemph%	-0.065***	-0.063***	-0.027	-0.080***
	(0.017)	(0.017)	(0.040)	(0.019)
$ESGscore_{MSCI}$	-0.093***	-0.094***	-0.085***	-0.097***

	(0.009)	(0.009)	(0.024)	(0.010)
(Intercept)	-7.442***	-6.913***	-6.243***	-6.339***
	(1.426)	(1.442)	(1.758)	(0.929)
Fixed industry and year effects	Yes	Yes	Yes	Yes
Num. obs.	6,170	6,170	1,114	5,056
Adj. R^2	0.261	0.262	0.215	0.279
Panel B: Model Comparison				
	RSS	**RSS**	**RSS**	**RSS**
Model with Topic Variables	11163.631	11152.337	2001.293	8466.224
Model without Topic Variables	11185.079	11185.079	2025.786	8487.863
Difference	-21.449***	-32.743***	-24.493***	-21.639***

Note: The table presents results concerning Hypothesis 3 and 4. In Part A, Model 1 and 2 present the relationship between ESG topic diversity and analysts dispersion. In Part B, Model 3 and 4 present the results for sub-sample analysis of relationships between E, S, and G topic variables and analyst dispersion between firs5t 7firm ESG reports and subsequent reports. Panel A presents the estimated coefficients for the independent variables–ESG topic diversity. As a goodness-of-fit measure, adjusted R2 are provided. Panel B reports the results from ANOVA tests comparing the sum of squared residuals (RSS) of models with topic-related variables (Eq. 4/5) and without (Eq. 3). ∗, ∗∗ and ∗∗∗ denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-sided t-tests.

Table 8: **Cross-sectional Analyses: Sub-sample Analyses on the Impact of E, S, and G Topic Variables**

PART A: ESG Performance and Disagreement

	$ESGscore_{MSCI}$		**Ratings Disagreement**	
	<median	**median≤**	**<median**	**median≤**
	$ADisperse_{t1}$	**$ADisperse_{t1}$**	**$ADisperse_{t1}$**	**$ADisperse_{t1}$**
Panel A: Aggregate-level E, S and G				
E	-0.011	-0.057***	-0.039*	-0.051**
	(0.018)	(0.016)	(0.020)	(0.022)
S	0.031**	0.041***	0.077***	0.043**
	(0.016)	(0.014)	(0.022)	(0.019)
G	0.043	0	-0.088*	0.015
	(0.031)	(0.034)	(0.049)	(0.043)
Controls	Yes	Yes	Yes	Yes
Num. obs.	3,047.000	3,123.000	1,602.000	1,625.000
Adj. R^2	0.264	0.234	0.342	0.306
Panel B: Model Comparison				
	RSS	**RSS**	**RSS**	**RSS [3pt]**
Model with topic variables	5653.373	5306.394	2493.857	2410.241
Model without Topic Variables	5,673.389	5,333.018	2,514.110	2,422.928
Difference	-20.016**	-26.624***	-20.253***	-12.687**

PART B: Informational Environment

	NoA		**INSTInvestors%**	
	<median	**median≤**	**<median**	**median≤**
	$ADisperse_{t1}$	**$ADisperse_{t1}$**	**$ADisperse_{t1}$**	**$ADisperse_{t1}$**
Panel A: Aggregate-level E, S and G				
E	-0.023	-0.032**	-0.050***	-0.012
	(0.020)	(0.014)	(0.015)	(0.018)
S	0.022	0.042***	0.028**	0.035**
	(0.018)	(0.012)	(0.013)	(0.017)
G	0.027	0.011	0.080***	-0.057*
	(0.035)	(0.029)	(0.030)	(0.033)
Controls	Yes	Yes	Yes	Yes
Num. obs.	2,765.000	3,405.000	3,167.000	3,003.000
Adj. R^2	0.208	0.336	0.298	0.269
Panel B: Model Comparison				
	RSS	**RSS**	**RSS**	**RSS**
Model with topic variables	5820.346	4865.099	5119.091	5406.034
Model without Topic Variables	5,827.775	4,887.255	5,152.742	5,415.735
Difference	-7.429	-22.156***	-33.651***	-9.701

PART C: Concurrent Disclosures

	$5DaysAfterEPS_d$	
	0	**1**
Panel A: Aggregate-level E, S and G	**$ADisperse_{t1}$**	**$ADisperse_{t1}$**
E	-0.040***	0.013
	(0.012)	(0.052)
S	0.039***	(0.051)
	(0.011)	(0.045)
G	0.028	0.025
	(0.024)	(0.076)
Controls	Yes	Yes

Num. obs.	5719	451
Adj. R^2	0.264	0.265

Panel B: Model Comparison

	RSS	**RSS**
Model with topic variables	10339.805	680.035
Model without Topic Variables	10378.865	682.489
Difference	-39.060***	-2.454

Note: The table presents results on the relationships between E, S, and G aggregate topics and analysts dispersion across different sub-sample specifications. In Part A, firms are grouped based on ESGscoreMSCI and RatingsDisagreement. Part B considers sub-samples based on NoA and INST Investors, while Part C splits the sample based on 5DaysAfterEPS. Panel A presents the estimated coefficients for the E, S, and G aggregate topics, along with adjusted R2 values as a measure of model fit. Panel B reports the results from ANOVA tests comparing the sum of squared residuals (RSS) of models with topic category variables (Eq. 4) and without (Eq. 3). $*$, $**$ and $***$ denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-sided t-tests.

Table 9: **Additional Analyses**

	$ESGscore_{MSCI}$	Ratings Disagreement	$ADisperse_{t1}$
	Model 1	**Model 2**	**Model 3**
Panel A: Regression results			
E	0.054***	-0.267***	
	-0.016	-0.076	
S	-0.001	0.025	
	-0.014	-0.074	
G	-0.022	0.243	
	-0.031	-0.167	
E *specificity*			-0.037
			-0.03
S *specificity*			-0.080*
			-0.043
G *specificity*			0.071
			-0.057
Controls	Yes	Yes	Yes
Fixed industry and year effects	Yes	Yes	Yes
Num. obs.	6,170	3,227	6,170
Adj. R^2	0.141	0.126	0.26

Panel B: Model Comparison			
	RSS	**RSS**	**RSS**
Model with Topic Variables	20,937.88	135,423.13	11,172.55
Model without Topic Variables	20,986.52	136,028.33	11,185.08
Difference	-48.647***	-605.198***	-12.534*

Note: The table presents regression results from additional analyses. Model (1) examines the relationship between ESG topic variables and ESGscoreMSCI; Model (2) relates ESG variables to RatingsDisagreement; and Model (3) assesses the association between ESG topic specificity variables and ADisperset1. Panel A reports the estimated coefficients for the independent variables: aggregate E, S, and G topics, and the specificity measures. Adjusted R2 values are provided as measures of model fit. Panel B reports ANOVA tests comparing the residual sum of squares (RSS) from models with ESG topic variables (Eq. 4) to those without (Eq. 3). ∗, ∗∗ and ∗∗∗ denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-sided t-tests.

10. Appendix

10.1. Validation of the sentLDA output

This Appendix details the procedures undertaken to assign and validate the topic labels. Topic labeling is a central component in the application of automated topic modeling procedures. Specifically, in sentLDA, the unsupervised machine-learning tool that the study uses, the topic outputs are obtained in the form of weighted lists of words. To translate these topics and determine their meaningful interpretations, we must assign these topic outputs with semantically meaningful labels. As such, the accuracy of these labels is central to the study analysis. Therefore, in this Appendix, we detail the rigorous steps we undertake to ensure the reliability of the assigned topic labels.

We undertake two procedures to assign and evaluate the topic labels. As the first procedure, we 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 sentences of typical length. We then sort these sentences based on the cosine similarity between them, emphasizing the sentences that are more similar to the other sentences. We also extract the 20 most frequent bi-grams (two-word phrases excluding stop words, numbers, and symbols) from the 334 mid-length sentences. We then evaluate the semantic meaning of the top 100 mid-length sentences and the top 20 bi-grams, and assign descriptive labels to each topic.

To illustrate and as shown in Figure 4, the topic with the highest-weighted words (*emission*, *energy*, *carbon*, and *gas*) is labelled as *EmissionsReduction*. From evaluating the assigned topic sentences, we find that (*renewable - energy*, *greenhouse - gas*, *gas - emission*, and *emission - reduction*) are the most frequent bi-grams. Furthermore, a sample of the highest weighted mid-length sentence reads as follows: "We are committed to reduction of CO_2 emissions by investing in renewable energy assets." As such, these further evaluations provide support for the assigned label. Figure 4 provides the word clouds representing the list and weights of the highest weighted words in each topic with their associated labels, while Table 4 provides the labels of each topic and briefly describes the 30 topics.[1515] Tables S1, S2, and S3 in the Supplementary Material (XLS file) provide the lists of 20 highest weighted words, the 20 most common bi-grams, 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, Lang, and Stice-Lawrence (2017), and Brown and Hillegeist (2007).

< Insert Figure 4 about here >

As a second evaluation, we graphically examine the similarities between the topics based on the words they emphasize. Figure 2a, 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 45% are not displayed). We observe several clusters. For instance, topics *Volunteering*, and *Education* show notable degree of correlations, meaning that the sentences discussing firms' volunteering activities often emphasize words that relate to education. Likewise, we find clusters between topics concerning the environment, such as *Recycling* and *EnergyEfficiency*, and also between topics related to various corporate governance issues, such as *CodeOfConduct*, and *CorporateGovernance*. Since our topic labels are discretionary, these linkages provide additional nuance and support for the interpretation of the topics.

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

---

[15] 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 *employee* is prominently related to the topics such as *EmployeeTraining*, *WorkerCondition*, and *Workforce*, which highlights the contextual nature of the word.

human judges. These labeling steps are then followed by descriptive analysis of the sample topics, which are detailed in Section 4.2 of the manuscript.

Figure 4.: **Wordclouds of ESG Disclosure Topics**

(1) ClimateStrategy

(2) EmissionsReduction

(3) Emissions

(4) EnvironmentalImpact

(5) Recycling

(6) WaterManagement

(7) BiodiversityClimate

(8) EnergyEfficiency

(9) ProductSustainability

(10)
EnvironmentalPerformance

(11) EmployeeSafety

(12) EmployeeWellbeing

(13) EmployeeTraining

(14) WorkerCondition

(15) Workforce

(16) HealthSafety

(17)
SupplyChainManagement

(18) Communities

(19) Education

(20) Volunteering

(21) SocialImpact

(22) DataProtection

(23) CustomerProduct

(24) Transparency

(25) CorporateGovernance

(26) StakeholderManagement

(27) CodeOfConduct

(28) SustainableFinancing

(29) OtherInformation 1

(30) OtherInformation 2

Note: Each wordcloud visualizes the most frequent words in one of the 30 ESG topics identified using the sentLDA model. Topics are grouped by thematic category (E, S, G, Other) and are ordered by interpretability and coherence.