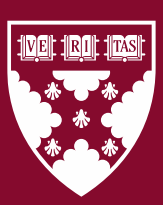
Working Paper 23-024

The Evolution of ESG Reports and the Role of Voluntary Standards

Ethan Rouen
Kunal Sachdeva
Aaron Yoon



Harvard Business School

The Evolution of ESG Reports and the Role of Voluntary Standards

Ethan Rouen
Harvard Business School

Kunal Sachdeva
Rice University

Aaron Yoon
Kellogg School of Management

Working Paper 23-024

Funding for this research was provided in part by Harvard Business School.

The Evolution of ESG Reports and the Role of Voluntary Standards[†]

Ethan Rouen[‡] **Kunal Sachdeva**[§] **Aaron Yoon**[||]

October 28, 2022

Abstract

We examine the evolution of ESG reports of S&P 500 firms from 2010 to 2021. The percentage of firms releasing these voluntary disclosures increased from 35% to 86% during this period, although the length of these documents experienced more modest growth. Using a semisupervised machine-learning approach and guided by voluntary standards that identified material ESG issues, we explore whether the content in these reports has become more relevant to investors. On average, firms devote most of their reports to topics that are material to their sector. The relative amount of material information increased by 11% after the release of voluntary standards. This increase was driven by firms that were not involved in the standard-setting process. Firms that helped develop the standards increased material disclosures at similar rates while the standards were being developed. These results provide new insights into how ESG reports evolved.

JEL classification: M14, M40

Keywords: *ESG Reports, SASB, Voluntary Disclosure, Textual Analysis, Topic Modelling*

[†]We thank Christian Leuz (editor), an anonymous referee, Richard Crowley, Jeffrey Hales, Jim Naughton, Bob Kaplan, Joseph Pacelli, Krishna Palepu, Kris Ramesh, George Serafeim, Rodrigo Verdi, Anthony Welsch, and workshop participants at the Journal of Accounting Research Registered Reports Conference, Columbia Business School, Harvard Business School, Singapore Management University, and the University of Texas-Austin for their thoughtful comments. We are indebted to Janine Guillot, Nicolai Lundy, and Arturo Rodriguez at the Value Reporting Foundation for providing access to SASB-related data and feedback. We thank RepRisk for sharing the raw incidents data. Note that Aaron Yoon serves on the academic advisory board of RepRisk. We acknowledge the excellent research assistance of James Enright, Jisang Kymm, Kunal Rai, Wesley Ramos Dos Santos, Andrew Robinson, Mei Tercek, Jo Wisch, and Irpan Yierpan.

[‡]Harvard Business School, erouen@hbs.edu (Corresponding author)

[§]Rice University, kunal.sachdeva@rice.edu

[||] Kellogg School of Management, aaron.yoon@kellogg.northwestern.edu

1 Introduction

At the start of the 21st century, almost no companies released ESG-related disclosures, but by 2021, most large publicly traded U.S. firms had converged around voluntary standalone ESG reports as a primary means of documenting their ESG activities (Serafeim, 2022).[1] This growth makes these reports among the fastest growing voluntary disclosures in history.

Despite the rapid adoption of ESG reports, there still exist almost no large-scale studies of the information disclosed within these reports.[2] One significant barrier to empirical analysis is that there exists no data source through which a large collection of ESG reports is available. Further, ESG reports are unstandardized in form and content, since they are not audited, mandated, or regulated in the United States (and most other jurisdictions), and the content of these reports continues to vary widely by firm and industry, as well as over time. To address this lack of consistency, regulators are considering reporting mandates to provide frameworks for disclosing ESG activities (IFRS, 2021; SEC, 2021). Mandates would likely improve transparency and comparability (Byard, Li, and Yu, 2011; Daske, Hail, Leuz, and Verdi, 2008; De Franco, Kothari, and Verdi, 2011; Hail, Leuz, and Wysocki, 2010; Tan, Wang, and Welker, 2011). As such, there is a need for careful empirical analyses of the potential introduction of ESG-related standards and potential disclosure mandates (Christensen, Hail, and Leuz, 2019, 2021).

This paper explores how the content of ESG reports has evolved in the absence of regulation, and how this content changed around the introduction of voluntary disclosure standards that defined a comprehensive set of financially material ESG issues. We study ESG reports at two units of analysis: the document-year level and document-topic-year level ("the topic level," going forward). At the document-year level, we report how

[1] ESG reports take various names, such as corporate sustainability reports or sustainability reports. In this paper, we refer to all reports that comprehensively discuss firms' ESG activities as "ESG reports."

[2] While we know of no studies that have examined the content of these reports on a large scale, several studies have examined the decision to disclose an ESG report and how it relates to firm characteristics (Hahn and Kühnen, 2013).

this disclosure choice evolved, providing evidence of the rapid growth in the number of firms reporting, the determinants of releasing ESG reports, and the change in the average length of the reports. We next study the content in these reports (i.e., topic-level analysis) to understand the rich heterogeneity within these documents. To do so, we employ a semisupervised model to learn the meanings of all words and phrases within ESG reports. We start our topic analysis by documenting how the language that firms use to discuss specific ESG topics has evolved with their operating environments. We then examine how much of these reports is devoted to providing financially material information, how the focus on material information changed around the introduction of voluntary standards, and how disclosure practices differed between two subsets of firms: those that participated in the development of these standards and those that were disclosing these reports prior to the release of the standards but were not involved in the standard setting process.

Our starting point is an intensive data collection exercise to collect ESG reports for all firms that were included at least once in the S&P 500 Index from 2010 to 2021. This focus on the largest U.S. firms holds much of the country and institutional settings constant while offering a robust disclosure environment. Given that there is no clearinghouse of ESG reports, we take several steps to ensure that our sample is largely complete.[3] We begin by hand collecting all available ESG reports from firms' current websites. We then do an exhaustive search of archived firm websites for older reports. If there are gaps in reporting years (e.g., we identify a report in 2010 and 2012 but not 2011), we search additional websites to locate the missing report. We compared our sample of 3,660 ESG reports to that of a commercially available set of ESG reports and found are coverage to be significantly greater.

Because these reports are complex PDF files, extracting their text into a machine-

[3]Because of the long time series and the lack of a clearinghouse for ESG reports, we cannot be sure that we collect every report. Still, our main analysis examines the content of these reports, not the choice to disclose, so it is unlikely that missing reports would bias our findings.

readable format poses an additional challenge. We employ several machine learning approaches as well as hand-verification to accurately extract the corpus of text from the PDFs. This process creates an initial document-year panel of the characteristics of ESG reports that is comparable across firms and time.

We use this panel of text to describe the properties of ESG reports and their contents in aggregate. We find that the percentage of firms releasing ESG reports increased monotonically from 35% in 2010 to 86% in 2020.[4] Despite the rapid growth in the percentage of reporting firms, growth in the length of these reports was modest and non-monotonic, with significant variation across sectors. This result starkly contrasts with the trend in regulatory filings, which have grown dramatically in length (e.g., Cohen, Malloy, and Nguyen (2020)). In addition, we find that firms with more negative ESG-related incidents and those with shorter 10-K filings are more likely to publish an ESG report, suggesting that these firms do so in response to negative attention and these disclosures serve to augment what is disclosed in regulatory filings.

Our primary research questions are whether and how disclosures in ESG reports evolved in the absence of regulation. At the document level, we find descriptive evidence that firms in the same sector increasingly use similar language over time, as do firms across sectors, meaning that firms may be coalescing around a common ESG vocabulary. Still, combined with our evidence that document length has not meaningfully changed in a decade of disclosure, this document-level analysis is insufficient to understand the rich heterogeneity *within* ESG reports. Therefore, we examine which ESG issues firms report on and whether the information disclosed is financially material to investors.

We investigate this disclosure heterogeneity in relation to financial materiality. To define materiality, we rely on sector-level guidance released by the Sustainable Accounting Standards Board (SASB). This guidance, which included detailed industry-level standards, was introduced in a staggered fashion by sector from 2013-2016 and was in-

[4]There is a slight decrease in the percentage of firms reporting in 2021 because, at the time of data collection in September 2022, some firms had yet to release their 2021 reports.

tended to be compatible with the 10-K to provide financially material ESG information to investors.[5]

Because of this focus on materiality at the sector level and the lack of a mandate for firms to include these disclosures in their 10-Ks, though, it is common for U.S. firms to discuss, in their ESG reports, how those reports conform to SASB standards. Given the thoroughness of the SASB framework, its focus on financial materiality, and the rapid adoption of its disclosure frameworks by firms in their ESG reports, the sector-level guidance is an appropriate set of guidelines for studying materiality in these reports.

Using SASB's industry-level standards, we apply to the full ESG report corpus of text ("the corpus") a semisupervised machine learning approach similar to that used by Li, Mai, Shen, and Yan (2021).[6] Specifically, we define topics based on the full set of ESG activities defined by SASB and provide a group of unique seed words for each topic from the standards to train a neural network model that learns the words and phrases used to describe each topic. This method is ideal for our setting because of the readily available set of topics and seed words provided by the SASB's guidance and provisional standards, which defined a universe of 26 ESG-related topics and identified which topics were financially material for each of the 11 SASB sectors. The algorithm allows us to quantify the semantics, as opposed to the syntax, of each document and produce our topic scores, calculated as a weighted-frequency count of all words on that topic within the ESG report, the term frequency-inverse document frequency (TF-IDF) (Li et al., 2021).

We first document that the language used within topics has evolved as the operating environment has changed. For example, when discussing customer welfare in ESG reports, terms like "vaccine" were uncommon in 2010 but were identified by the algorithm

[5]As we note in Section 2, SASB is not the only organization that has given guidance on material ESG issues. For example, the Global Reporting Initiative (GRI) has provided guidance on ESG materiality from stakeholders' perspectives. Our focus on SASB's definition of materiality allows us to focus on ESG topics that are material to investors.

[6]SASB defined standards for 77 distinct industries and included each of those industries in one of 11 sectors.

as among the most important terms in 2020, a reflection of the COVID-19 pandemic. In addition to providing unique descriptive evidence of how disclosure language changes, this analysis shows the flexibility of the machine-learning approach in measuring text as language evolves.

We next study the topic scores' relation to each other, to firm-level covariates, and to commercial ESG ratings in order to validate their relevance.[7] When examining correlations between the topic scores, we find little evidence that disclosure of any one topic strongly correlates with disclosure of another, which suggests that we are identifying unique characteristics of each topic. The data also reveal that firms disclose more on topics that relate directly to their business activities, suggesting that these reports discuss real firm behaviors.[8] Importantly, we find strong correlations between our topic-level scores and commercial ESG ratings.[9] This series of results provides confidence that we are appropriately measuring relevant disclosures related to specific ESG activities.

Our approach to quantifying ESG reports allows us to examine whether these reports focus on information that is material to investors. We find that, on average, firms disclose 48% more on material topics relative to immaterial topics. In addition, at the sector level, we frequently find that the most discussed topics are those deemed material. These results provide initial evidence that firms disclose ESG information that aligns with a set of standards defining materiality.

Turning to the evolution of the content in ESG reports, we examine whether the changes in ESG reporting relate to the public release of SASB's provisional standards. To do so, we exploit the staggered sector-level introduction of these standards and use a difference-in-differences empirical specification to uncover whether the amount of infor-

[7] We also manually inspect the words identified as relevant to each topic as a final check of the accuracy of the algorithm. Beyond a basic sanity check, this provides further insight into how firms discuss ESG issues.

[8] For example, we find a strong positive correlation between research and development expenses and the topics "data security" and "product design and lifecycle management." We also find a strong correlation between advertising expense and the topic "selling practices and product labeling."

[9] In our main analysis, our results are robust to controlling for ESG ratings, providing evidence that we are capturing information incremental to that provided by commercial ratings agencies.

mation related to material and immaterial topics changes as SASB provides voluntary disclosure guidance on ESG's financial materiality. In all of our analyses, we include a host of firm-level controls and restrictive fixed effects that greatly reduce concerns about omitted variables, such as investor preferences, the evolution of ESG reporting, concurrent changes in regulatory reporting (i.e., growth in the 10-K), time-invariant sector characteristics, and firm characteristics.

We find statistically and economically meaningful evidence that firms increased their disclosure of material ESG information after the release of SASB standards, with the proportion of material information increasing by an average of 11.0% after the standards. These findings are among the first to suggest that well-defined voluntary standards and guidance can help improve ESG disclosures and should be of interest to investors and regulators. We acknowledge that this evidence cannot be interpreted causally, since the introduction of SASB standards and ESG disclosure choices were both likely shaped by stakeholders (e.g., investors and firms) and other factors in the disclosure environment. Still, as we progressively add controls and fixed effects in our specifications, the R-squared increases, but our estimates do not change, suggesting that there is not an omitted variable problem (Oster, 2019). We also use a stacked difference-in-differences research design in all of our main tests, which has been shown to address recent concerns about biased estimates in the staggered difference-in-differences research design (Baker, Larcker, and Wang, 2022; Barrios, 2021; Cengiz, Dube, Lindner, and Zipperer, 2019).

Given the voluntary nature of both ESG disclosures and the adoption of SASB standards, understanding *how* firms converge toward material disclosures remains an outstanding but important question. Prior research provides insights into how firms' voluntary financial disclosures converge, but ESG reports are vastly different from financial disclosures, given that they cover numerous subjects and appeal to several stakeholders (Bourveau, Breuer, and Stoumbos, 2020; Einhorn and Ziv, 2008; Leuz, 2000). To provide insights into the question of how convergence arises, we examine a subset of firms that

were involved in the standard-setting process (i.e., potential disclosure leaders) and compare their disclosure practices to those of other early disclosers. This analysis, in a sense, echoes the work of Daske, Hail, Leuz, and Verdi (2013), which recognized considerable firm-level discretion in implementing IFRS and separated credible adopters from "label" ones. In our analysis, we exploit a novel aspect of SASB standard-setting: the inclusion of a set of firms participating in Industry Working Groups (IWGs), sector-level groups of investors, auditors, regulators, and companies that collaborated with SASB.

We split our sample between members of the IWGs and other early disclosers. We find a stark difference between the two groups in the timing of changes in their disclosure content. In the years preceding the release of the standards, IWG firms dramatically increased the amount of material disclosure in their reports, but their material disclosures remained largely unchanged in the post-period. On the other hand, the material information in the reports of other early disclosers was unchanged in the years leading up to the release of the standards but increased in the post-period in a way that resembled the IWG firms in the pre-period. These results provide evidence that standards, even voluntary ones, can serve as powerful guidance when a large sample of firms chooses voluntary disclosure in the absence of regulation. The analysis also suggests that disclosure leaders (i.e., IWG firms) can learn while doing, while other firms may be equally quick to respond once standards are known.

This paper contributes to several streams of literature. First, it contributes to the ESG literature specifically — and the unregulated disclosure literature more broadly — that emphasizes the need to analyze voluntary disclosure and the consequences of potential disclosure mandates (Bochkay, Hales, and Serafeim, 2021; Christensen, Hail, and Leuz, 2019, 2021; Leuz, 2018). Hail, Tahoun, and Wang (2018) shows that regulatory interventions may be delayed and ineffective and may have unintended consequences. Given the likelihood that ESG disclosure mandates, such as the ESG Disclosure Simplification Act of 2021 proposed by the U.S. Congress, will be implemented in lieu of regulation

of ESG-related behavior, our paper provides evidence about how firms' disclosures may converge in response to a mandate (Christensen et al., 2021; Hail et al., 2010; Leuz and Wysocki, 2016).

Relatedly, our paper extends research on the evolution of unregulated and unmandated disclosures (Bourveau et al., 2020; Daske et al., 2008; Leuz, 2000). We add to this literature by studying ESG reports, which present a new challenge to researchers, given that they lack a formal template, contain nonfinancial and textual information, and offer managers discretion about whether and what to disclose. To that end, we are among the first researchers to use state-of-the-art machine learning to clean, parse, quantify, and investigate not just the disclosure choice but also the heterogeneous content of these disclosures (Brown et al., 2020; Cao et al., 2021; Li et al., 2021). The resulting dataset provides detailed, transparent measures of ESG disclosures and will be made freely available to researchers interested in this topic.

Our paper also contributes to the literature that examines the impact of mandatory disclosure and informs regulators contemplating ESG disclosure mandates (Chen, Lewis, Schipper, and Zhang, 2017; Christensen, Hail, and Leuz, 2013; Einhorn, 2005; Grewal, Riedl, and Serafeim, 2019; Leuz, 2018; Leuz and Wysocki, 2016; Yip and Young, 2012). Further, our paper extends the recent literature related to ESG disclosure mandates outside of the United States, which has shown that regulation is positively associated with ESG disclosure quantity, analyst forecast accuracy, and real firm activities (Christensen, Floyd, Liu, and Maffett, 2017; Fiechter, Hitz, and Lehmann, 2022; Ioannou and Serafeim, 2019; Krueger, Sautner, Tang, and Zhong, 2021; Lin, Shen, Wang, and Yu, 2021). Unlike prior and concurrent research, we provide evidence that the content within ESG reports can converge toward material information in the absence of regulation, and that both firms and standard setters can play important roles in this process.

2 Institutional Setting

In this section, we discuss two important aspects of our institutional setting. First, we describe the growth in demand for ESG reporting, the literature that has examined ESG disclosures, and the challenges facing researchers in this area. Second, we discuss SASB standard-setting and explain its appeal for our research design.

2.1 ESG Reports and Ratings

Interest in ESG practices and reporting in business has swelled in recent years, with investors and firms devoting significant attention to the issue. Globally, in the early 1990s, fewer than 20 publicly traded firms issued reports that included ESG data. By 2014, the number around the world providing some information on ESG issues had increased to nearly 6,000 (Serafeim, 2013). In the United States, 83% of companies registered with the Securities and Exchange Commission (SEC) in 2017 disclosed some sustainability information in their regulatory filings (SASB, 2017).

Early firm-initiated disclosures of sustainability information tended to be reactive, with firms often disclosing in press releases or on company websites after high-profile scandals (Christensen et al., 2021). These practices became recognized as industry best practices and served as guidelines. In response to the recent growth in demand for ESG information from investors, stakeholders, and regulators, firm ESG disclosures began being centralized in a single document: the ESG report. Yet the content in these reports varied widely by firm, by industry, and over time, in part because ESG reports are not audited, mandated, or regulated in many jurisdictions, including the United States. This heterogeneity creates significant hurdles for researchers.

Still, a stream of academic literature has examined the firm-level determinants and contents of ESG reports (see Hahn and Kühnen (2013) for a detailed review). Though many of these studies relied on small samples or used settings outside of the United

States, they raise important concerns about the usefulness and consistency of the reports. Papers such as Simnett, Vanstraelen, and Chua (2009) and Manetti and Becatti (2009) document that firms seek voluntary assurance of their non-financial reports to enhance credibility but that there is great variation in whether firms seek assurance, the amount of information reported, and the validity of that information (Perego and Kolk, 2012). Other small-scale studies find that firms avoid discussing negative ESG events and exclude discussion of targets and performance (Boiral, 2013; Hubbard, 2011).

The literature on ESG ratings also captures the challenges of analyzing voluntary disclosures that allow for wide discretion. Research has found significant disagreements in ESG ratings from different data vendors (Berg, Kölbel, and Rigobon, 2022; Chatterji, Durand, Levine, and Touboul, 2016). This disagreement is likely due, in part, to the vendors basing their analyses on publicly available ESG information, including ESG reports, that provide different information for firms within the same sector. Further, these ratings are subjective because there are differences in how data vendors define, weigh, and measure ESG, which causes confusion among investors as to what ESG entails (Berg et al., 2022; Serafeim and Yoon, 2022).

2.2 Sustainability Accounting Standards Board

Partly in response to growing concerns about the lack of uniformity in ESG reporting in the face of increasing demand, organizations began developing and disseminating voluntary ESG reporting guidelines. Their aim has been to address the inconsistency of corporate ESG disclosures by improving or harmonizing reporting practices.

SASB is among the most prominent organizations providing guidance on materiality in ESG information. A nonprofit, it was founded in 2011 to develop and disseminate sustainability accounting standards so that publicly listed corporations had guidance on the financially material factors they should disclose in compliance with SEC requirements. SASB standards are designed for the disclosure of financially material sustainability is-

sues in mandatory SEC filings, such as the Form 10-K and 20-F.[10]

SASB's board, which oversaw the development of the standards, contained a mix regulators, academics, lawyers, and investors. Its standards were developed via a multi-stakeholder process, consisting of research supported by Bloomberg technology, data, and analytical tools; multi-stakeholder industry working groups (IWGs); a public comment period; and reviews by an independent Standards Council of experts in standards development, securities law, environmental law, metrics, and accounting.

Using the guidelines developed with the IWGs, SASB produced an initial set of provisional standards that defined a universe of 26 ESG-related topics. The goal of this effort was to create a comprehensive set of disclosures covering all aspects of ESG that firms could encounter within their operations. The materiality guidelines (i.e., the materiality map) reported in Figure 1 span five dimensions: environment, social capital, human capital, business model and innovation, and leadership and governance. SASB also determined which topics were material for each of 11 sectors with significant heterogeneity across sectors. In Figure 1, topics deemed most material to a sector are reported in black and those that are somewhat material are in gray. Topics in white are not material to firms in the sector. For example, greenhouse gas emissions are determined to be material for the extractive minerals and processing sector but not for the financial sector. The provisional standards were published by sector between July 2013 and March 2016. In Table IA.1, we report the publication dates for each sector's standards. For a detailed description of the standard development process, see Appendix A.1.

We note that GRI also provides widely recognized standards for ESG reporting. It was founded in 1997 and required that a GRI-compliant reports cover issues that reflect an organization's significant economic, environmental, and social impacts or issues that substantively influence the assessments and decisions of stakeholders. The key differences between SASB and GRI are the period when their standards were introduced and

[10]SASB officials have said that the reluctance to include this information in regulatory filings is driven in large part by litigation risk.

that SASB has an investor focus while GRI has a multi-stakeholder focus.[11]

We rely on SASB's voluntary standards as our benchmark for the identification and quantification of material disclosures because these standards provide the following unique advantages for our setting (i.e, the U.S. market) and period. First, SASB focuses on identifying appropriate disclosures that highlight the link between ESG issues and shareholder value. These disclosures aim to facilitate decision-making among investors and capital providers and were developed to be consistent with and incorporated in regulatory filings like the 10-K. This goal of financial materiality starkly contrasts with GRI's broader focus on stakeholders.

SASB also has become the most adopted comprehensive voluntary standards provider on material ESG issues in the United States.[12] Further bolstering the claim of SASB's primacy as a capital-markets-focused standard setter, following the merger of SASB and GRI, the new organization was consolidated into the IFRS Foundation in 2022, with the SASB standards serving as the starting point for the development of IFRS's Sustainability Disclosure Standards.[13] While these standards were originally created to be included in regulatory filings, in the absence of regulation firms have rapidly adopted SASB standards in their ESG reports. For example, Goldman Sachs includes in its 2021 ESG report a "SASB Index" section that begins as follows: "This report is evidence of our ongoing commitment to provide disclosures under the [SASB] standards."

Finally, from a research design perspective, SASB's standards offer two advantages. The staggered introduction of the standards allows us to identify treated and control firms in a difference-in-differences empirical specification. Also, unlike GRI, these standards were introduced when many firms were already releasing ESG reports, which allows for a large sample that spans both the pre- and post-standards periods.

[11] As disclosure of ESG issues is now of global interest, SASB and GRI merged into a unified organization, the Value Reporting Foundation, in 2022.

[12] https://www.thecaq.org/sp-500-and-esg-reporting/

[13] https://www.sasb.org/wp-content/uploads/2022/07/IFRS-Foundation-completes-consolidation-with-VRF_Release_Final.pdf.

3 Collection and Cleaning of ESG Reports

In this section, we describe our methodology for collecting, parsing, and measuring the content of ESG reports. The dataset and the underlying code described in this section will be made available to researchers and practitioners.

3.1 Collecting ESG Reports

The focus of this paper is the ESG reports of all firms that were included in the S&P 500 at least once from 2010 to 2021 — resulting in a sample of 613 unique firms with available data — which allows us to create a consistent time series for each firm. Because there is no centralized database of ESG reports, we begin with large-scale hand collection of data. We choose the S&P 500 firms for this exercise since larger firms are more likely to have the necessary resources to disclose and these firms provide a level of comparability within sectors.[14] As a starting point to conduct a complete search for ESG reports of all companies, we downloaded all available reports from firms' websites. We also conducted exhaustive searches of archival websites (i.e., archived versions of firms' websites and other publicly available sources). Note that, in more recent years, firms have commonly released numerous types of reports related to ESG activities, such as climate-specific reports and Equal Employment Opportunity Commission reports (EEO-1s). To ensure that we have a sample that can be consistently compared in the cross-section and over time, we only include aggregated reports that are described as "ESG reports," "sustainability reports," "CSR reports," or other similar terms. When we are uncertain as to which report to use, we manually inspect all reports and select the one that most comprehensively describes ESG activities. If no such report exists (e.g., if a firm releases only its EEO-1), we exclude that firm-year from our sample.

[14] Bourveau, Chowdhury, Rouen, and Le (2022) finds similar disclosure rates for S&P 500 firms but finds that fewer than 20% of non-S&P 500 firms (i.e., smaller firms) release ESG reports.

3.2 Extracting and Cleaning Text from ESG Reports

Having collected a total of 3,660 ESG reports for all firms in our sample, we next extract the text from these reports. Because the reports are unstructured, not standardized, and in PDF format, extracting text poses a significant challenge. To ensure accurate extraction, we used several machine learning techniques to preserve the textual contents of the documents as disclosed in the original file.

As a first step, we iterated through each page of each document and saved each page separately, which allowed us to create an optical scan of the page, resulting in high-quality images. From testing, we found that this strategy was superior to relying on commonly used libraries to directly identify the text in PDFs since PDF encoding introduces noise that results in excess and incorrect words. Next we used Google's open-source optical character recognition (OCR) engine, Tesseract, to extract the text from each page. Because the OCR process can introduce errors in the documents, we used Microsoft's natural language processing (NLP) engine, which studies the context of each sentence to correct misspellings in the text and remove any artifacts introduced during this procedure. We next used regular expression matching to remove any remaining special characters, numbers, website addresses, and other unnecessary information. Finally, we hand-verified the extracted text to ensure that we accurately constructed a corpus representing the ESG reports (the corpus). We then merged the extracted text of each page into a single file for each report. The result of these steps is a firm-year panel of the text of ESG reports. Note that, while we employed several research assistants for hand verification in our initial data collection and cleaning steps, the methodology we have developed results in a process that is largely automated.

4 Document-Level Analysis of ESG Reports

The rest of this paper is devoted to examining trends in ESG reporting and discussing our main empirical findings. We begin by documenting the growth in reporting and the firm characteristics associated with the decision to release an ESG report. We next turn to the content of these reports, examining changes in the aggregate text over time before turning to the topic-level analysis, where we analyze how both the amount of space devoted to a topic and the language used to describe that topic have evolved. We conclude by examining how topic-level reporting changed around the release of the SASB standards and whether firms involved in standard-setting (i.e., IWGs) had different reporting practices than those not involved in the process.

4.1 The Growth of ESG Reports

Anecdotal evidence suggests that ESG reports have become the main means by which firms disclose ESG-related activities, but, due to the absence of a clearinghouse for these reports, there is little empirical evidence that this is the case. Figure 2 provides empirical evidence for the claim of ESG reports' rapid growth. In aggregate, from 2010 to 2020, the percentage of firms releasing ESG reports grew from 35% to 86%. There is a slight decrease in reports in 2021 because a subset of firms that released 2020 reports had yet to do so for 2021 at the time of data collection. At the sector level, as reported in Table 1, Panel A, we see similar near-monotonic trends across most sectors.

While ESG reports are becoming more frequent, our evidence suggests that they are not becoming much larger in terms of word count. Table 1, Panel B, shows that the average number of words in these reports has grown from 7,312 in 2010 to 10,403 in 2020, with significant variation across sectors. This result is surprising, given that the scope of ESG issues has grown during this period and that other filings, such as the 10-K, have grown significantly (Cohen et al., 2020).

4.2 Determinants of Releasing an ESG Report

Small-scale studies have examined some of the determinants of voluntarily disclosing an ESG report, but the factors that are associated with the disclosure choice remain unexplored to some extent (Hahn and Kühnen, 2013). In this section, we examine firm-level covariates we expect to relate to the choice to publish an ESG report. Because these variables are likely to be associated not just with the choice to disclose but also the content of that disclosure, we include them as control variables in our main analysis.

4.2.1 Descriptive Statistics

The decision to disclose — and what to disclose in — ESG reports is likely driven by various firm characteristics. To that end, we consider a set of firm-level covariates. Firm Size is the natural logarithm of the market value of equity. Larger firms have greater resources and a more robust disclosure environment, so we expect that these firms are more likely to disclose and to disclose more information. Market-to-book is the market value of equity divided by the book value of equity. Growth firms may need to disclose more to help justify their higher market valuations. We include ROE, the firm's return on equity, as a measure of financial performance. While it is difficult to predict the relation between firm performance and ESG disclosures, Bourveau et al. (2022) finds that firms with weaker performance are more likely to provide human capital disclosures, perhaps to justify this cost to investors. R&D/Sales is R&D expense over sales. Capex/PP&E is capital expenditure divided by property, plant, and equipment. SG&A/Sales is selling, general, and administrative expense over sales. Adv Exp/Sales is advertising expense over sales. These four measures capture aspects of the firm's operating structure and investments and are likely to be associated with the disclosure choice, although making directional predictions is challenging. Leverage is the long-term-debt-to-total-assets ratio. Less financially constrained firms may have more resources to dedicate to ESG investments and disclosures.

We also include the following variables: ESG Score is the ESG rating of the firm from the Refinitiv database. This variable is to account for the possibility that firms with more ESG investments would be more inclined to disclose their ESG endeavors. Incidents is the sum of all ESG scandal events at the firm in the prior year, as defined by the RepRisk database, and is included because scandals may encourage firms to disclose more ESG information. Institutional Ownership is the percentage of shares held by institutional investors and accounts for the pressure from these investors to incorporate and disclose ESG activities. Number of Words is the total number of words in the firm's 10-K, which is downloaded from the Notre Dame Software Repository for Accounting and Finance (SRAF) (Loughran and McDonald, 2016). We include this variable to control for concurrent growth in the length of regulatory disclosures. Assurance is an indicator equal to 1 if the firm's ESG report received outside assurance, as identified by Refinitiv. We include this variable because prior literature has shown that firms' choice to have their reports assured is related to endogenous firm ESG activities (Simnett et al., 2009).[15]

Summary statistics for these variables are reported in Table 2. For the 3,660 firm-years, we find that the average market value of the firm is more than $23 billion with a market-to-book ratio of 4.38 and a return on equity of 0.17. These firms devote 3% of their sales to R&D and 1% to advertising. They are largely controlled by institutional owners, with an average of 80% of shares being held by institutions. As measured by Assurance, 31% of the ESG reports in our sample received outside assurance to some degree.

[15]We also note that the percentage of ESG reports receiving some level of assurance, while still small, has increased rapidly during our sample period. In untabulated analysis, we find that the percent of ESG reports receiving assurance increased from 15% in 2010 to 38% in 2020.

4.2.2 Firm characteristics and disclosure choice

What firm characteristics are associated with the decision to release an ESG report? We examine this relation by implementing the following equation:

$$Disclose_{i,t} = \boldsymbol{X}_{i,t}\gamma + \delta_i + \delta_t + \epsilon_{i,t}, \tag{1}$$

where Disclose is an indicator equal to 1 if firm i releases an ESG report in year t, and 0 otherwise. $\boldsymbol{X}_{i,t}$ is a vector of all of the firm-year variables described above. We include firm and year fixed effects to control for time-invariant firm characteristics and unobservable trends in reporting. The confidence interval is set to 95%. We scale all variables to have means of 0 and standard deviations of 1 to allow for comparability.

The results of this analysis are reported in Figure 3. While we cannot draw causal conclusions from this analysis, it is interesting to note that Incident and Number of Words are both significantly correlated with the disclosure decision. Firms with more negative ESG incidents are more likely to release an ESG report, which is consistent with prior literature that argued that firms often disclosed ESG information in response to negative ESG events (Christensen et al., 2021). Interestingly, firms with fewer words in their 10-Ks, as measured by Number of Words, are more likely to publish an ESG report. This relation suggests that firms with less-detailed regulatory filings may shift disclosure into the unaudited ESG report. Lastly, although institutional owners have helped drive firms toward more ESG disclosures, we find no statistical relation between releasing a report and institutional holdings. This is likely due to the average institutional holdings in our sample being 80% of all shares.

4.3 Aggregate Changes in the Content of ESG Reports

We now shift from examining the disclosure choice to examining the content of these reports, beginning by quantifying the content in aggregate. To do so, we document how

the similarity of the text between ESG reports changes during our sample period along two dimensions: the similarity of ESG reports for firms in the same sector and year (i.e., within sector) and the similarity of reports across sectors within the same year (i.e., across sectors).

Figure 4 examines the average annual cosine similarities within and across sectors. The blue line shows that firms in the same sector increasingly use similar language in their ESG reports, while the red line shows similar trends across sectors. These results, while descriptive, suggest that firms are converging toward a common ESG vocabulary. Combined with evidence that the length of these reports has not significantly changed during our sample period, these findings point to firms learning to disclose not more but potentially more comparable information. Still, examining disclosures at the document level is insufficient to support this claim, and so we next turn to our topic-level analysis to better understand whether the content of firms' ESG reports is becoming more focused on issues material to investors.

5 Topic-Level Analysis of ESG Reports

As mentioned previously, ESG reports are largely text-based, meaning that effective analysis of these reports requires researchers to quantify the text in a way that is comparable across firms and across time. In this section, we begin by describing how we use the SASB standards to quantify the language within ESG reports at the topic level. Next, we show how the language within these topics has evolved across time in important ways that would be difficult to detect without large-scale analysis. This section concludes with a set of analyses that provide us with confidence that our topic measures reflect the underlying disclosures we seek to quantify.

5.1 Quantifying Content within ESG Reports

In our previous analysis, we quantified ESG reports at the document level, a common approach to textual analysis that treats words as independent and tends to ignore context but one that has not been applied to ESG reports. Research in accounting and finance has used this type of approach to examine the narrative components of financial disclosures (Brown and Tucker, 2011; Cohen et al., 2020; Li, 2008). This level of analysis efficiently examines trends in ESG disclosures but ignores the rich heterogeneity within these reports, so in our main empirical analyses, we use a neural network model to quantify the content (i.e., the meaning of what is being written) within these disclosures. These two levels of measurement offer an opportunity to provide rich descriptive and empirical evidence of reporting trends and detailed disclosure choices in both how much information on a topic is being disclosed and what language is being used to describe each topic.

We argue that textual analysis (as opposed to an analysis of reported quantitative metrics) of ESG reports is the most appropriate way to study these disclosures. Despite the rise in ESG reports, there remains a lack of consistency and comparability in what is being disclosed and the metrics used in the disclosures (Amel-Zadeh and Serafeim, 2018). This lack of quantitative uniformity poses a challenge for researchers seeking to conduct large-scale analysis, given that extracting metrics to judge firms' disclosure choices is infeasible.[16] In the cross-section, it is a challenge both to judge the appropriateness of the metrics used and to compare the activities those metrics describe.

Our textual analysis approach, on the other hand, provides an opportunity to create comparable measures of what is being discussed and whether it matters to investors. The largely qualitative nature of ESG reports lends itself well to textual analysis, and examining the text of these disclosures is likely to be a more fruitful endeavor at this

[16]For example, Bourveau et al. (2022) focuses exclusively on human capital metrics and finds nine categories of metrics. There is significant variation in the metrics used within these categories as well as the channels through which they are disclosed.

point, compared to a comparison of inconsistent, unaudited metrics.

To quantify the text within ESG reports, we rely on a *word2vec* topic modelling algorithm, a semisupervised word embedding neural network model that identifies the meaning of words and phrases to associate them with broad topics.[17] Given our strong priors about the topics we are interested in modelling (i.e., the 26 ESG topics defined by SASB), this approach is likely to be superior to an unsupervised approach in that it allows us to label topics to ease interpretation. Further, we can provide priors (i.e., seed words) to the algorithm as a starting point to help uncover additional related terms. Unlike the common document-level approach to textual analysis, our word embedding approach vectorizes words and phrases, identifying relationships between them to quantify their meaning. In other words, applying the *word2vec* algorithm to ESG reports results in the quantification of the semantics of the content.

As an example of how the algorithm works, we examine one of the 26 ESG topics, business ethics. To uncover an initial set of words associated with this topic, we first provide a small set of seed words from the SASB provisional standards that uniquely describe the topic. For business ethics, SASB provides the terms "conduct," "corruption," and "bribery" (among others). These words provide initial conditions for the *word2vec* algorithm. Using our seed words, the algorithm analyzes the full corpus of text and generates an initial expanded dictionary of words that are associated with the topic. We then hand-check this newly created dictionary of words generated by the algorithm and remove words that are erroneously added to the topics but do not clearly reflect the topic's subject. This is done, in part, to help reduce the Type I errors introduced by the algorithm. Using this expanded dictionary, we re-run our topic model to identify and quantify the content of each topic.

Table 3 reports the five words that the algorithm identifies within the expanded dictionary as most important for each of the ESG topics. We report the top 50 most im-

[17] The algorithm and code are described in detail by Li et al. (2021)

portant words for each topic in Table IA.2. For business ethics, the words include those identified from the SASB standards as well as others like "money_laundering," and "financial_crime." In addition, in Figure 5, we report word clouds that show the most commonly occurring words for the topics business ethics and GHG emissions (word clouds for the other 24 topics are reported in Figure IA1). In Panel A, the words most frequently used when discussing business ethics are "risk," "subject," and various words describing codes of conduct as well as "violation" and "corruption."

Note that, for a specific ESG report, the words within a topic do not need to mention the seed words for the algorithm to identify and quantify discussion of that topic. This approach improves upon prior unsupervised strategies in that it allows us to specify topics, yet it gives us the flexibility to uncover previously unidentified key words associated with each topic.

5.2 Measuring Topic Disclosure

The ESG dictionary we create allows us to measure the amount of discussion of each of the 26 ESG topics at the document-year level. Our dataset is organized at the topic-document-year level. In other words, there are 26 different measurements for each document-year, one for each topic. We follow Li et al. (2021) in using the term frequency-inverse document frequency (TF-IDF) as our primary measure at the topic level. The term frequency is a simple count of the number of words within a topic-document. This measure fails to account for the importance or relevance of words, so we weight the term frequency by the document frequency of each word, measured as the number of documents in the corpus that include the word. The resulting weighted measure accounts for the relative importance of each word or phrase to the topic.

Table 4 reports the summary statistics for TF-IDF at the topic level and provides insights into the most commonly discussed ESG topics within ESG reports. The most common topics are those related to risk factors and resource management. Specifically,

the top five topics, in order, are business model resilience, supply chain management, energy management, systemic risk management, and management of the legal and regulatory environment. Interestingly, all of these topics relate directly to operational risks within the firm, topics that are frequently discussed in regulatory reports like the 10-K.

5.3 Within Topic Analysis

Given that the language in ESG reports has evolved over time, it is likely that terms are introduced into the reports while others become less frequent. For example, the COVID-19 pandemic introduced a significant risk to firms' various stakeholders that likely introduced new terms into certain topics in 2020 disclosures that had not been discussed in previous ESG reports. This evolution of language poses challenges to techniques like bag-of-words analysis, which would require an update to the dictionary to capture changes to language. The *word2vec* algorithm, though has significant advantages when quantifying topics with changing language. Given our strong priors, we can define topics and provide seed words, yet allow the algorithm to determine other words that are associated with these topics and terms. Further, we are able to train the algorithm over the entire sample so that it identifies new terms that are introduced into the business lexicon, as well as the topics to which these new terms belong.

To underline these advantages, we consider how the language has changed *within* each topic over time. In other words, we explore in Figure 6 how the words that comprise each topic have evolved as the issues covered within those topics have changed. To do so, we show a ranked partial list of terms associated with a topic in the first column (the 10 most important terms in 2010, as measured by their relative term frequency, and a subset of other words that became among the most important words in 2020). We then document how the importance of these words to a topic changed from 2010 to 2020.

Panel A focuses on the Consumer Welfare topic and exemplifies the importance of our flexible algorithm when measuring language that is evolving. The words "vaccine,"

"pharmaceutical," and "animal_welfare" were not in the top 10 in 2010 but were the eighth, ninth, and tenth most important words in 2020 for this topic. We also see an upward migration of "medicine" and a downward migration of "disease." This evolution likely reflects firms changing their discussion in response to their business and stakeholder demands, including responses to the COVID-19 pandemic (a virus, not a disease) and the increased possibility of regulation related to the treatment of farm animals.

We see another form of language evolution when studying the Customer Privacy topic in Panel B of Figure 6. Terms such as "intellectual_property," "information_technology," and "enterprise_risk" became less important from 2010 to 2020, while the terms "cybersecurity" and "datum_privacy" were not in the top 10 in 2010 but became the sixth and ninth most important words in 2020. Unlike the previous example, which showed firms shifting their language in response to events, the evidence in Panel B suggests that firms shifted their language as catchall terms like cybersecurity worked their way into the lexicon, replacing similar but possibly less specific terms.

To further document these evolutions in language, we report alluvial graphs for the rest of the topics in Figure IA2. Taken together, this analysis not only documents how language evolves, but also provides us with confidence that our approach to measuring topic-level disclosures is effective at capturing that evolution.

5.4 Validating the Topic-Level Measures

To provide us with additional confidence that the TF-IDF scores are relevant and appropriately reflect the topics we define, we next examine the correlations among topics, the correlations between topics and firm-level covariates, and the correlations between topics and commercial ESG ratings. Figure 7 provides a correlation map of the relations among topics, with dark shades of blue signifying more positively correlated and darker shades of red signifying more negatively correlated.

Most topics are not strongly correlated with others, meaning that the algorithm is

likely capturing unique characteristics of each topic. There are limited strong positive correlations between related topics, such as "materials sourcing and efficiency" and "waste and hazardous materials management." There are also limited strong negative correlations between unrelated topics, like "employee engagement, diversity, and inclusion" and "GHG emissions." These negative correlations could reflect the trade-off between disclosure of different types of information, or they could reflect that, in sectors where employee engagement is material, emissions often are not.

Figure 8 reports correlations between each topic and firm-level characteristics. Among the strongest correlations are those where the topic discussed relates directly to the firm characteristic. For example, there is a strong positive relation between Adv Exp/Sales and the topic "selling practices and product labeling." R&D/Sales is strongly associated with "data security" and "product design and lifecycle management." These associations suggest that firms use their ESG reports, in part, to discuss issues directly related to their operations.

While recent literature finds that commercial ESG ratings often contradict one another, in our setting, it is worth exploring their relation to our topic since ratings are based on firm disclosures and attempt to measure underlying activity, not disclosure itself (Berg et al., 2022). Given that firms are more likely to disclose positive ESG information, we expect commercial ratings to be positively associated with our topic scores (Boiral, 2013).

Consistent with this expectation, in Table IA.3 we find positive associations between our topic scores and commercial ratings from Refinitiv and MSCI. Importantly, we do not find significant relations between the ESG scores and the interaction term TF-IDF $\times$ All Material Topics, where All Material Topics is an indicator equal to 1 if the topic measured by TF-IDF is material for that sector and 0 otherwise. This suggests that, while our measures are capturing some of the information contained in commercial ESG ratings, these ratings fail to capture sector-level variation in materiality. Taken together,

this series of results provides us with confidence that we are appropriately measuring disclosure related to specific ESG activities.

6 Are ESG Reports Evolving Toward Material Disclosures?

A large body of literature in accounting and finance has quantified the text of firm disclosures, with much of the analysis conducted at the document level, using an approach that treats each word as independent. When examining whether firms are releasing ESG information that is material to investors, this approach is infeasible because it ignores the heterogeneity within the documents, which discuss numerous topics in various ways. Our semisupervised neural network model approach to measuring topics within the reports overcomes this challenge and allows us to investigate whether ESG reports evolved toward more material disclosures.

6.1 Sector-Level Content in ESG Reports

Before examining the evolution of the content of these disclosures, we first look within sector at the topics firms discussed and the amount of content devoted to these topics. Table 5 reports the five most material and immaterial (as defined by SASB) topics that are discussed in each sector. The average percentage of the report devoted to each topic, defined as the sum or words for that topic divided by total words in the ESG report, is reported in parentheses. On average, firms devote more of their reports to material topics than they do to immaterial ones. For example, as reported in Panel A, the most discussed topic for firms in the consumer goods sector is supply chain management, which is material. Firms, on average, devote 13.5% of their reports to this topic, 66% more than they do to the most-discussed immaterial topic, business model resilience (8.11%).

That being said, there is significant heterogeneity in discussion of material topics at

the sector level. As reported in Panel C, the most common topic in the financials sector is business model resilience, which is deemed immaterial by SASB. This analysis, though, is done at the aggregate level and does not account for the fact that firms were learning how to disclose during this time and that ESG reporting was evolving. In our next set of analyses, we examine how these reports have evolved and whether firms have gravitated toward material disclosure.

6.2 Examining Disclosure Materiality

With SASB's definitions of material topics as our guideposts, we ask whether firms are guided by financial materiality in their ESG disclosures. There are at least three reasons why firms may not be disclosing material information. First, they may not deem sector-level materiality, as defined by SASB, to be appropriate guidance. Second, firms may avoid discussing material topics if the activities they describe depict them in a bad light. Third, investors may not be the target audience for these disclosures since they are already the target audience for significant disclosures (e.g., SEC filings and conference calls) and ESG reports discuss firms' impact on multiple stakeholders. Still, it is important to note that ESG reports often are posted on firms' investor relations websites, suggesting that investors are the intended audience.

In Table 6, we examine whether firms' ESG disclosures, on average, focus on material information by implementing the following regression.

$$TF-IDF_{i,t,j} = \beta_1 I(AllMaterialTopics)_{j,k} + \delta_{i,t} + \delta_{j,t} + \epsilon_{i,j,t} \quad (2)$$

where j indexes a topic disclosed by firm i in sector k at time t. TF-IDF is our topic measure, and All Material Topics is an indicator equal to 1 if the topic is material to that sector (and 0 if not). In addition to a series of firm-level controls, we include high-dimensional fixed effects to address unobserved heterogeneity. We use topics-by-year

effects, $\delta_{j,t}$, to address concerns that our results are driven by some topics becoming more important over time, such as if firms, on average, respond to investors' shifting preferences by disclosing more on certain topics in different years. In our most restrictive specification, we add firm-by-year fixed effects, $\delta_{i,t}$, that absorb time variation at the firm level to control for factors, such as financial performance and changes to regulatory disclosures like the documented growth in 10-Ks (Cohen et al., 2020).

Estimating Equation 2, we find that, on average, firms are guided by materiality in their ESG disclosures. Our baseline result is that, on average, firms disclose 48% more on material topics relative to immaterial topics.[18] In all columns, the coefficient on All Material Topics is positive and statistically significant at 1%, including in Column (4), which examines a narrow window from four years before to four years after the release of standards to increase the likelihood that the results are driven by the event of interest. These results provide our first set of evidence that firms disclose ESG information that aligns with a set of standards defining materiality. We are careful to avoid making claims of causality in any of our analyses, given the numerous potential omitted variables that could bias our results. However, when we progressively add controls and fixed effects, the R-squared increases, but our estimates do not change, suggesting that there are not omitted variable problems (Oster, 2019).

6.3 The Evolution of Material Disclosures

We next examine graphically at the sector-topic level how firms' emphasis on each topic in their ESG reports evolves. Figure 9 reports heat maps for two sectors, financials and extractives and minerals processing, to document the relative information devoted to each topic in each year in our sample. Green (red) bars represent topics that are material (immaterial) to the sector, and darker shading means that firms, on average, devote more

[18] The sample mean of TF-IDF is 60.23, so economic materiality is calculated as 29.00/60.23, where 29.00 is the coefficient in Column (3).

of the ESG report to a topic. In the financials sector, firms increased their disclosures of the material topics "access and affordability," "employee engagement, diversity, and inclusion," and "customer privacy." They also increased disclosures related to "business model resilience" and "management of the legal and regulatory rnvironment," even though SASB deemed these immaterial.

The extractives and minerals processing sector saw increases in numerous material topics, including "employee health and safety" and "business model resilience." The industry also increased disclosures on immaterial topics related to employee engagement and systemic risk management. These results suggest that firms, on average, increased discussion of material topics. We report heatmaps for the rest of the 24 topics in Figure IA3. In the next section, we examine whether these increases were concentrated around the release of the SASB standards.

6.4 Material Disclosures Around the Release of the SASB Standards

SASB released its provisional standards in a staggered fashion at the sector level from 2013 to 2016, with the health care sector having the first public standards in June 2013 and the infrastructure sector having the last in March 2016 (see Table IA.1). In the following analyses, we exploit this feature of standards-setting to analyze how firms' disclosures changed around the publication of standards. Specifically, we ask whether firms increased their material disclosures after the release of the SASB standards. To examine this question, we use a stacked difference-in-differences approach, implementing the following equation:

$$\begin{aligned} TF-IDF_{i,j,t} &= \beta_1 I(AllMaterialTopics)_{j,k} \\ &\quad +\beta_2 I(AllMaterialTopics)_{j,k} \times I(Post)_{k,t} + \delta_{i,t} + \delta_{j,t} + \delta_{j,k} + \epsilon_{i,j,t} \end{aligned} \tag{3}$$

where j indexes a topic disclosed by firm i in sector k at time t. TF-IDF is our topic measure, Post is an indicator equal to 1 if the firm is in a sector where standards have been published (and 0 otherwise), and All Material Topics is an indicator equal to 1 if the topic is material to that sector (and 0 if not). All Material Topics $\times$ Post is the interaction of the two and our main variable of interest. As in our prior analysis, we include all of the firm-level control variables and saturate the model with high-dimensional fixed effects, similar to Equation 2. We also include include topic-by-sector effects, $\delta_{j,k}$, to address the level of importance of each topic for each sector.

Recent research has found that staggered difference-in-differences designs can produce biased estimates under certain conditions (Borusyak and Jaravel, 2017; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021). Groups treated toward the end of the period could be controls at the beginning, and those treated toward the beginning could serve as controls toward the end. In addition, Barrios (2021) points out that, in staggered difference-in-differences with two-way fixed effects, estimates are variance-weighted average effects with some weights being negative. We mitigate this concern of bias by using stacked regressions, following the suggestions of Cengiz et al. (2019) and Baker et al. (2022). In this design, we create groups that are event specific and include controls that are unaffected by the event within the estimation window. These groups are then stacked in relative time across all events, resulting in a simulation where all treatments occur at once.

Table 7 reports the results from implementing Equation 3. In all specifications, the coefficient on All Material Topics $\times$ Post is positive and statistically significant. In Column (5), with our most restrictive fixed effects structure, the coefficient, 6.610, is positive and significant at the 1% level. Economically, this result suggests that firms increased their relative discussion of material topics by 11.0% after the release of the SASB standards.[19] In addition, using the narrow window of four years before to four years after the release

[19]The sample mean of TF-IDF is 60.23, so economic materiality is calculated as 6.610/60.23, where 6.610 is the coefficient in Column (5).

of the standards to reduce concerns about other events influencing our findings, our results, as reported in Column (6), remain unchanged. As in our last analysis, as we add controls and fixed effects, our R-squared increases, but our estimate does not change, suggesting that there is not an omitted variable problem (Oster, 2019). In summary, our evidence suggests that the amount of financially material information included in ESG reports significantly increased after the introduction of standards defining materiality, an important finding given that ESG reports remain unaudited and that the SASB standards remain the dominant guide for firms reporting ESG information.

6.5 ESG Disclosure Leaders and Followers

We exploit another feature of standards setting to understand how firms learn about materiality and the role of firms that take the lead in creating standards. Specifically, we reexamine the analysis we describe in Equation 3, but we do so separately for firms that were part of IWGs and other firms that were also disclosing prior to the release of the standards but were not involved in the process, which we call non-IWG firms.

Table 8 reports the results from our analysis for these two groups of firms. Examining IWG firms in Panel A, we find little evidence that these firms significantly increased their material disclosures after the release of the standards. Non-IWG firms, though, significantly increased their material disclosures after the release of the standards. In all specifications, the coefficients on All Material Topics $\times$ Post are positive and significant at the 1% level.

This regression framework raises a question: Did IWG firms not follow their own advice? Another possibility is that they used the standard-setting process to improve their disclosures in expectation of the publication of the standards, while aspects of that process induced non-IWG firms to change their ESG disclosure. This is an important question and has similarities to those posed by previous studies on unmandated disclosure. Leuz (2000) and Bourveau et al. (2020) show that capital market forces can drive

firms to a near-full disclosure equilibrium in the medium to long run through repetition and learning. Relatedly, Daske et al. (2013) note considerable firm-level discretion in implementing IFRS and separate credible adopters from label adopters, providing evidence on the differential capital market consequences for the two groups.

We conduct dynamic analysis of the previous tests and report the results in Figure 10. Figure 10 shows how the amount of material information changed for IWG firms and non-IWG firms from four years before to four years after the release of the standards. The differences between these groups is stark. Consistent with the findings in Table 8, we find that the amount of material information reported by IWG firms remained largely unchanged after the passage of the standards. In the period leading up to the standards, though, these firms significantly increased the amount of material information in their reports. On the other hand, material information in the reports of non-IWG firms was unchanged from four years before the release of the standards until the standards were released. In the four years after the release, though, the amount of material information in their reports increased monotonically. These results are echoed in Table 9.

Taken together, these results provide a compelling portrait of the relation between standard-setting and firms' disclosures. On average, firms gravitated toward material disclosures once materiality was defined, but this response was not uniform across firms. Those involved in standard-setting appeared to learn while doing, gradually increasing their material disclosures as the process continued. Other firms appeared to respond to the release of the standards and increased their material disclosures at a slower pace over a similar period.

6.6 Robustness and additional analyses

6.6.1 Robustness of results to alternative topic measures

One concern in our analysis is that TF-IDF does not accurately capture the discussion in the documents. Specifically, N occurrences of a specific term in a document may not truly carry N times the significance of that term, which is what TF-IDF measures. As such we consider a variant of our topic measure that uses the logarithm of the term frequency to compute the weighted frequency-inverse document frequency or WF-IDF. Using WF-IDF as our dependent variable, we find similar results to those in prior analysis using TF-IDF. We re-estimate Equation 3 and report the results in Table IA.4. Inspecting the coefficient on the interaction term, we again find evidence of firms increasing their disclosures of material topics following the introduction of provisional standards. In addition, in untabulated results, we re-run our main analysis using the unweighted term frequency and again find statistically significant results consistent with out main findings. These results assuage concerns that our results are driven by how we measure disclosures in ESG reports.

6.6.2 The influence of missing reports

Given the challenges of collecting all ESG reports and the fact that we are attempting to collect reports from the previous 12 years, there may be reports that were released but are no longer available online, which raises concerns that unobservable features of the sample are driving our results. We address this concern by randomly dropping 20% of our sample for each year and then conducting our entire analysis again, beginning with the training of the algorithm. Table IA.5 reports the replication of Table 7 for this reduced sample. Our inferences remain unchanged, mitigating concerns that availability of ESG reports is driving our results.

7 Conclusion

This paper documents the evolution of one of the fastest growing voluntary disclosures in history, the standalone ESG report. We show that, during the last 11 years, these reports have been widely adopted among the largest firms in the United States and that the language used within them has become increasingly similar across firms over time. Applying a semisupervised neural network model to the text of these reports to measure the amount of financially material information, as defined by SASB, we find that the language in these reports has evolved to reflect changes to the operating environment and that firms, on average, disclose a large amount of material information. Importantly, the public release of the SASB standards was associated with a meaningful increase in the amount of disclosed material information, although this increase was not uniform across our sample. Firms that played a role in the development of the SASB standards gradually increased their material disclosures during the development period, while other firms increased material disclosures at a similar pace once the standards were released.

These results show how disclosures across a diffuse set of ESG topics can be harmonized in the absence of regulation, which is of growing interest, due to the current frustrations with the ESG disclosure landscape and the nascent efforts by regulators to mandate these disclosures (IFRS, 2021; SEC, 2021). This paper also is among the first to provide large-scale analysis of ESG reports. Still, given the increasing importance of ESG disclosures to the investment community and others, there remain many unanswered questions about the content of ESG reports and their usefulness to various stakeholders. To help speed the analysis of these reports and further advance the study of this disclosure phenomenon, we make our data and underlying code publicly available.

References

Amel-Zadeh, Amir, and George Serafeim, 2018, Why and how investors use ESG information: Evidence from a global survey, *Financial Analysts Journal* 74, 87–103.

Baker, Andrew C, David F Larcker, and Charles CY Wang, 2022, How much should we trust staggered difference-in-differences estimates?, *Journal of Financial Economics* 144, 370–395.

Barrios, John Manuel, 2021, Staggeringly problematic: A primer on staggered did for accounting researchers, Available at SSRN 3794859.

Berg, Florian, Julian F Kölbel, and Roberto Rigobon, 2022, Aggregate Confusion: The Divergence of ESG Ratings, *Review of Finance* .

Bochkay, Khrystyna, Jeffrey Hales, and George Serafeim, 2021, Disclosure standards and communication norms: Evidence of voluntary disclosure standards as a coordinating device for capital markets, Available at SSRN 3928979.

Boiral, Olivier, 2013, Sustainability reports as simulacra? a counter-account of a and a+ gri reports, *Accounting, Auditing & Accountability Journal* .

Borusyak, Kirill, and Xavier Jaravel, 2017, Revisiting event study designs, Working Paper.

Bourveau, Thomas, Matthias Breuer, and Robert C Stoumbos, 2020, Learning to disclose: Disclosure dynamics in the 1890s streetcar industry, Available at SSRN 3757679.

Bourveau, Thomas, Maliha Chowdhury, Ethan Rouen, and Anthony Le, 2022, Human capital disclosures, Available at SSRN.

Brown, Nerissa, Richard Crowley, and W. Brooke Elliott, 2020, What are you saying? using topic to detect financial misreporting, *Journal of Accounting Research* 58, 237–291.

Brown, Stephen V, and Jennifer Wu Tucker, 2011, Large-sample evidence on firms' year-over-year md&a modifications, *Journal of Accounting Research* 49, 309–346.

Byard, Donal, Ying Li, and Yong Yu, 2011, The effect of mandatory ifrs adoption on financial analysts' information environment, *Journal of Accounting Research* 49, 69–96.

Callaway, Brantly, and Pedro HC Sant'Anna, 2021, Difference-in-differences with multiple time periods, *Journal of Econometrics* 225, 200–230.

Cao, Sean, Yongtae Kim, Angie Wang, and Houping Xiao, 2021, From words to syntax: Identifying context-specific information in textual analysis, Available at SSRN 3568504.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer, 2019, The effect of minimum wages on low-wage jobs, *Quarterly Journal of Economics* 134, 1405–1454.

Chatterji, Aaron K, Rodolphe Durand, David I Levine, and Samuel Touboul, 2016, Do ratings of firms converge? implications for managers, investors and strategy researchers, *Strategic Management Journal* 37, 1597–1614.

Chen, Qi, Tracy R Lewis, Katherine Schipper, and Yun Zhang, 2017, Uniform versus discretionary regimes in reporting information with unverifiable precision and a coordination role, *Journal of Accounting Research* 55, 153–196.

Christensen, Hans B, Eric Floyd, Lisa Yao Liu, and Mark Maffett, 2017, The real effects of mandated information on social responsibility in financial reports: Evidence from mine-safety records, *Journal of Accounting and Economics* 64, 284–304.

Christensen, Hans B, Luzi Hail, and Christian Leuz, 2013, Mandatory ifrs reporting and changes in enforcement, *Journal of Accounting and Economics* 56, 147–177.

Christensen, Hans B, Luzi Hail, and Christian Leuz, 2019, Adoption of CSR and sustainability reporting standards: Economic analysis and review, ECGI Working Paper Series.

Christensen, Hans B, Luzi Hail, and Christian Leuz, 2021, Mandatory csr and sustainability reporting: economic analysis and literature review, *Review of Accounting Studies* 26, 1176–1248.

Cohen, Lauren, Christopher Malloy, and Quoc Nguyen, 2020, Lazy prices, *Journal of Finance* 75, 1371–1415.

Daske, Holger, Luzi Hail, Christian Leuz, and Rodrigo Verdi, 2008, Mandatory ifrs reporting around the world: Early evidence on the economic consequences, *Journal of Accounting Research* 46, 1085–1142.

Daske, Holger, Luzi Hail, Christian Leuz, and Rodrigo Verdi, 2013, Adopting a label: Heterogeneity in the economic consequences around ias/ifrs adoptions, *Journal of Accounting Research* 51, 495–547.

De Franco, Gus, Sagar P Kothari, and Rodrigo S Verdi, 2011, The benefits of financial statement comparability, *Journal of Accounting Research* 49, 895–931.

Einhorn, Eti, 2005, The nature of the interaction between mandatory and voluntary disclosures, *Journal of Accounting Research* 43, 593–621.

Einhorn, Eti, and Amir Ziv, 2008, Intertemporal dynamics of corporate voluntary disclosures, *Journal of Accounting Research* 46, 567–589.

Fiechter, Peter, Jörg-Markus Hitz, and Nico Lehmann, 2022, Real effects of a widespread csr reporting mandate: Evidence from the european union's csr directive, *Journal of Accounting Research* .

Grewal, Jody, Edward J Riedl, and George Serafeim, 2019, Market reaction to mandatory nonfinancial disclosure, *Management Science* 65, 3061–3084.

Hahn, Rüdiger, and Michael Kühnen, 2013, Determinants of sustainability reporting: a review of results, trends, theory, and opportunities in an expanding field of research, *Journal of Cleaner Production* 59, 5–21.

Hail, Luzi, Christian Leuz, and Peter Wysocki, 2010, Global accounting convergence and the potential adoption of ifrs by the us (part i): Conceptual underpinnings and economic analysis, *Accounting Horizons* 24, 355–394.

Hail, Luzi, Ahmed Tahoun, and Clare Wang, 2018, Corporate scandals and regulation, *Journal of Accounting Research* 56, 617–671.

Hubbard, Graham, 2011, The quality of the sustainability reports of large international companies: An analysis, *International Journal of Management* 28, 824.

IFRS, 2021, IFRS foundation announces international sustainability standards board, consolidation with cdsb and vrf, and publication of prototype disclosure requirements, (Available at: https://www.ifrs.org/news-and-events/news/2021/11/ifrs-foundation-announces-issb-consolidation-with-cdsb-vrf-publication-of-prototypes/.

Ioannou, Ioannis, and George Serafeim, 2019, The Consequences of Mandatory Corporate Sustainability Reporting, in *The Oxford Handbook of Corporate Social Responsibility: Psychological and Organizational Perspectives* (Oxford University Press).

Krueger, Philipp, Zacharias Sautner, Dragon Yongjun Tang, and Rui Zhong, 2021, The effects of mandatory ESG disclosure around the world, *European Corporate Governance Institute–Finance Working Paper* 21–44.

Leuz, Christian, 2000, The development of voluntary cash flow statements in germany and the influence of international reporting standards, *Schmalenbach Business Review* 52, 182–207.

Leuz, Christian, 2018, Evidence-based policymaking: promise, challenges and opportunities for accounting and financial markets research, *Accounting and Business Research* 48, 582–608.

Leuz, Christian, and Peter D Wysocki, 2016, The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research, *Journal of Accounting Research* 54, 525–622.

Li, Feng, 2008, Annual report readability, current earnings, and earnings persistence, *Journal of Accounting and Economics* 45, 221–247.

Li, Kai, Feng Mai, Rui Shen, and Xinyan Yan, 2021, Measuring corporate culture using machine learning, *Review of Financial Studies* 34, 3265–3315.

Lin, Yan, Rui Shen, Jasmine Wang, and Y. Julia Yu, 2021, Global evidence on mandating ESG disclosure in annual reports, Working Paper.

Loughran, Tim, and Bill McDonald, 2016, Textual analysis in accounting and finance: A survey, *Journal of Accounting Research* 54, 1187–1230.

Manetti, Giacomo, and Lucia Becatti, 2009, Assurance services for sustainability reports: Standards and empirical evidence, *Journal of Business Ethics* 87, 289–298.

Oster, Emily, 2019, Unobservable selection and coefficient stability: Theory and evidence, *Journal of Business & Economic Statistics* 37, 187–204.

Perego, Paolo, and Ans Kolk, 2012, Multinationals' accountability on sustainability: The evolution of third-party assurance of sustainability reports, *Journal of Business Ethics* 110, 173–190.

SASB, 2017, The state of disclosure 2017: An analysis of the effectiveness of sustainability disclosure in sec filings.

SEC, 2021, SEC announces enforcement task force focused on climate and ESG issues, (Available at: https://www.sec.gov/news/press-release/2021-42).

Serafeim, George, 2013, The role of the corporation in society: An alternative view and opportunities for future research, Available at SSRN 2270579.

Serafeim, George, 2022, *Purpose and Profit: How Business Can Lift Up the World* (HarperCollins Leadership).

Serafeim, George, and Aaron Yoon, 2022, Stock price reactions to ESG news: The role of ESG ratings and disagreement, *Review of Accounting Studies* 1–31.

Simnett, Roger, Ann Vanstraelen, and Wai Fong Chua, 2009, Assurance on sustainability reports: An international comparison, *The Accounting Review* 84, 937–967.

Sun, Liyang, and Sarah Abraham, 2021, Estimating dynamic treatment effects in event studies with heterogeneous treatment effects, *Journal of Econometrics* 225, 175–199.

Tan, Hongping, Shiheng Wang, and Michael Welker, 2011, Analyst following and forecast accuracy after mandated ifrs adoptions, *Journal of Accounting Research* 49, 1307–1357.

Yip, Rita WY, and Danqing Young, 2012, Does mandatory ifrs adoption improve information comparability?, *The Accounting Review* 87, 1767–1789.

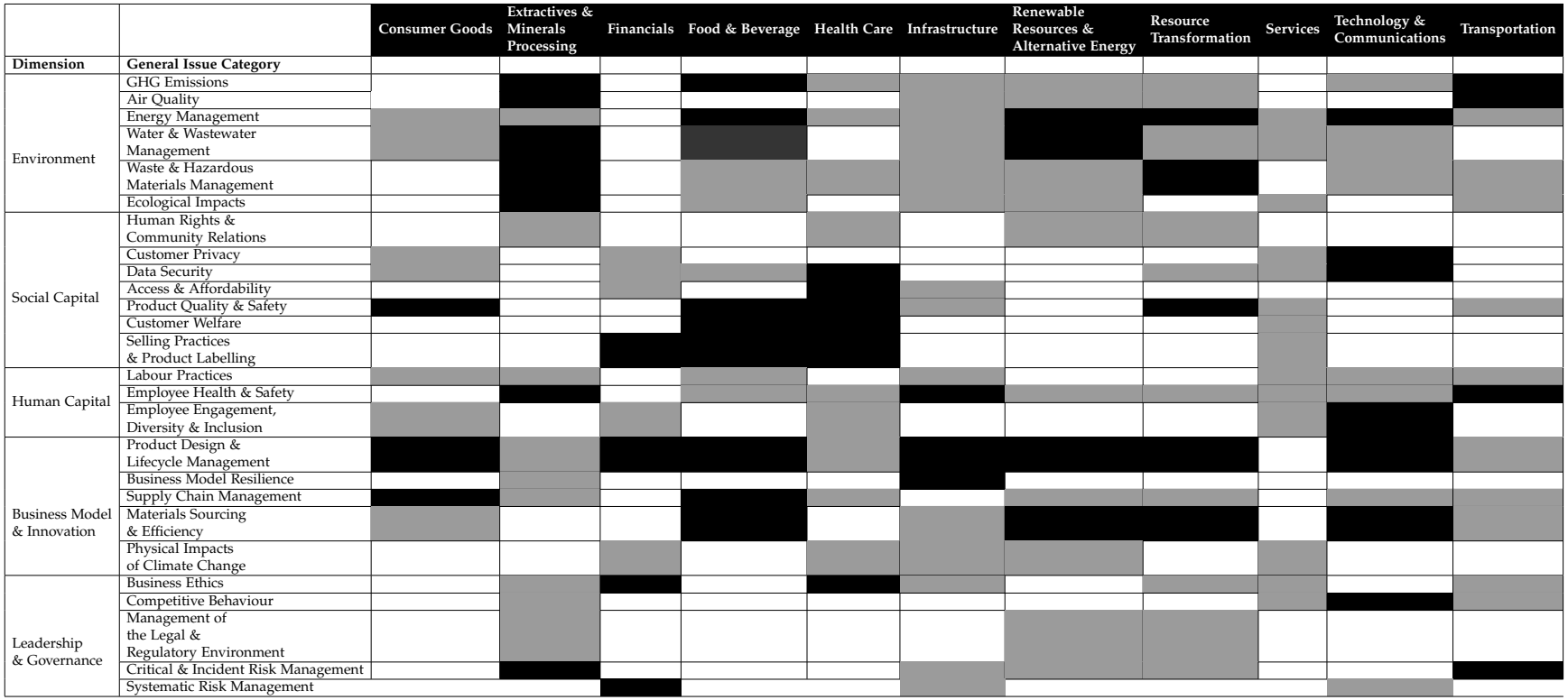


FIGURE 1: SASB MATERIALITY MAP

Figure 1 reports at the sector level the level of materiality for all topics examined in our study. The map is provided by SASB. Dark (light) grey color means that the issue is most (somewhat) material for the sector. White means that the issue is not material for the sector.

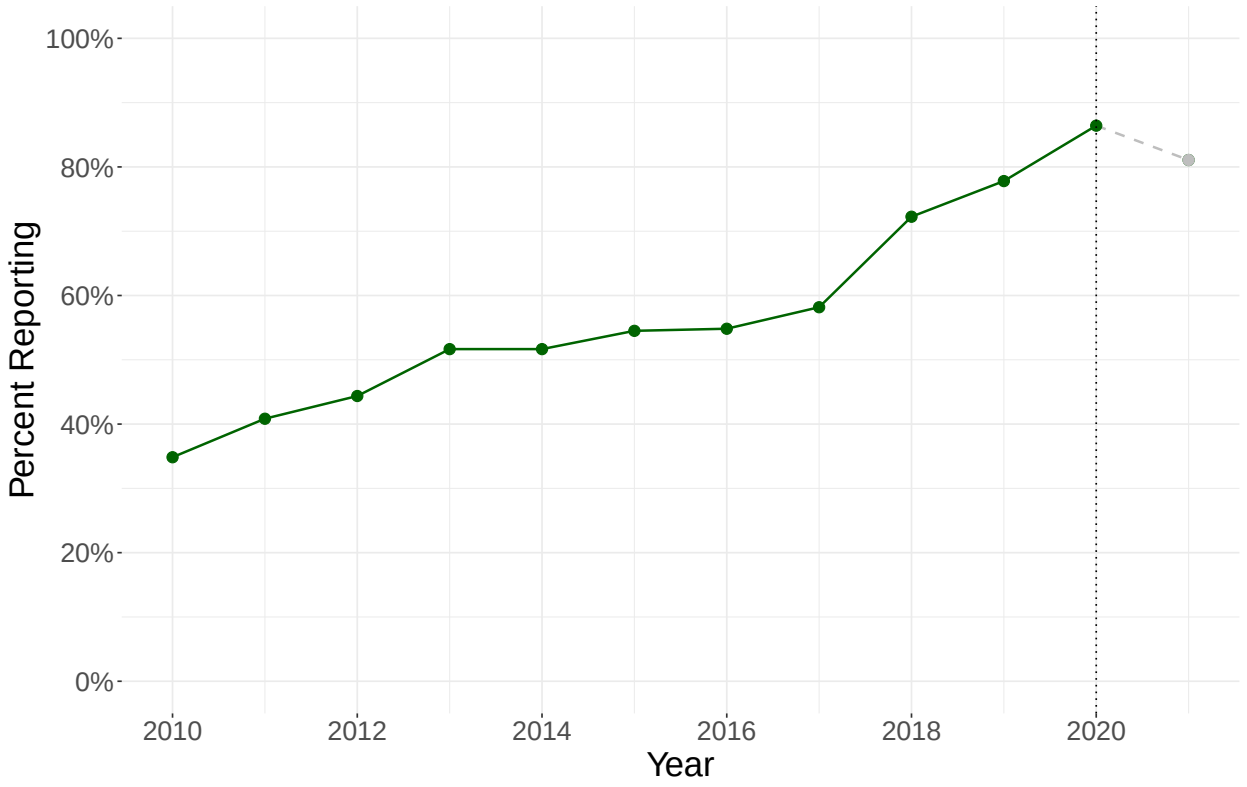


FIGURE 2: PERCENT OF S&P 500 FIRMS PUBLISHING ESG REPORTS

Figure 2 presents, by year, the percent of firms in our sample that released an ESG report. The sample consists of 613 unique firms that were included at least once in the S&P 500 from 2010-2021.

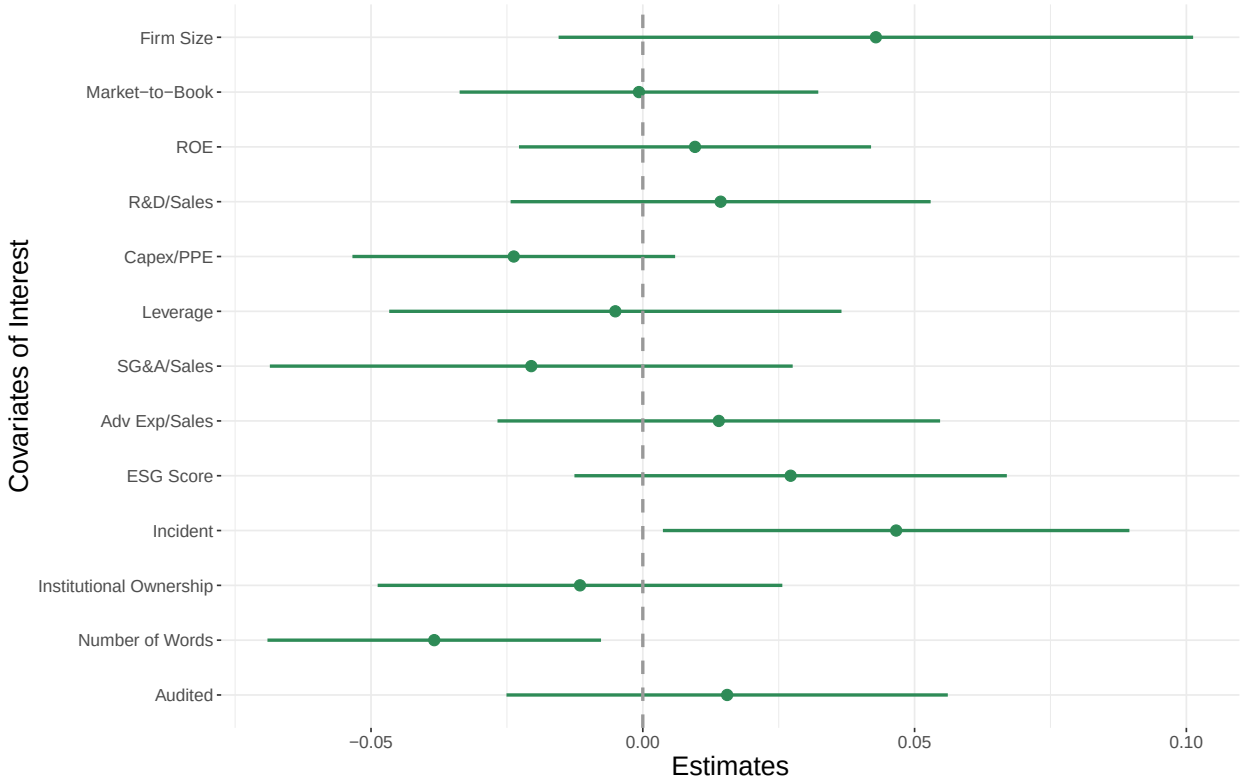


FIGURE 3: DETERMINANTS OF PUBLISHING AN ESG REPORT

Figure 3 presents the coefficient estimates from the hazard model described in Equation 1, which examines the determinants of issuing ESG reports. Firm Size is the natural logarithm of the market value of equity. Market-to-Book is the market value of equity divided by the book value of equity. ROE is net income over average shareholder equity. R&D/Sales is R&D expense divided by sales. Capex/PP&E is capital expenditure divided by property, plant, and equipment. Leverage is the long-term debt to total assets ratio. SG&A/Sales is selling, general, and administrative expense divided by sales. Adv Exp/Sales is advertising expense divided by sales. ESG Score is the ESG rating of the firm by the Thomson Reuters database. Incidents counts the number of sustainability incidents within a given year for a firm. Institutional Ownership is the percentage of shares held by institutional investors. Number of Words is the count of words in the 10-K, using data from the SRAF. Assurance is an indicator equal to 1 if the firm's ESG report received outside assurance, as identified by Refinitiv. All variables are normalized to have a mean of 0 and a standard deviation of 1 to allow for comparability.

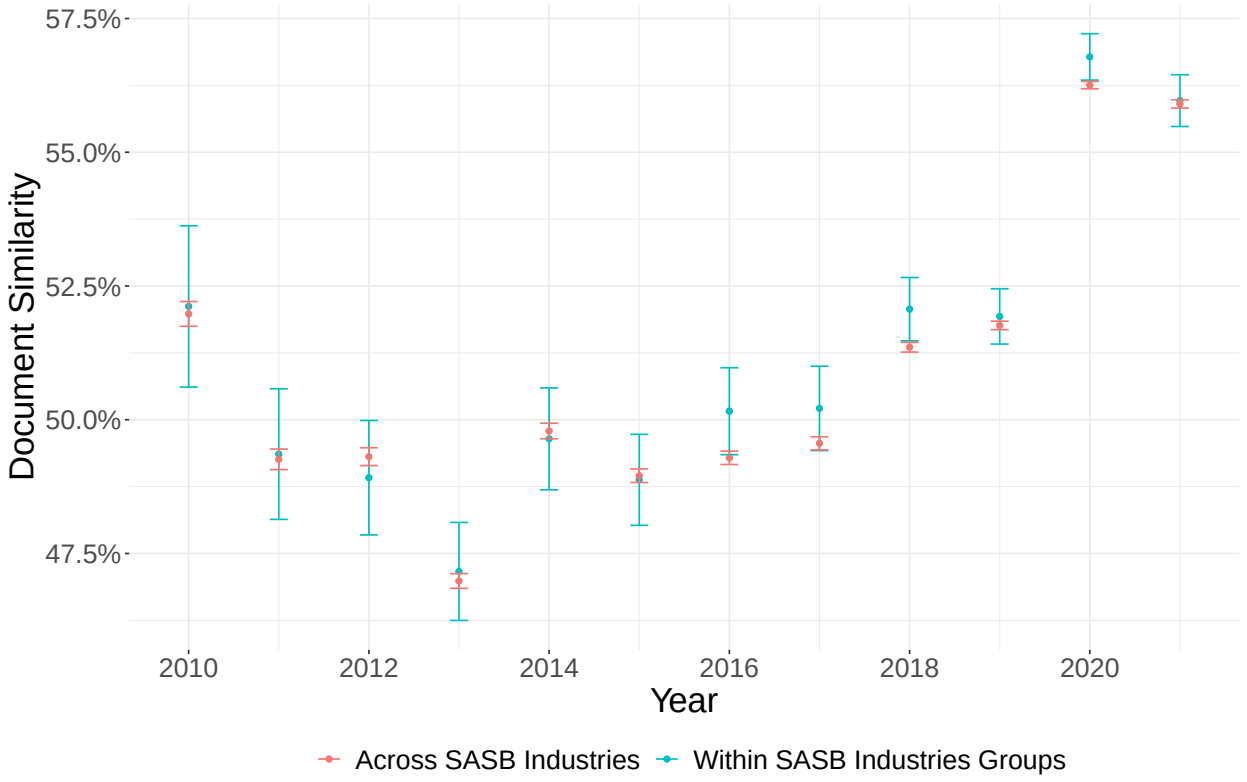
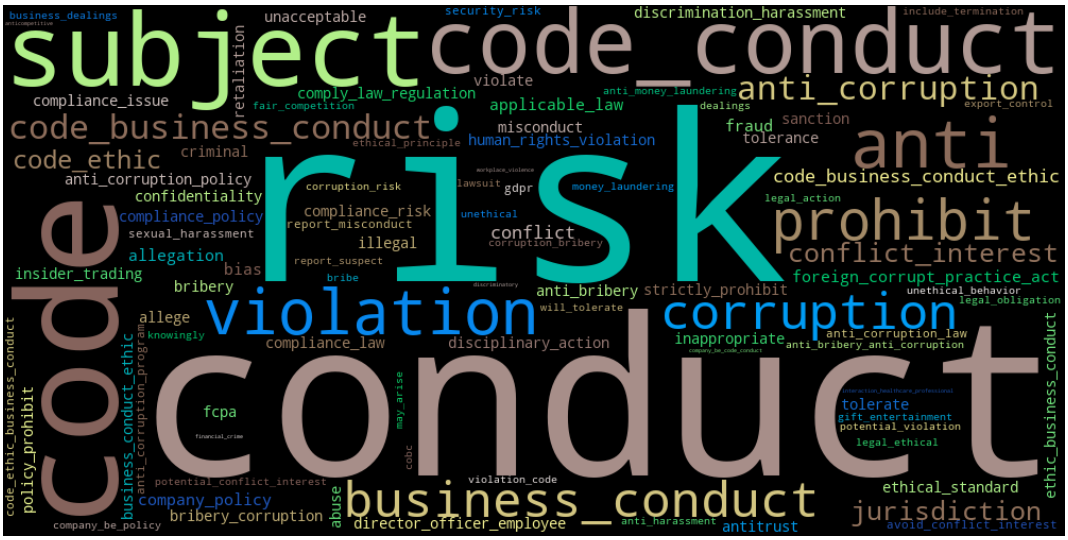
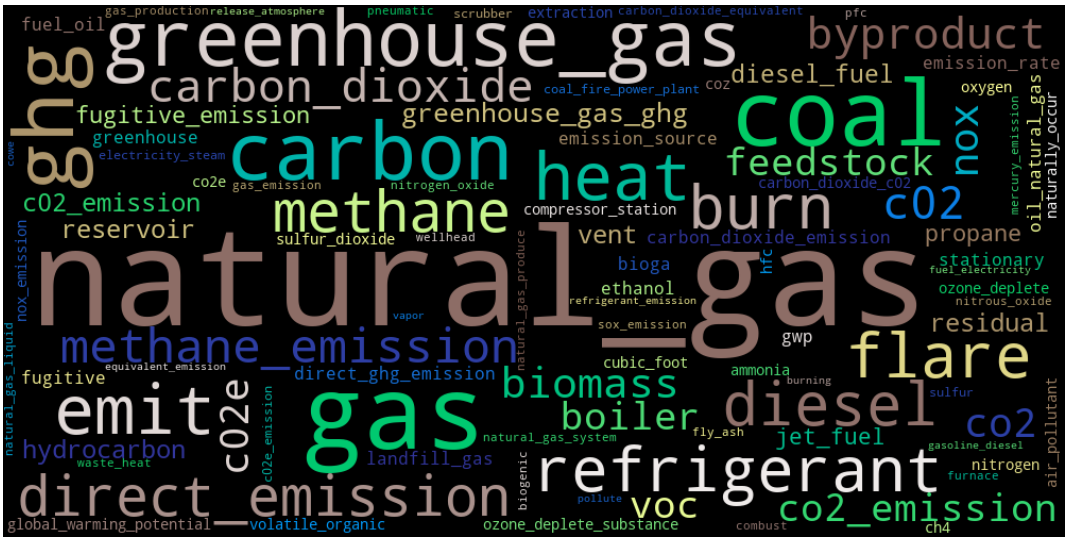


FIGURE 4: CONVERGENCE AT THE DOCUMENT LEVEL

This figure plots histograms of textual cosine similarity of ESG reports within and across sectors. The blue (red) line measures the cosine similarity for all pairwise combinations of firms in the same year and in the same sector (the same year).



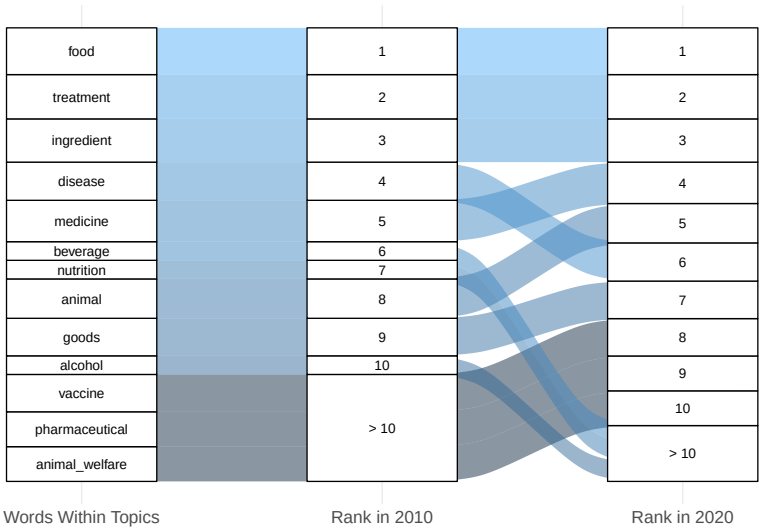
(a) Ethical Word Cloud



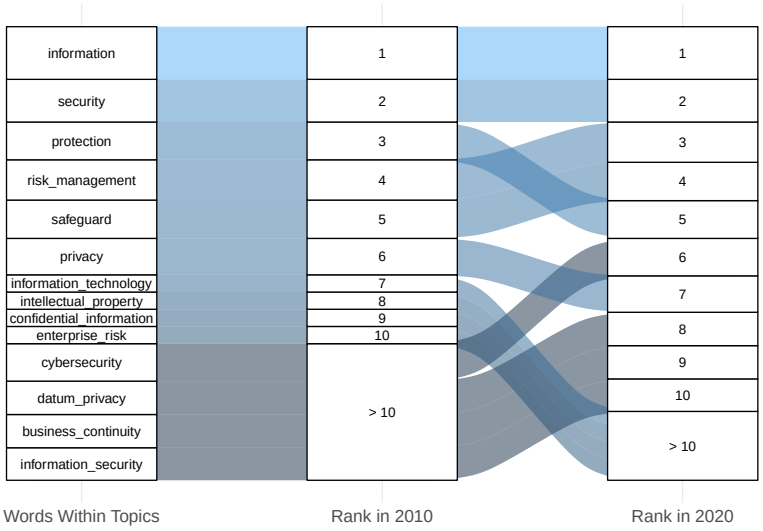
(b) Greenhouse Gas Word Cloud

FIGURE 5: SAMPLE WORD CLOUDS FOR TWO TOPICS

Figure 5 provides a visual representation of the frequency of words detected by our neural network model for two topics, Business Ethics and GHG Emissions. These are the words the algorithm identified as related to the topic. The word's relative size represents the word's frequency within a given topic. Panel A presents the word cloud related to Business Ethics. Panel B presents the word cloud related to GHG Emissions. They are among the 26 ESG topics identified by SASB. Word clouds for all other topics are reported in Figure IA1.



(a) Customer Welfare



(b) Customer Privacy

FIGURE 6: CHANGES IN LANGUAGE OVER TIME

Figure 6 provides a visual representation of the importance of words, as measured by the term frequency, within a topic in 2010 and 2020 reports. Panel A presents the Customer Welfare topic while Panel B presents the Customer Privacy topic. They are among the 26 ESG topics identified by SASB. The ranking is based on their relative term-frequency, in a given year. The blue color bands represent the relative importance of each word within a given topic. Alluvial plots for each topics is reported in Figure IA2.

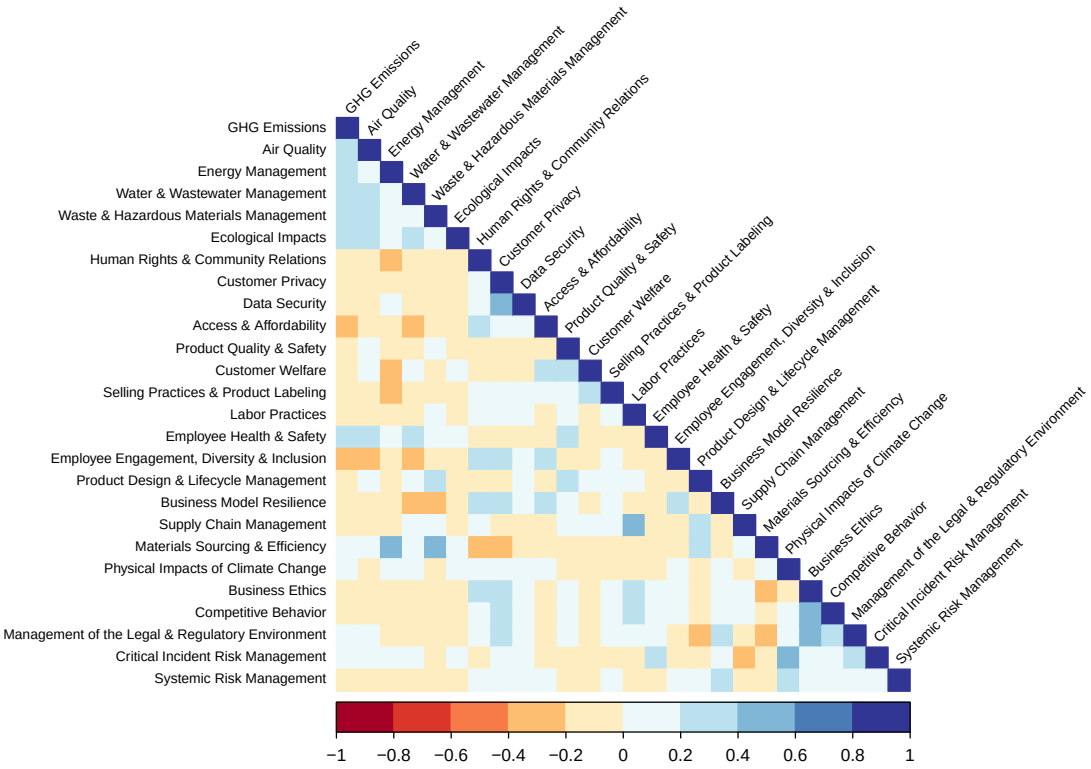


FIGURE 7: CORRELATIONS AMONG ESG TOPICS

Figure 7 presents the correlations of the variable TF-IDF among the 26 ESG topics identified by SASB. TF-IDF is the term frequency-inverse document frequency of all words within a topic. Darker shades of blue (red) represent a more positive (negative) correlation, with yellow representing little or no correlation.

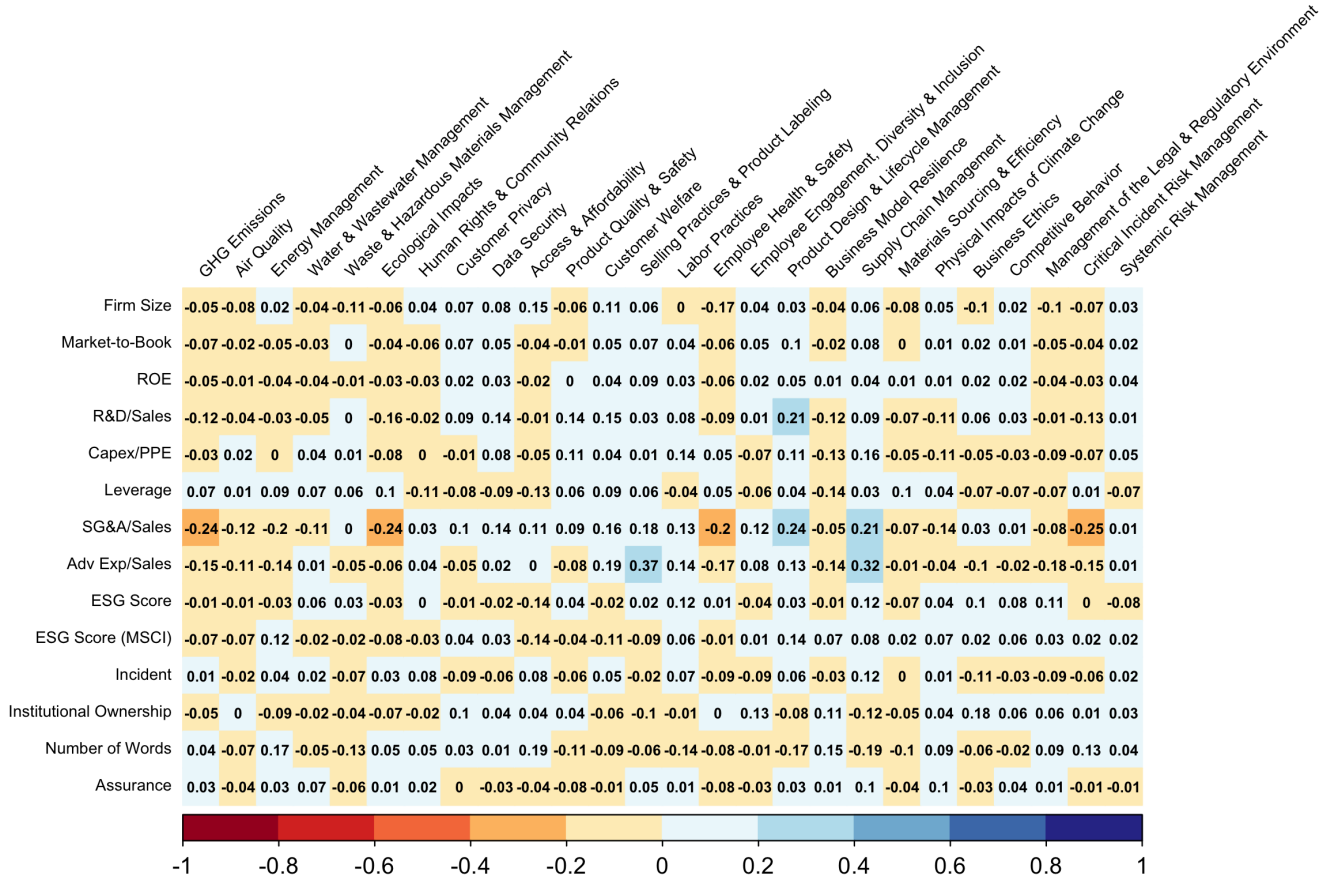
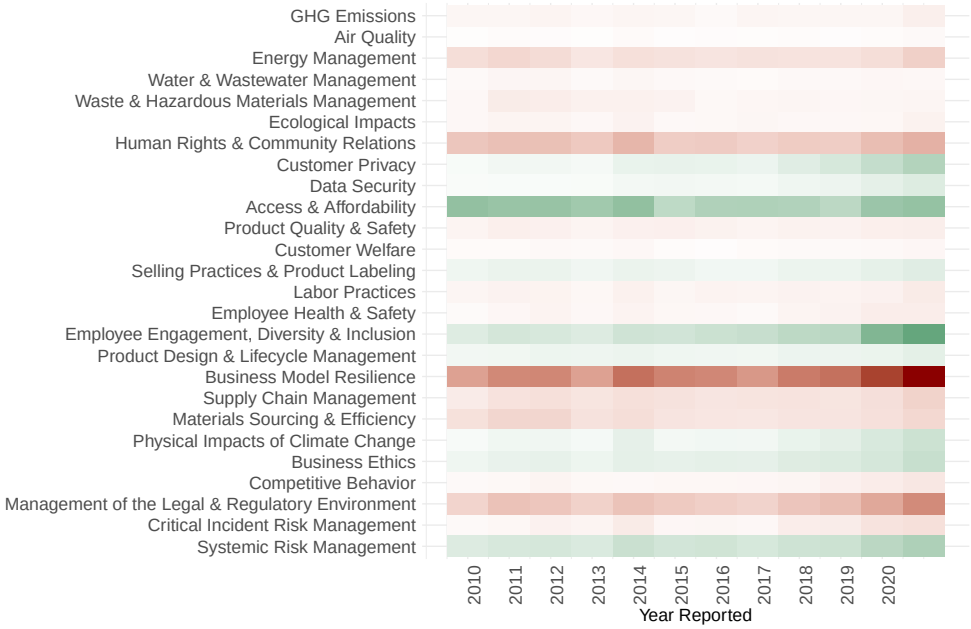
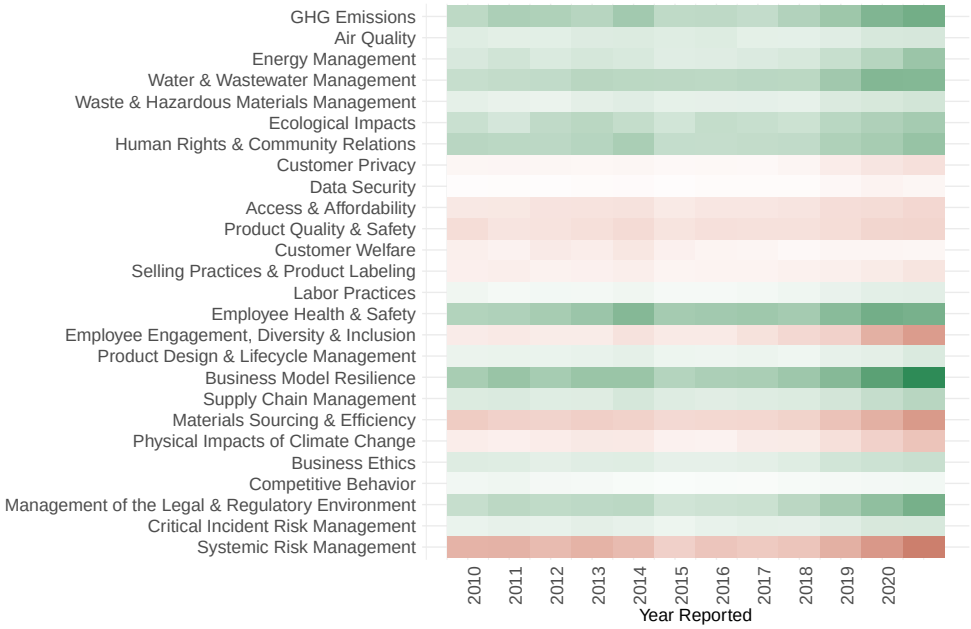


FIGURE 8: CORRELATIONS AMONG TOPICS AND FIRM VARIABLES

Figure 8 presents the correlations among TF-IDF for each of the 26 ESG topics identified by SASB and the firm level covariates. TF-IDF is the term frequency-inverse document frequency of all words within a topic. Firm Size is the natural logarithm of the market value of equity. Market-to-Book is the market value of equity divided by the book value of equity. ROE is net income over average shareholder equity. R&D/Sales is R&D expense divided by sales. Capex/PP&E is capital expenditure divided by property, plant, and equipment. Leverage is the long-term debt to total assets ratio. SG&A/Sales is selling, general, and administrative expense divided by sales. Adv Exp/Sales is advertising expense divided by sales. Incidents counts the number of sustainability incidents within a given year for a firm. ESG Score is the ESG rating of the firm by the Refinitiv database. Institutional Ownership is the percentage of shares held by institutional investors. Number of Words is the count of words in the 10-K, using data from the SRAF. Assurance is an indicator equal to 1 if the firm's ESG report received outside assurance, as identified by Refinitiv. Darker shades of blue (red) represent a more positive (negative) correlation, with yellow representing little or no correlation.



(a) Financial



(b) Extractives & Minerals Processing

FIGURE 9: EVOLUTION OF TOPICS BY SECTOR

Figure 9 displays sector-level heatmaps of the relative amount of content in ESG reports for each of the 26 ESG topics identified by SASB. Panel A plots the average level of the content discussed in ESG reports in the Financial sector, while Panel B plots the same for the Extractives & Minerals Processing sector. Topics with green (red) bars are material (immaterial) for the sector. Darker (lighter) shades represent topics that are discussed more (less) within ESG reports. White bars represent topics that have relatively little content discussed in the documents, on average. Heatmaps for the other 24 topics are reported in Figure IA3.

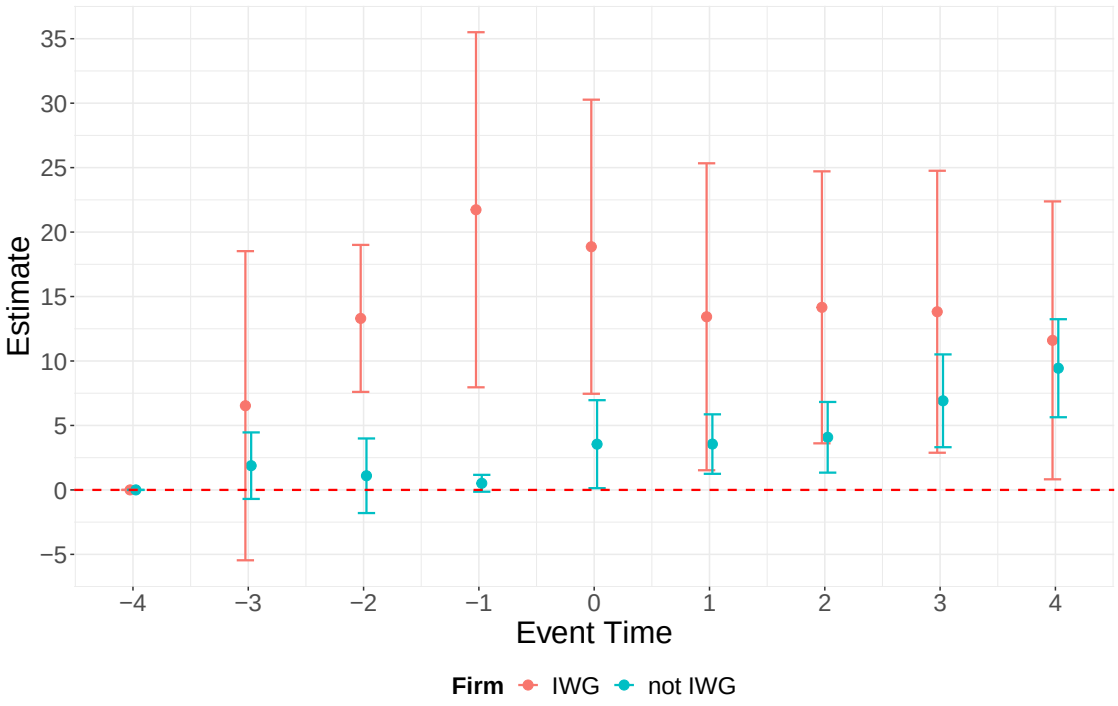


FIGURE 10: COMPARISON OF MATERIAL DISCLOSURES OF IWG AND NON-IWG FIRMS AROUND THE RELEASE OF SASB STANDARDS

Figure 10 plots the dynamic coefficients when regressing TF-IDF on the interaction term between All Material Topics and an indicator for the year relative to the publication of the SASB standards, from t-4 to t+4. TF-IDF is the term frequency-inverse document frequency of all words within a topic. All Material Topics is an indicator equal to 1 if the topic being examined is material to the firm's sector, and 0 otherwise. Red (blue) whisker plots are the coefficients for the sub-sample of firms that participated (did not participate) in IWGs.

TABLE 1: AGGREGATE TRENDS IN ESG REPORTS OVER TIME

Table 1 presents aggregate trends in ESG reports over time in total and by sector. Panel A presents the number of ESG reports issued by all firms in the sample by sector and year. Panel B presents the average number of words per document by sector and year. CG denotes the Consumer Goods sector; EMP denotes the Extractives & Minerals Processing sector; FB denotes the Food & Beverage sector; FIN denotes the Financials sector; HC denotes the Health Care sector; INF denotes the Infrastructure sector; RR denotes the Renewable Resources & Alternative Energy sector; RT denotes the Resource Transformation sector; SERV denotes the Services sector; TECH denotes the Technology & Communications sector; and TRANS denotes the Transportation sector.

Panel A: Number of ESG Reports

YEAR	CG	EMP	FB	FIN	HC	INF	RR	RT	SERV	TECH	TRANS	Total
2010	16	14	20	18	15	12	4	16	5	26	13	159
2011	16	17	17	21	22	19	4	28	6	28	13	191
2012	15	22	19	23	27	23	5	25	10	32	13	214
2013	22	21	23	28	29	29	8	22	15	38	19	254
2014	20	21	21	21	32	28	7	30	18	40	17	255
2015	22	28	21	27	33	35	11	28	17	36	15	273
2016	24	31	23	31	29	34	6	33	15	40	16	282
2017	26	30	25	35	33	37	8	32	15	39	20	300
2018	33	37	26	49	41	49	10	46	20	53	21	385
2019	34	37	29	59	44	51	8	43	26	66	27	424
2020	42	39	28	62	55	56	9	50	31	72	28	472
2021	34	37	26	59	55	51	7	52	32	74	24	451
Total	304	334	278	433	415	424	87	405	210	544	226	3660

Panel B: Mean Number of Words Per ESG Report

YEAR	CG	EMP	FB	FIN	HC	INF	RR	RT	SERV	TECH	TRANS	Mean
2010	8159	7898	5610	5950	4521	10522	5596	6187	7082	7785	11453	7312
2011	8557	8957	5533	7369	5908	10076	9188	5539	6356	7524	17366	7989
2012	9128	7759	7682	7268	9353	7827	10889	7456	11563	9431	12882	8781
2013	6807	8400	8692	5918	8465	6206	6624	5734	8288	8266	13961	7874
2014	6404	8714	10208	8121	8833	8385	8212	7130	7071	9010	10169	8413
2015	6558	6827	7983	6284	7623	8821	10868	7096	5564	7468	14663	7819
2016	6702	7291	7986	6079	8205	8204	9210	6807	6429	7398	8221	7375
2017	6919	7418	6632	5960	7192	7215	7849	5802	5123	8113	9651	7053
2018	8008	7913	9546	6909	7793	8165	8859	6877	6108	7190	8038	7637
2019	8509	10162	9572	7007	8104	7579	9506	8157	6398	7362	7432	7969
2020	9827	12275	10005	8946	9159	9572	9036	9719	7427	7662	10005	9312
2021	10553	14709	9120	10886	9184	10317	8177	12249	8067	9630	8870	10403
Mean	8198	9354	8369	7577	8169	8528	8791	7851	7074	8091	10563	8319

TABLE 2: SUMMARY STATISTICS OF FIRM-LEVEL VARIABLES

Table 2 presents the summary statistics of the firm level variables used in our analysis. Firm Size is the natural logarithm of the market value of equity. Market-to-Book is the market value of equity divided by the book value of equity. ROE is net income over average shareholder equity. R&D/Sales is R&D expense divided by sales. Capex/PP&E is capital expenditure divided by property, plant, and equipment. Leverage is the long-term debt to total assets ratio. SG&A/Sales is selling, general, and administrative expense divided by sales. Adv Exp/Sales is advertising expense divided by sales. ESG Score is the ESG rating of the firm by the Refinitiv database. Incidents counts the number of sustainability incidents within a given year for a firm. Institutional Ownership is the percentage of shares held by institutional investors. Number of Words is the count of words in the 10-K, using data from the SRAF. Assurance is an indicator equal to 1 if the firm's ESG report received outside assurance, as identified by Refinitiv.

Variable	N	Mean	Std. Dev.	Pctl. 25	Median	Pctl. 75
Firm Size	3660	10.06	1.17	9.3	9.96	10.76
Market-to-Book	3660	4.38	10.14	1.38	2.55	4.82
ROE	3660	0.17	0.58	0.07	0.14	0.25
R&D/Sales	3660	0.03	0.06	0	0	0.03
Capex/PPE	3660	0.08	0.05	0.05	0.07	0.1
Leverage	3660	0.31	0.17	0.19	0.3	0.42
SG&A/Sales	3660	0.19	0.16	0.05	0.17	0.29
Adv Exp/Sales	3660	0.01	0.02	0	0	0.02
ESG Score	3660	56.19	13.48	46.71	56.47	65.5
Incidents	3660	11.31	27	0	2	10
Institutional Ownership	3660	0.8	0.14	0.72	0.81	0.89
Number of Words	3660	82257.23	60863.44	51482.75	68077	92280.75
Assurance	3660	0.31	0.46	0	0	1

TABLE 3: MOST IMPORTANT WORDS IDENTIFIED BY ALGORITHM

Table 3 presents the five most important words or phrases identified by the neural network model for each of the 26 ESG topics as defined by SASB. The algorithm was seeded at the topic level with unique words included in the SASB standards and then identified related phrases and words. It was then reseeded using this more comprehensive dictionary, which was used to calculate TF-IDF. Appendix Table IA.2 reports the 50 most important words identified by the algorithm for each topic.

Topics	**Top Five Words**
Environment	
GHG Emissions	gas, carbon_dioxide, c02, natural_gas, potent_greenhouse_gas
Air Quality	pollution, pollutant, ppm, contaminant, particulate_matter
Energy Management	grid, energy_use, energy_consumption, utility, energy
Water & Wastewater Management	wastewater, fresh_water, water, freshwater_source, groundwater
Waste & Hazardous Materials Management	recycling, hazardous_waste, nonhazardous, non_hazardous, waste
Ecological Impacts	biodiversity, topsoil, natural_habitat, wetland_habitat, wetland
Social Capital	
Human Rights & Community Relations	political, rights, social, labor, law
Customer Privacy	cyber_security, datum_security, information_security, datum_privacy, datum_protection
Data Security	it_infrastructure, phishing, encrypted, access_management, cyber_intrusion
Access & Affordability	affordability, underserve, underserved_community, affordable, low_moderate_income_community
Product Quality & Safety	quality, safety_quality, quality_safety, product_performance, product_testing
Customer Welfare	nutrition, counterfeit, antimicrobial, medicine, pathogen
Selling Practices & Product Labeling	labeling, marketing, transparency, marketing_practice, marketing_communication
Human Capital	
Labor Practices	freedom_association, child_force_labor, force_child_labor, force_labor, labor_child_labor
Employee Health & Safety	accident, safety, safety_incident, incident, injury_illness
Employee Engagement, Diversity & Inclusion	diversity, gender, racial, inclusion_diversity, discrimination
Business Model & Innovation	
Product Design & Lifecycle Management	innovation, product_packaging, packaging_design, packaging_solution, product
Business Model Resilience	responsiveness, planning, execution, approach, strategy
Supply Chain Management	supplier, supply_base, throughout_supply_chain, supply_chain_partner, chain
Materials Sourcing & Efficiency	recycle, efficiency, renewable_energy, recycled, renewable_resource
Physical Impacts of Climate Change	climate_change, rise_sea_level, change_climate, sea_level_rise, extreme_weather_event
Leadership & Governance	
Business Ethics	bribery, bribery_corruption, corruption_bribery, money_laundering, financial_crime
Competitive Behavior	bargaining_power, price_fix, embezzlement, corruption_extortion, payment_kickback
Management of the Legal & Regulatory Environment	compliance, regulatory_environment, regulation, law_regulation, legislative_regulatory
Critical Incident Risk Management	catastrophe, storm_damage, such_as_hurricane, landslide, natural_disaster_such_as_hurricane
Systemic Risk Management	complex, complicated, highly_complex, uncertainty, scale_complexity

TABLE 4: TOPIC SUMMARY STATISTICS

Table 4 reports descriptive statistics for TF-IDF for each of the 26 ESG topics identified by SASB. TF-IDF is the term frequency-inverse document frequency of all words within a topic.

Variable	N	Mean	Std. Dev.	Pctl. 25	Median	Pctl. 75
Environment						
GHG Emissions	3660	48.38	83.78	5.03	18.45	53.82
Air Quality	3660	26.02	42.28	3.37	11.66	32.98
Energy Management	3660	96.37	145.08	24.67	53.59	103.72
Water & Wastewater Management	3660	52.2	83.13	5.77	21.47	65.4
Waste & Hazardous Materials Management	3660	46.7	59.25	10.9	28.12	60.45
Ecological Impacts	3660	43.44	81.41	4.12	13.8	46.13
Social Capital						
Human Rights & Community Relations	3660	80.66	85.37	25.81	56.38	106.88
Customer Privacy	3660	30.37	45.01	2.14	12.1	41.23
Data Security	3660	19.21	31.28	0	7.73	23.89
Access & Affordability	3660	80.43	118.29	18.62	42.56	91.66
Product Quality & Safety	3660	46.96	56.98	12.38	29.17	61.55
Customer Welfare	3660	46.5	124.52	2.49	9.54	30.64
Selling Practices & Product Labeling	3660	37.55	54.46	8.49	22.24	47.36
Human Capital						
Labor Practices	3660	38	63.9	3.56	16.55	45.59
Employee Health & Safety	3660	72.65	91.85	12.57	39.7	97.04
Employee Engagement, Diversity & Inclusion	3660	87.51	95.72	22.24	57.18	119.91
Business Model & Innovation						
Product Design & Lifecycle Management	3660	71.79	98.31	17.12	40.11	85.76
Business Model Resilience	3660	157.28	150.33	57.87	117.47	207.61
Supply Chain Management	3660	101.91	164.91	21.4	53.55	115.27
Materials Sourcing & Efficiency	3660	79.7	90.03	28.4	55.63	100.65
Physical Impacts of Climate Change	3660	33.05	47.13	5.91	17.6	40.97
Leadership & Governance						
Business Ethics	3660	47.48	51.49	10.25	33.08	67.06
Competitive Behavior	3660	14.02	25.2	0	6.08	18.56
Management of the Legal & Regulatory Environment	3660	88.42	94.5	24.81	62.06	116.99
Critical Incident Risk Management	3660	26.72	40.14	4.1	13.53	32.99
Systemic Risk Management	3660	92.58	98.28	32.14	67.53	118.43

TABLE 5: TOP FIVE MATERIAL AND IMMATERIAL TOPICS BY SECTOR

Table 5 reports the top five material and immaterial topics in ESG reports for each of the 11 SASB sectors. The rankings are based on the percent of words within the average ESG report devoted to that topic, which is reported in parentheses. The topics are from the 26 ESG topics identified by SASB.

Panel A: Consumer Goods

	Material	*Immaterial*
1.	Supply Chain Management (13.5%)	Business Model Resilience (8.11%)
2.	Product Design & Lifecycle Management (7.12%)	Systemic Risk Management (5.72%)
3.	Employee Engagement, Diversity & Inclusion (6.72%)	Human Rights & Community Relations (5.16%)
4.	Materials Sourcing & Efficiency (6.66%)	Management of the Legal & Regulatory Environment (4.43%)
5.	Labor Practices (4.59%)	Waste & Hazardous Materials Management (3.42%)

Panel B: Extractives & Minerals Processing

	Material	*Not-Material*
1.	Business Model Resilience (9.98%)	Systemic Risk Management (5.81%)
2.	Employee Health & Safety (9.3%)	Materials Sourcing & Efficiency (3.92%)
3.	GHG Emissions (7.16%)	Employee Engagement, Diversity & Inclusion (3.1%)
4.	Water & Wastewater Management (7.11%)	Product Quality & Safety (2.56%)
5.	Management of the Legal & Regulatory Environment (7.04%)	Physical Impacts of Climate Change (2.02%)

Panel C: Financials

	Material	*Not-Material*
1.	Access & Affordability (11.28%)	Business Model Resilience (15.8%)
2.	Employee Engagement, Diversity & Inclusion (9.67%)	Management of the Legal & Regulatory Environment (7.15%)
3.	Systemic Risk Management (7.09%)	Human Rights & Community Relations (7%)
4.	Customer Privacy (4.86%)	Materials Sourcing & Efficiency (3.37%)
5.	Business Ethics (3.85%)	Energy Management (3.36%)

Panel D: Food & Beverage

	Material	*Not-Material*
1.	Supply Chain Management (13.52%)	Systemic Risk Management (5.31%)
2.	Customer Welfare (10.93%)	Human Rights & Community Relations (4.46%)
3.	Business Model Resilience (7.42%)	Employee Engagement, Diversity & Inclusion (3.89%)
4.	Product Design & Lifecycle Management (5.82%)	Management of the Legal & Regulatory Environment (3.71%)
5.	Materials Sourcing & Efficiency (5.71%)	Access & Affordability (3.23%)

Panel E: Health Care

	Material	*Not-Material*
1.	Customer Welfare (12.95%)	Business Model Resilience (8.62%)
2.	Access & Affordability (10.38%)	Systemic Risk Management (5.91%)
3.	Employee Engagement, Diversity & Inclusion (5.49%)	Management of the Legal & Regulatory Environment (5.58%)
4.	Product Quality & Safety (4.82%)	Materials Sourcing & Efficiency (3.14%)
5.	Supply Chain Management (4.77%)	Water & Wastewater Management (1.99%)

continued on next page...

Panel F: Infrastructure

	Material	*Not-Material*
1.	Energy Management (13.49%)	Management of the Legal & Regulatory Environment (6.02%)
2.	Business Model Resilience (10.39%)	Employee Engagement, Diversity & Inclusion (4.48%)
3.	Materials Sourcing & Efficiency (6.47%)	Human Rights & Community Relations (4.32%)
4.	Systemic Risk Management (5.69%)	Supply Chain Management (3.4%)
5.	Employee Health & Safety (4.97%)	Customer Privacy (1.88%)

Panel G: Renewable Resources & Alternative Energy

	Material	*Not-Material*
1.	Energy Management (16.35%)	Business Model Resilience (8.88%)
2.	Management of the Legal & Regulatory Environment (6.75%)	Systemic Risk Management (4.98%)
3.	Employee Health & Safety (6.63%)	Employee Engagement, Diversity & Inclusion (3.31%)
4.	Materials Sourcing & Efficiency (5.8%)	Access & Affordability (2.92%)
5.	Ecological Impacts (5.64%)	Product Quality & Safety (2.7%)

Panel H: Resource Transformation

	Material	*Not-Material*
1.	Product Design & Lifecycle Management (7.5%)	Business Model Resilience (9.32%)
2.	Supply Chain Management (6.3%)	Systemic Risk Management (5.73%)
3.	Materials Sourcing & Efficiency (6.19%)	Employee Engagement, Diversity & Inclusion (4.83%)
4.	Energy Management (5.81%)	Ecological Impacts (2.91%)
5.	Employee Health & Safety (5.54%)	Access & Affordability (2.46%)

Panel I: Services

	Material	*Not-Material*
1.	Employee Engagement, Diversity & Inclusion (9.96%)	Business Model Resilience (11.27%)
2.	Energy Management (6.11%)	Human Rights & Community Relations (7.23%)
3.	Selling Practices & Product Labeling (4.02%)	Systemic Risk Management (6.59%)
4.	Business Ethics (3.09%)	Supply Chain Management (5.73%)
5.	Customer Privacy (2.9%)	Access & Affordability (5.35%)

Panel J: Technology & Communications

	Material	*Not-Material*
1.	Supply Chain Management (9.36%)	Business Model Resilience (9.31%)
2.	Product Design & Lifecycle Management (7.36%)	Human Rights & Community Relations (5.74%)
3.	Employee Engagement, Diversity & Inclusion (6.19%)	Management of the Legal & Regulatory Environment (5.11%)
4.	Systemic Risk Management (5.93%)	Access & Affordability (3.83%)
5.	Energy Management (5.89%)	Business Ethics (3.1%)

Panel K: Services

	Material	*Not-Material*
1.	Materials Sourcing & Efficiency (7.31%)	Business Model Resilience (9.43%)
2.	Employee Health & Safety (6.88%)	Systemic Risk Management (6.83%)
3.	Energy Management (6.86%)	Management of the Legal & Regulatory Environment (5.7%)
4.	Supply Chain Management (5.99%)	Human Rights & Community Relations (5.03%)
5.	GHG Emissions (4.36%)	Employee Engagement, Diversity & Inclusion (4.34%)

TABLE 6: MATERIAL ESG DISCLOSURES

Table 6 documents the the relative amount of material information disclosed in ESG reports, where materiality is defined by the SASB sector-level standards that defined material topics. TF-IDF is the term frequency-inverse document frequency of all words within a topic. All Material Topics equals one if the topic measured by TF-IDF is material, and 0 otherwise. Column 4 includes a sub-sample from four years before to four years after the release of the SASB standards. The control variables included in column 2 are the full set described in Table 2. Standard errors are robust to heteroskedasticity and clustered at the firm and year levels.

	TF-IDF			
	Full	Full	Full	Narrow Window
	(1)	(2)	(3)	(4)
All Material Topics	29.001***	29.001***	29.001***	25.713***
	(1.940)	(1.940)	(1.940)	(1.637)
Firm FE	Yes	Yes	No	No
Year FE	Yes	Yes	No	No
Firm × Year FE	No	No	Yes	Yes
Topic × Year FE	Yes	Yes	Yes	Yes
Controls	No	Financials and Text	Subsumed by FE	Subsumed by FE
Observations	95,160	95,160	95,160	62,556
Adjusted R^2	0.346	0.348	0.455	0.450

TABLE 7: CHANGES IN MATERIAL DISCLOSURES AROUND STANDARDS

Table 7 documents changes in the relative amount of material information disclosed in ESG reports around the release of the SASB sector-level standards that defined material topics. TF-IDF is the term frequency-inverse document frequency of all words within a topic. All Material Topics equals one if the topic measured by TF-IDF is material, and 0 otherwise. Post is an indicator equal to 1 in the years after SASB published its standards for the firm's sector, and 0 otherwise. Firm Size is the natural logarithm of the market value of equity. Market-to-Book is the market value of equity divided by the book value of equity. ROE is net income over average shareholder equity. R&D/Sales is R&D expense divided by sales. Capex/PPE is capital expenditure divided by property, plant, and equipment. Leverage is the long-term debt to total assets ratio. SG&A/Sales is selling, general, and administrative expense divided by sales. Adv Exp/Sales is advertising expense divided by sales. ESG Score is the ESG rating of the firm by the Refinitiv database. Incidents counts the number of sustainability incidents within a given year for a firm. Institutional Ownership is the percentage of shares held by institutional investors. Number of Words is the count of words in the 10-K, using data from the SRAF. Assurance is an indicator equal to 1 if the firm's ESG report received outside assurance, as identified by Refinitiv. Column 6 reports the results for a subsample from four years before to four years after the release of the SASB standards. Standard errors are robust to heteroskedasticity and clustered at the firm and year levels.

	TF-IDF					
	Full	Full	Full	Full	Full	Narrow Window
	(1)	(2)	(3)	(4)	(5)	(6)
All Material Topics x Post	5.141**	5.638***	5.861***	5.860***	6.610***	3.339***
	(2.356)	(1.913)	(1.913)	(1.911)	(1.837)	(1.212)
All Material Topics	25.138***					
	(2.171)					
Firm Size			−1.335	−1.350		
			(2.349)	(2.350)		
Market-to-Book			−0.084	−0.084		
			(0.143)	(0.143)		
ROE			2.148	2.139		
			(3.030)	(3.036)		
R&D/Sales			−2.379**	−2.387**		
			(1.032)	(1.035)		
Capex/PPE			−0.632*	−0.633*		
			(0.342)	(0.343)		
Leverage			−0.163	−0.161		
			(0.135)	(0.135)		
SG&A/Sales			−0.276	−0.274		
			(0.246)	(0.244)		
Adv. Expense/Sales			1.895	1.909		
			(1.959)	(1.945)		
ESG Score			0.261***	0.262***		
			(0.079)	(0.079)		
Incident			−0.154	−0.155		
			(0.191)	(0.191)		
Institutional Ownership			9.628	10.011		
			(15.168)	(15.191)		
Log(Words in 10K)				−1.067		
				(2.016)		
Assurance			7.998***	7.967***		
			(2.775)	(2.783)		
Topic × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Topic × Sector FE	No	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	No	No
Year FE	Yes	Yes	Yes	Yes	No	No
Firm × Year FE	No	No	No	No	Yes	Yes
Controls	No	No	Financial	Financial and Text	Subsumed by FE	Subsumed by FE
Observations	95,160	95,160	95,160	95,160	95,160	62,556
Adjusted R^2	0.344	0.421	0.422	0.422	0.534	0.527

TABLE 8: CHANGE IN MATERIAL DISCLOSURES BY IWG FIRMS AND OTHER EARLY DISCLOSERS

Table 8 documents changes in the relative amount of material information disclosed in ESG reports around the release of the SASB sector-level standards that defined material topics for two subsamples. Panel A reports the results for firms that took part in the standard-setting process through the IWGs. Panel B reports the results for all other firms that disclosed prior to the release of the standards but were not part of IWGs. TF-IDF is the term frequency-inverse document frequency of all words within a topic. All Material Topics equals one if the topic measured by TF-IDF is material, and 0 otherwise. Post is an indicator equal to 1 in the years after SASB published its standards for the firm's sector, and 0 otherwise. Control variables used in columns 3 and 4 are those from Table 2. Column 6 reports the results for a subsample from four years before to four years after the release of the SASB standards. Standard errors are robust to heteroskedasticity and clustered at the firm and year levels.

Panel A: Industry Working Group Members

	TF-IDF					
	Full	Full	Full	Full	Full	Narrow Window
	(1)	(2)	(3)	(4)	(5)	(6)
All Material Topics x Post	2.612	3.232	3.714	3.725	4.783	2.238
	(3.505)	(2.900)	(2.792)	(2.793)	(3.037)	(2.311)
All Material Topics	34.575***					
	(4.365)					
Industry Working Group	Yes	Yes	Yes	Yes	Yes	Yes
Topic × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Topic × Sector FE	No	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	No	No
Year FE	Yes	Yes	Yes	Yes	No	No
Firm × Year FE	No	No	No	No	Yes	Yes
Controls	No	No	Financial	Financial and Text	Subsumed by FE	Subsumed by FE
Observations	29,900	29,900	29,900	29,900	29,900	21,372
Adjusted R^2	0.339	0.432	0.438	0.439	0.557	0.549

Panel B: Other Early Disclosers

	TF-IDF					
	Full	Full	Full	Full	Full	Narrow Window
	(1)	(2)	(3)	(4)	(5)	(6)
All Material Topics x Post	7.548***	7.856***	8.043***	8.054***	8.812***	4.703***
	(2.477)	(2.260)	(2.223)	(2.222)	(2.080)	(1.435)
All Material Topics	19.848***					
	(1.825)					
Industry Working Group	No	No	No	No	No	No
Topic × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Topic × Sector FE	No	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	No	No
Year FE	Yes	Yes	Yes	Yes	No	No
Firm × Year FE	No	No	No	No	Yes	Yes
Controls	No	No	Financial	Financial and Text	Subsumed by FE	Subsumed by FE
Observations	65,260	65,260	65,260	65,260	65,260	41,184
Adjusted R^2	0.355	0.432	0.434	0.434	0.529	0.522

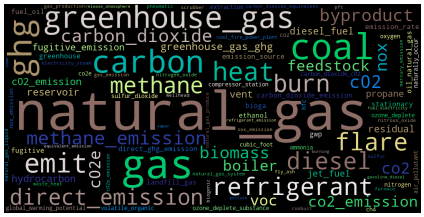
TABLE 9: DYNAMIC ANALYSIS OF CHANGES IN ESG REPORTS FOR IWG FIRMS AND OTHER EARLY DISCLOSERS

Table 9 reports the results of dynamic analysis of the relative amount of material information in ESG reports around the release of the SASB sector-level standards that defined material topics for two subsamples. Column (1) reports the results for firms that took part in the standard-setting process through the IWGs. Column (2) reports the results for all other firms that disclosed prior to the release of the standards but were not part of IWGs. TF-IDF is the term frequency-inverse document frequency of all words within a topic. All Material Topics equals one if the topic measured by TF-IDF is material, as defined by SASB, and 0 otherwise. All Material Topics is interacted with indicators for each of the years around the release of the standards from t-3 to t+4. Standard errors are robust to heteroskedasticity and clustered at the firm and year levels.

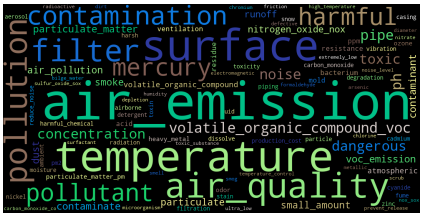
	TF-IDF	
	IWG	Non-IWG
	(1)	(2)
All Material Topics $\times$ $I(t-3)$	6.532	1.880
	(5.382)	(1.158)
All Material Topics $\times$ $I(t-2)$	13.304***	1.098
	(2.559)	(1.298)
All Material Topics $\times$ $I(t-1)$	21.732***	0.514
	(6.181)	(0.294)
All Material Topics $\times$ $I(t)$	18.865***	3.548**
	(5.119)	(1.532)
All Material Topics $\times$ $I(t+1)$	13.431**	3.556***
	(5.343)	(1.036)
All Material Topics $\times$ $I(t+2)$	14.162**	4.081***
	(4.735)	(1.230)
All Material Topics $\times$ $I(t+3)$	13.821**	6.913***
	(4.909)	(1.615)
All Material Topics $\times$ $I(t+4)$	11.602**	9.440***
	(4.837)	(1.706)
Firm $\times$ Year FE	Yes	Yes
Topic $\times$ Sector FE	Yes	Yes
Topic $\times$ Year FE	Yes	Yes
Controls	Subsumed by FE	Subsumed by FE
Observations	21,372	41,184
Adjusted R^2	0.549	0.522

INTERNET APPENDIX
FOR ONLINE PUBLICATION

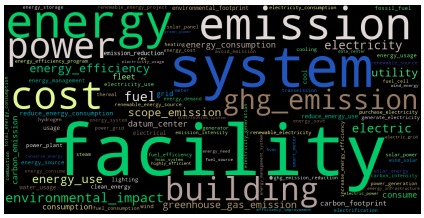
The Evolution of ESG Reports and the Role of Voluntary Standards



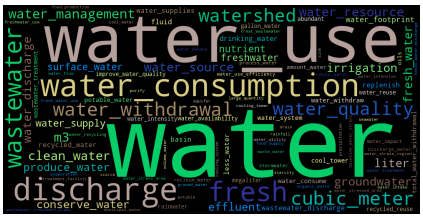
GHG Emissions



Air Quality

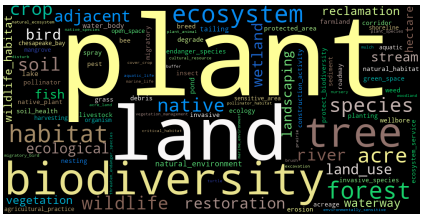


Energy Management



Water & Wastewater Management

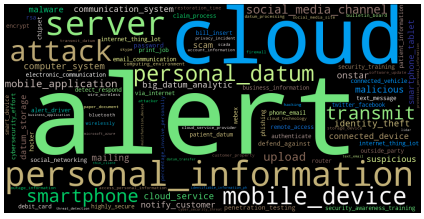
Waste & Hazardous Materials Management



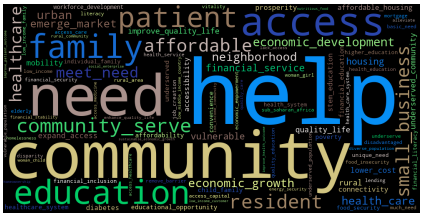
Ecological Impacts

Human Rights & Community Relations

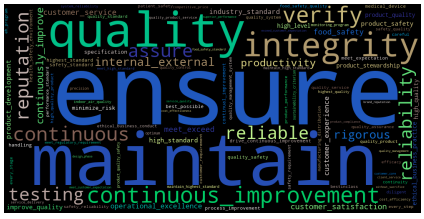
Customer Privacy



Data Security



Access & Affordability



Product Quality & Safety

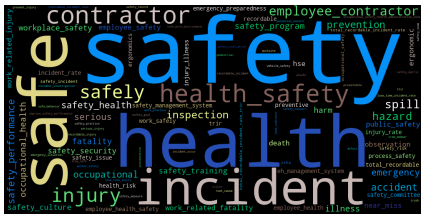
Customer Welfare

FIGURE IA. 1: WORD CLOUDS FOR ADDITIONAL 24 TOPICS

Figure 1 provides a visual representation of the frequency of words detected by our neural network model for the 24 topics not presented in the main body of the paper. Larger words are those that appear more frequently.

Selling Practices & Product Labeling

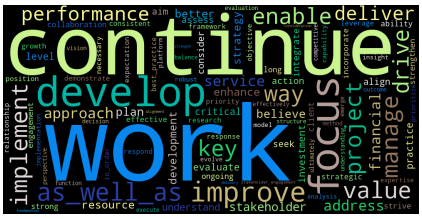
Labor Practices



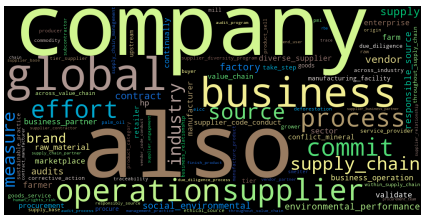
Employee Health & Safety

Employee Engagement, Diversity & Inclusion

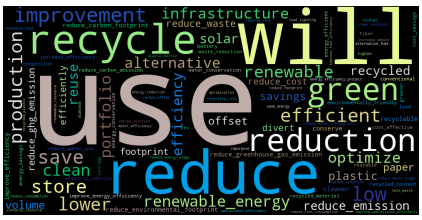
Product Design & Lifecycle Management



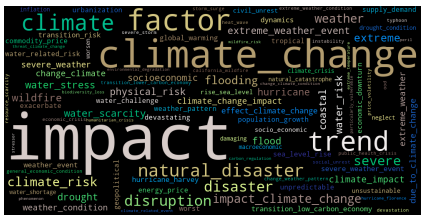
Business Model Resilience



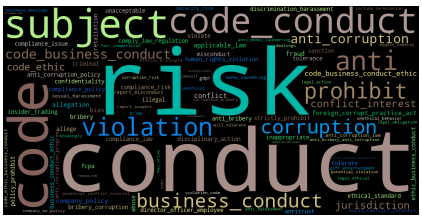
Supply Chain Management



Materials Sourcing & Efficiency



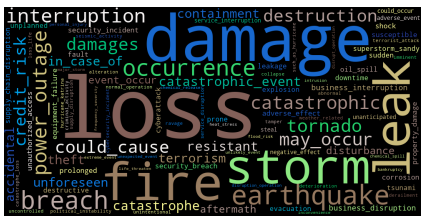
Physical Impacts of Climate Change



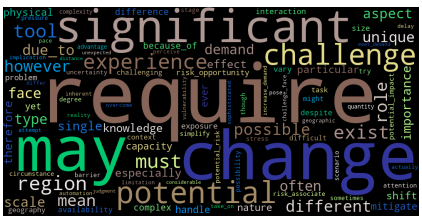
Business Ethics

Competitive Behavior

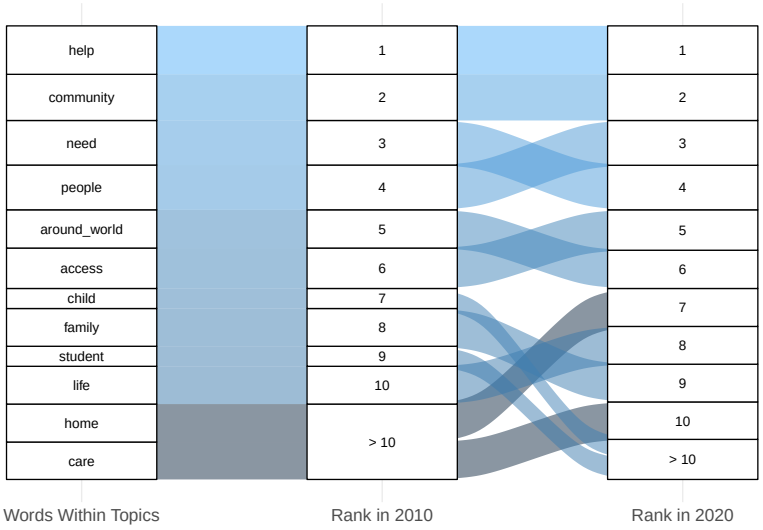
Management of the Legal & Regulatory Environment



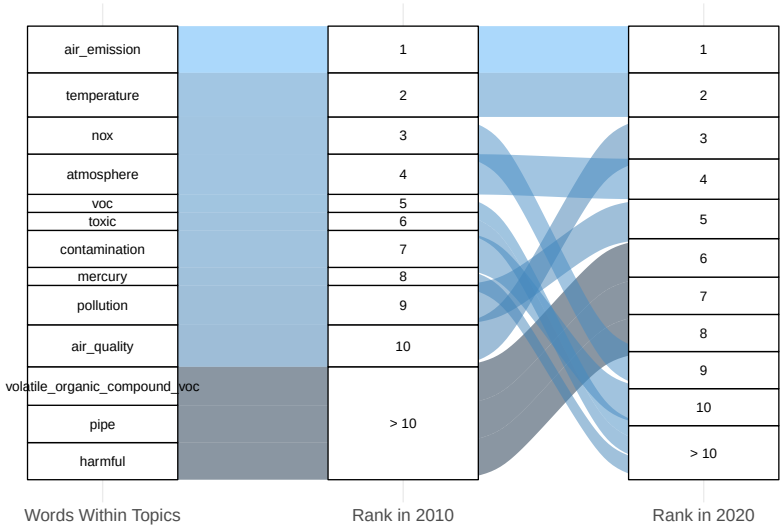
Critical Incident Risk Management



Systemic Risk Management



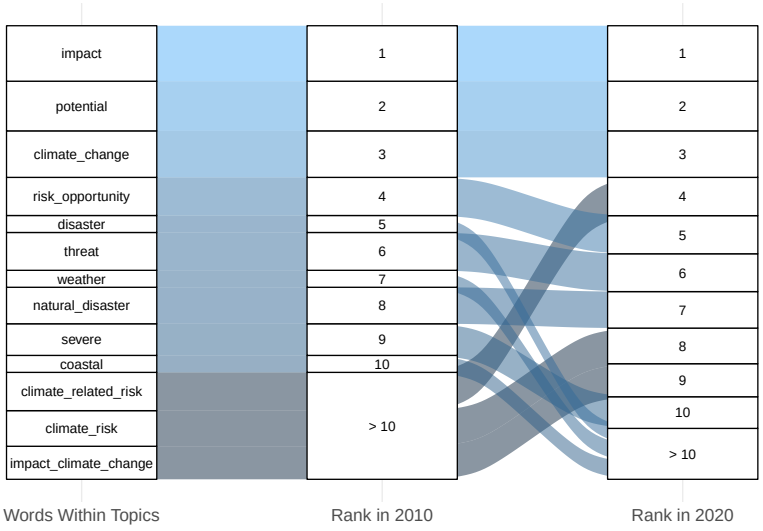
Access & Affordability



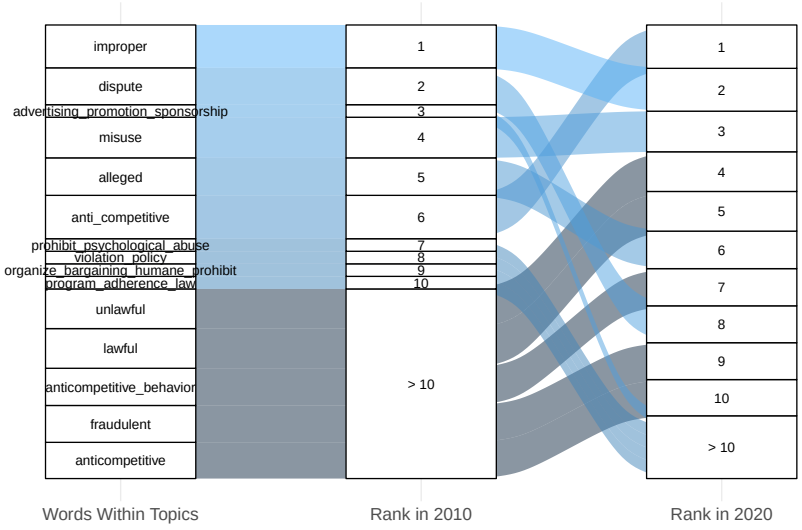
Air Quality

FIGURE IA. 2: CHANGES IN LANGUAGE OVER TIME

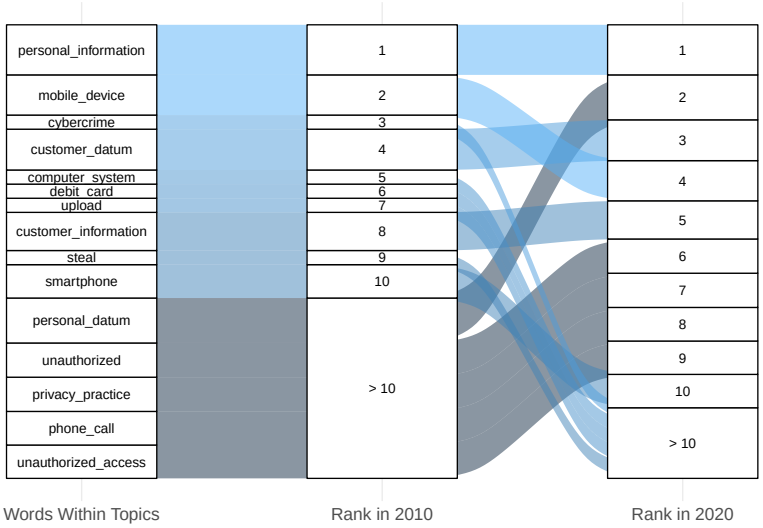
Appendix Figure IA 2 expands on Figure 6 and displays plots documenting the change in the 10 most important words, as measured by the term frequency, for the remaining 24 of 26 topics identified by SASB.



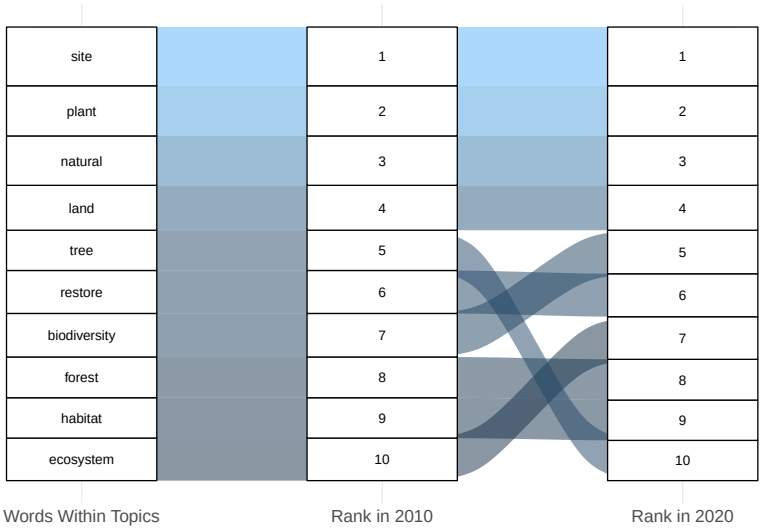
Physical Impacts of Climate Change



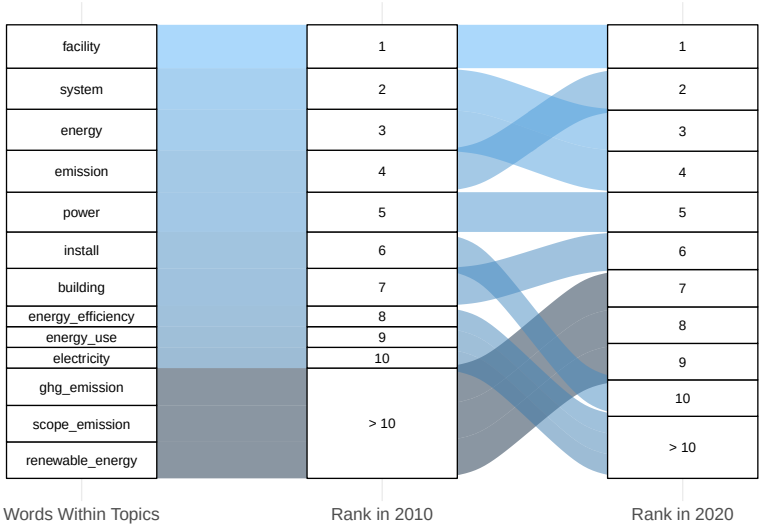
Competitive Behavior



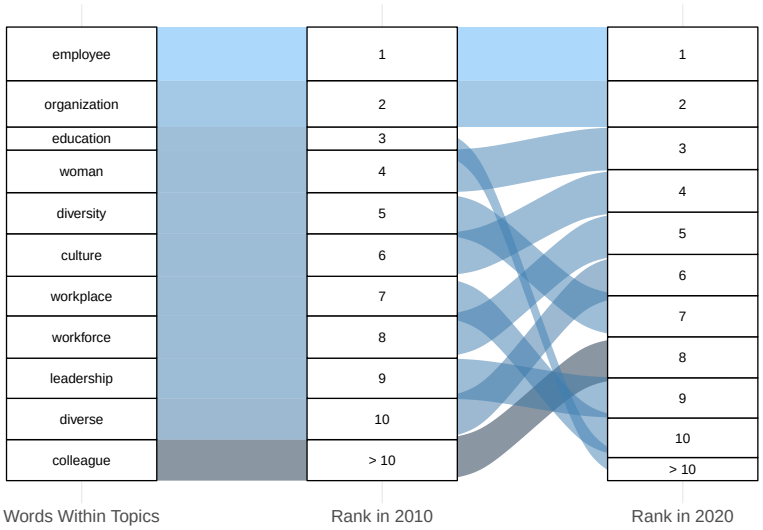
Data Security



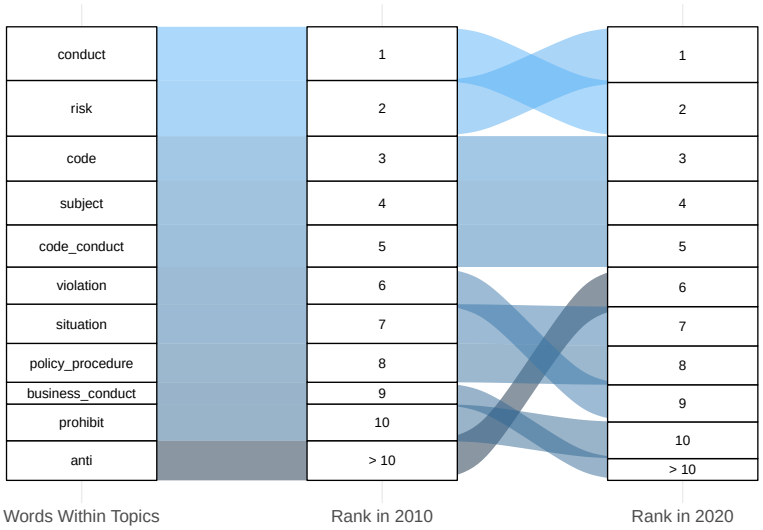
Ecological Impacts



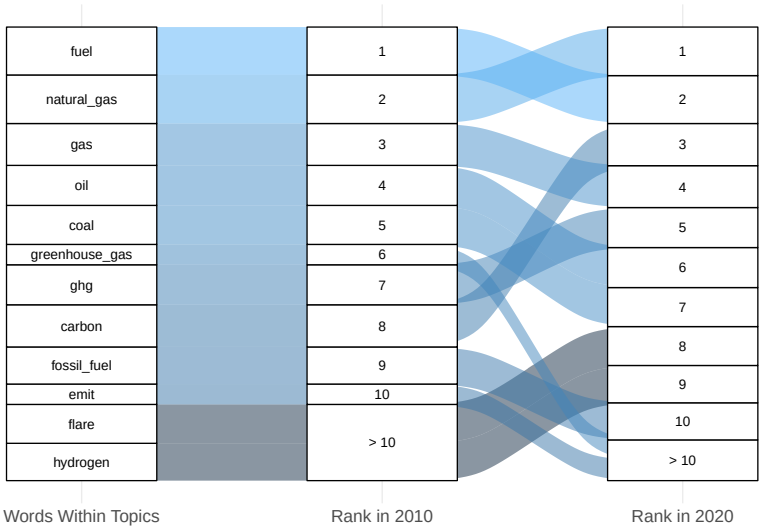
Energy Management



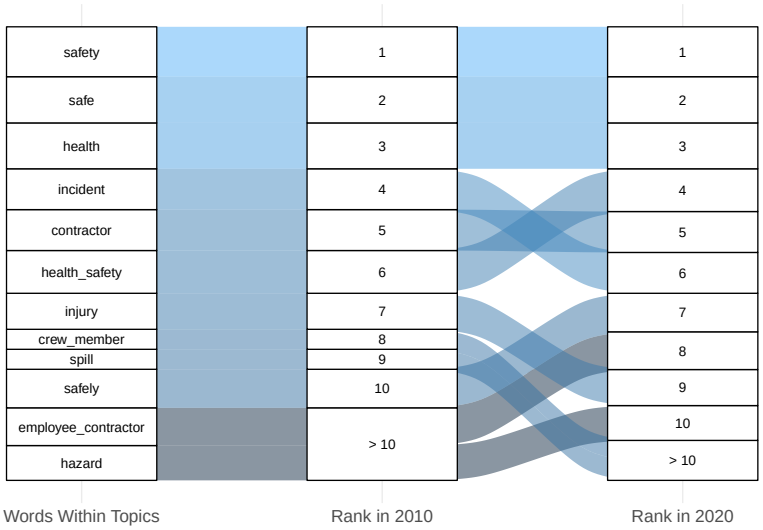
Employee Engagement, Diversity & Inclusion



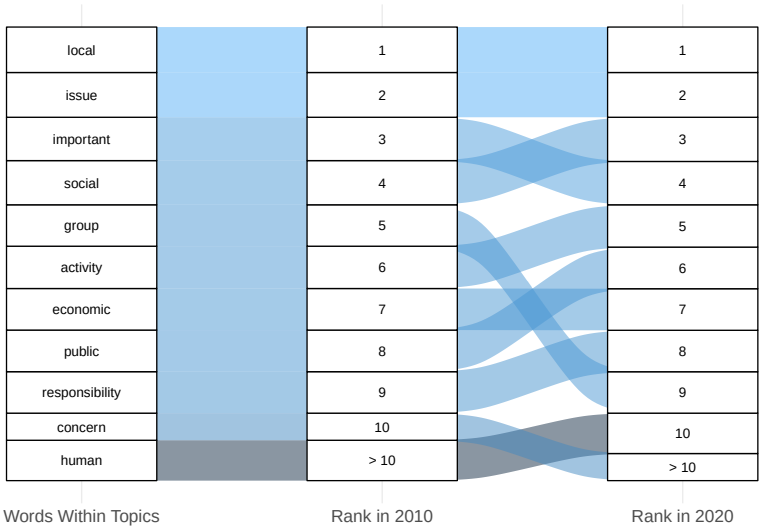
Business Ethics



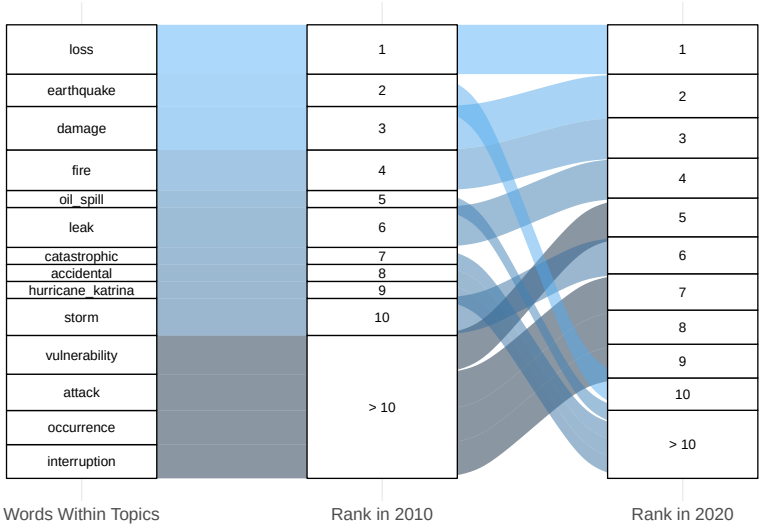
GHG Emissions



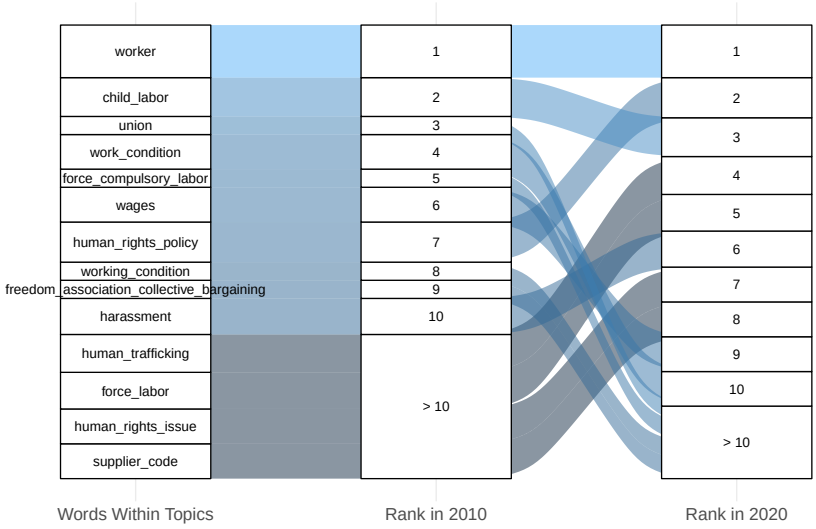
Employee Health Safety



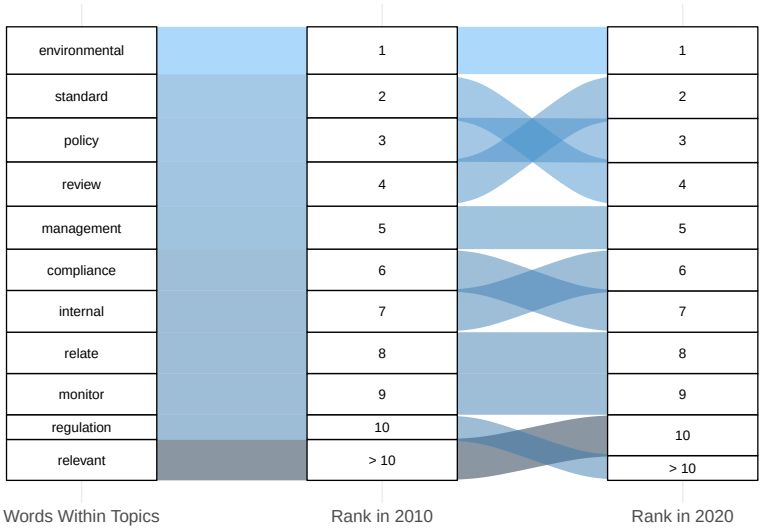
Human Rights Community Relations



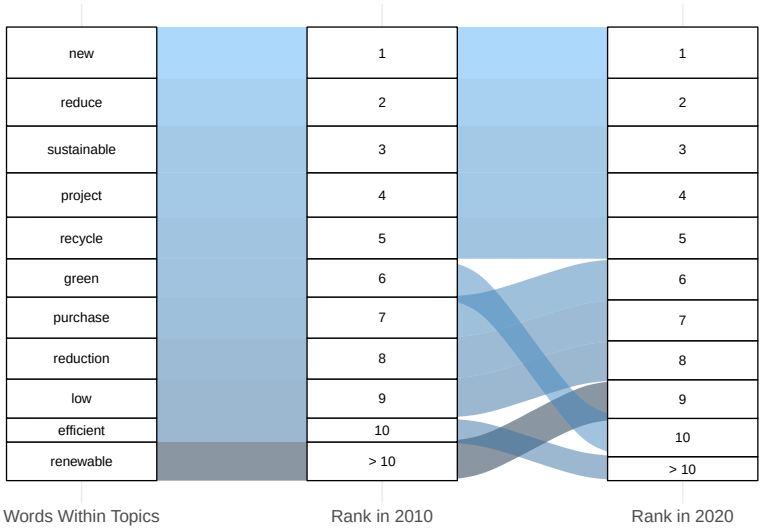
Critical Incident Risk Management



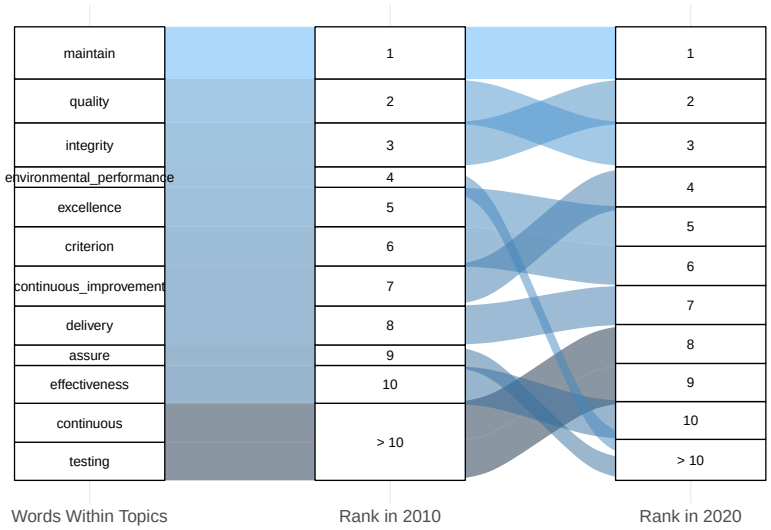
Labor Practices



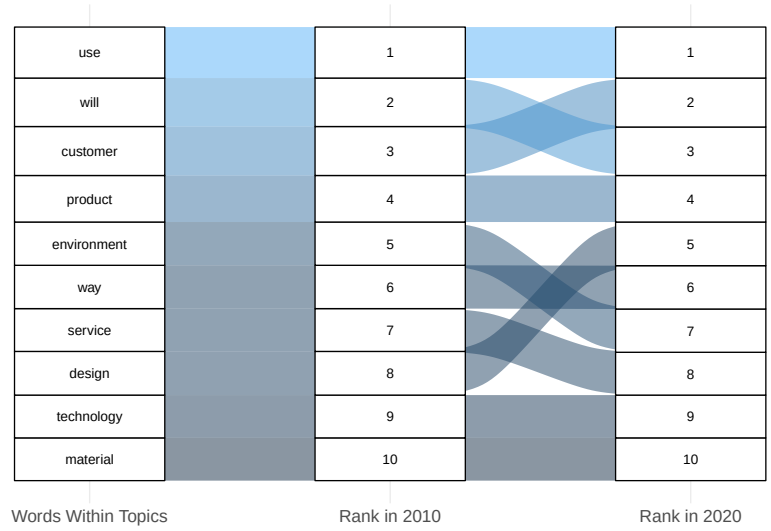
Management of the Legal & Regulatory Environment



Materials Sourcing & Efficiency

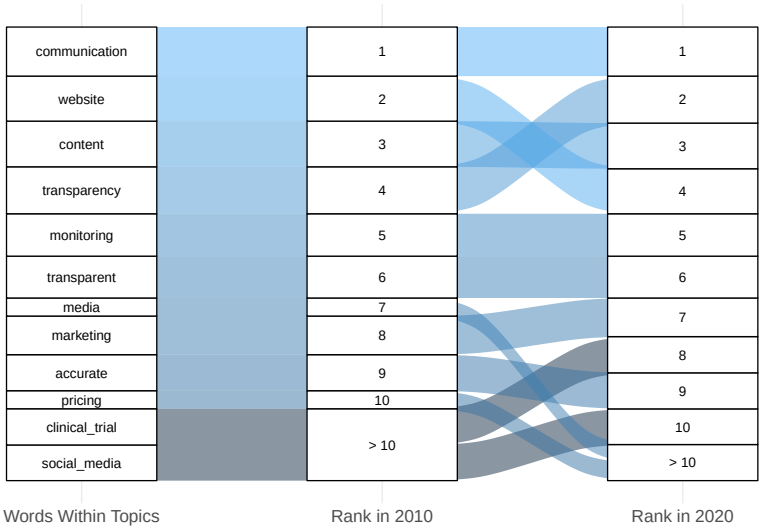


Product Quality & Safety

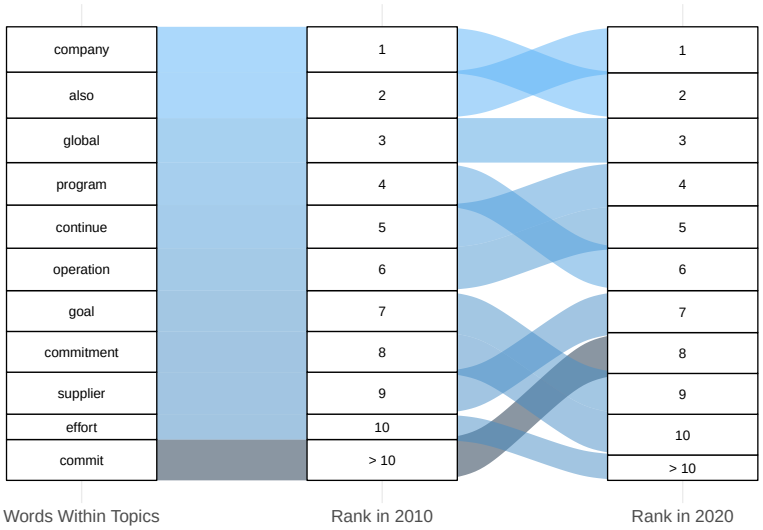


Product Design & Lifecycle Management

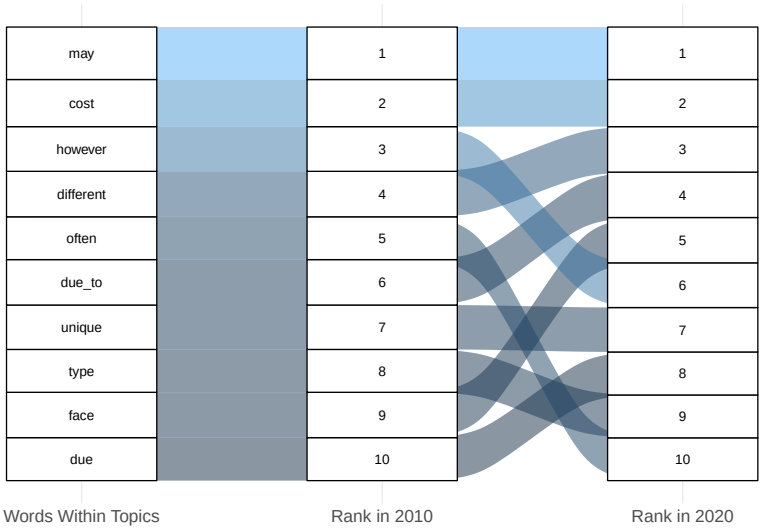
Business Model Resilience



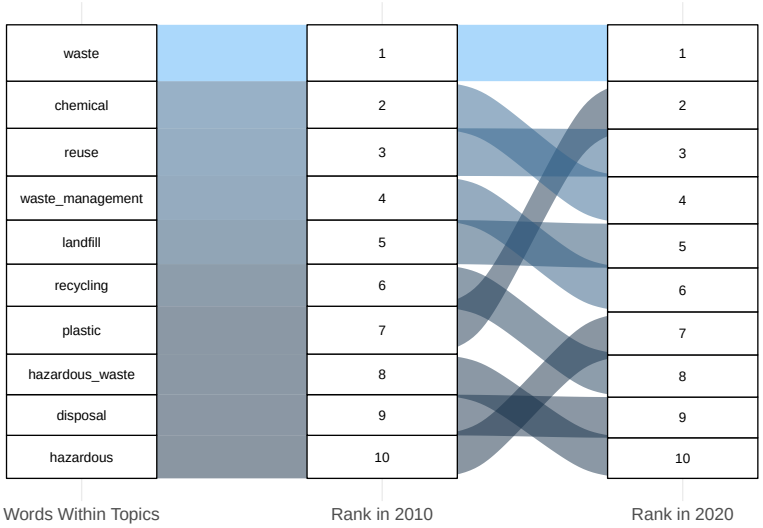
Selling Practices & Product Labeling



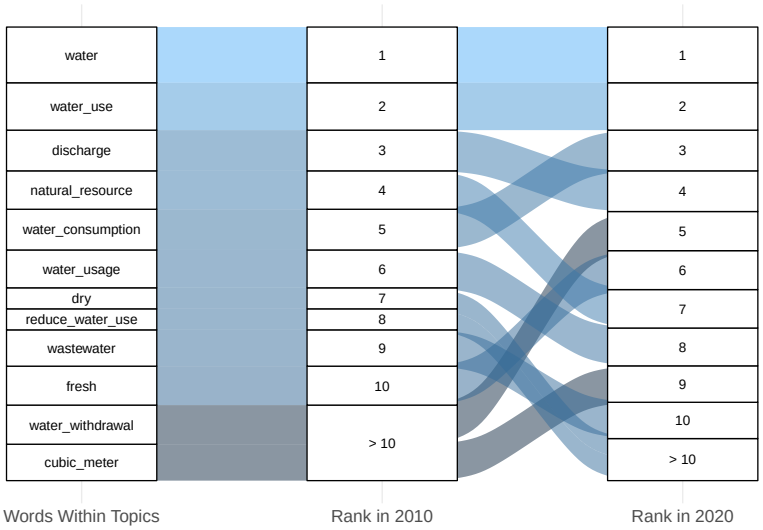
Supply Chain Management



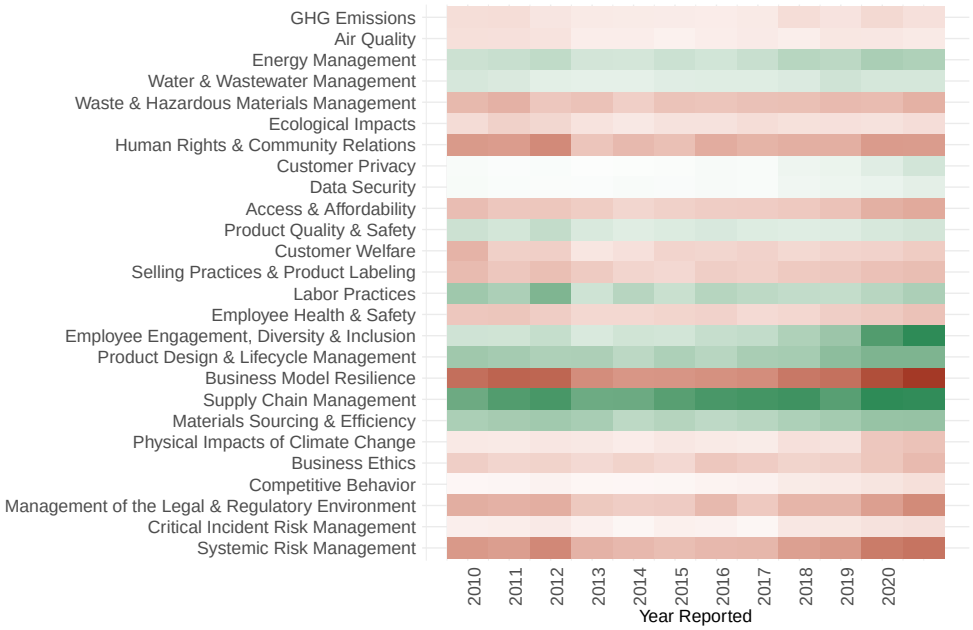
Systemic Risk Management



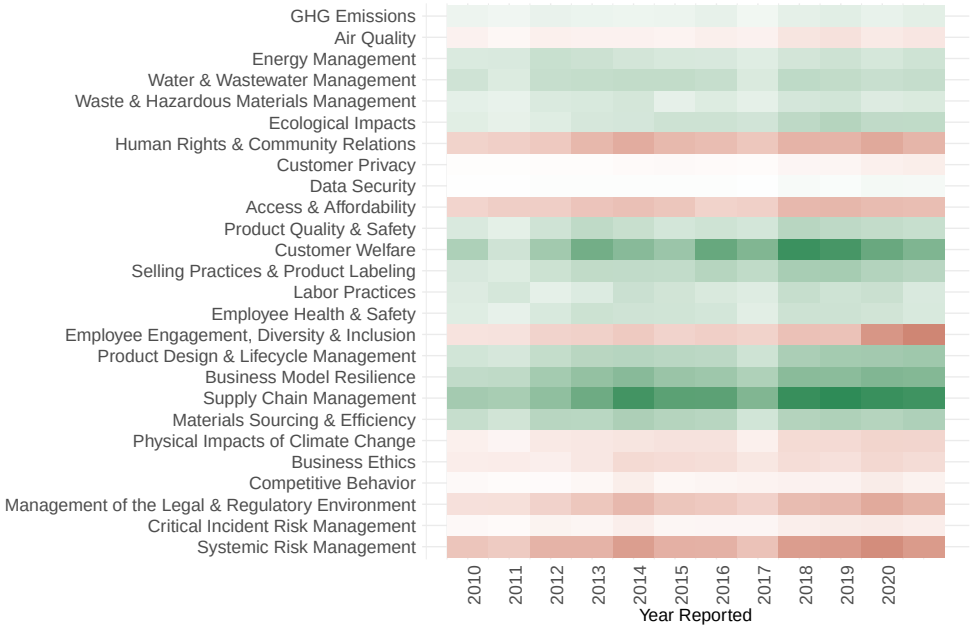
Water & Wastewater Management



Water & Wastewater Management



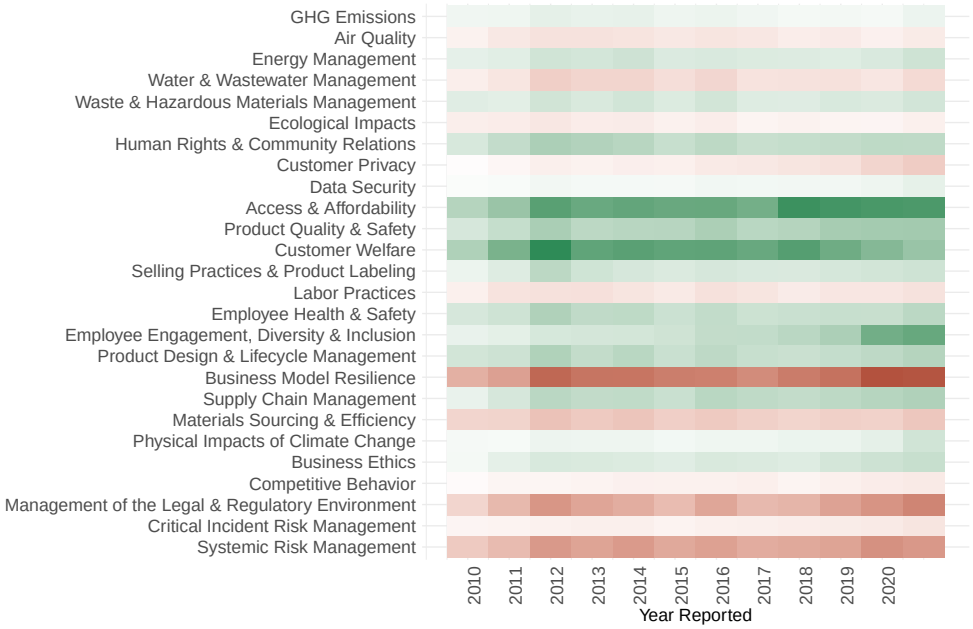
Consumer Good



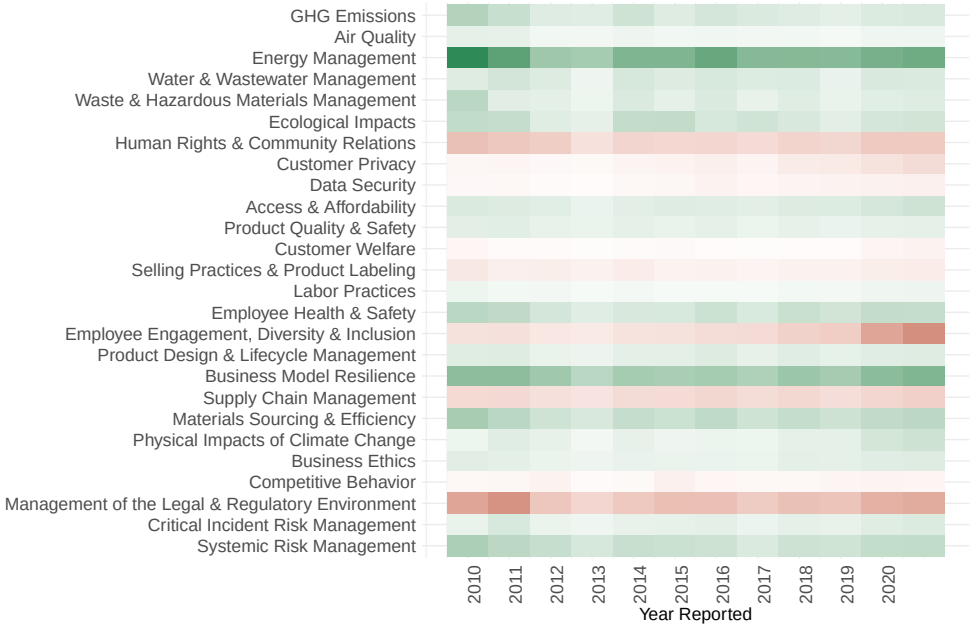
Food & Beverage

FIGURE IA. 3: EVOLUTION OF TOPICS BY SECTOR

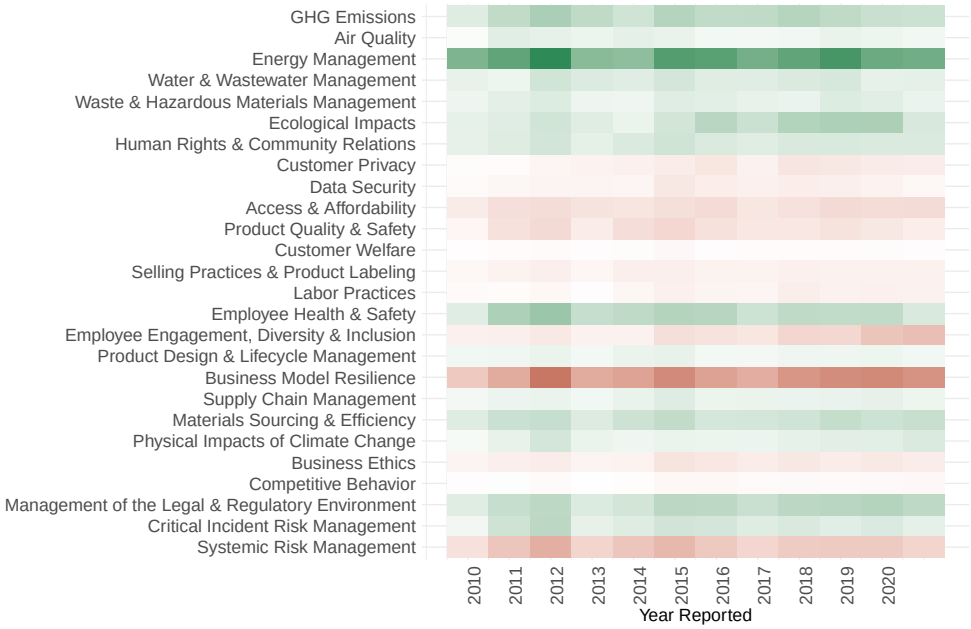
Appendix Figure IA 3 expands on Figure 9 and displays sector-level heatmaps of the relative amount of content in ESG reports for each of the 26 ESG topics identified by SASB. Topics with green (red) bars are those that are material (immaterial) for the sector. Darker (lighter) shades represent topics that are discussed more (less) within ESG reports. White bars represent topics that have relatively little content discussed in the documents, on average.



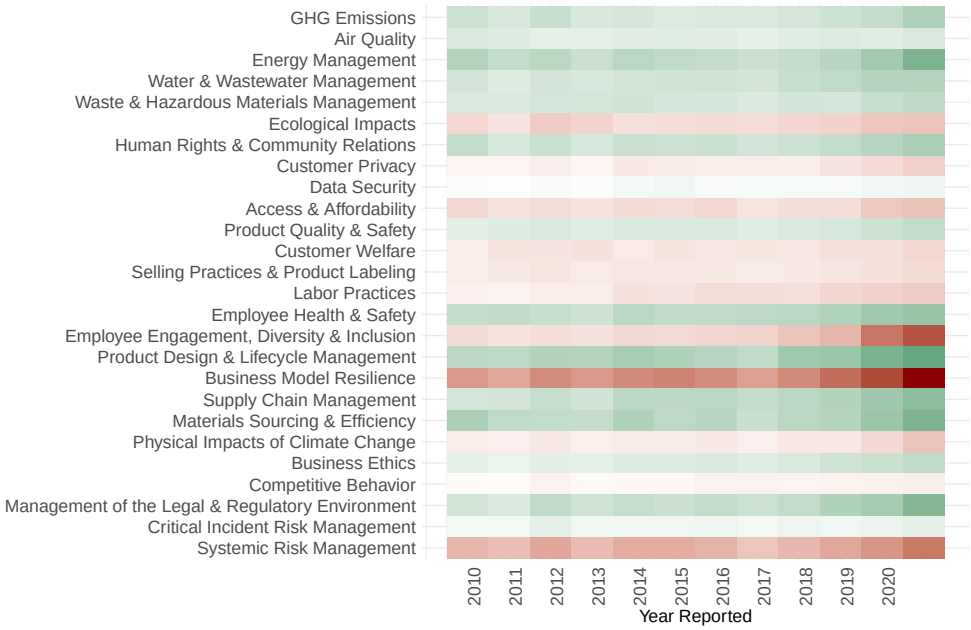
Health Care



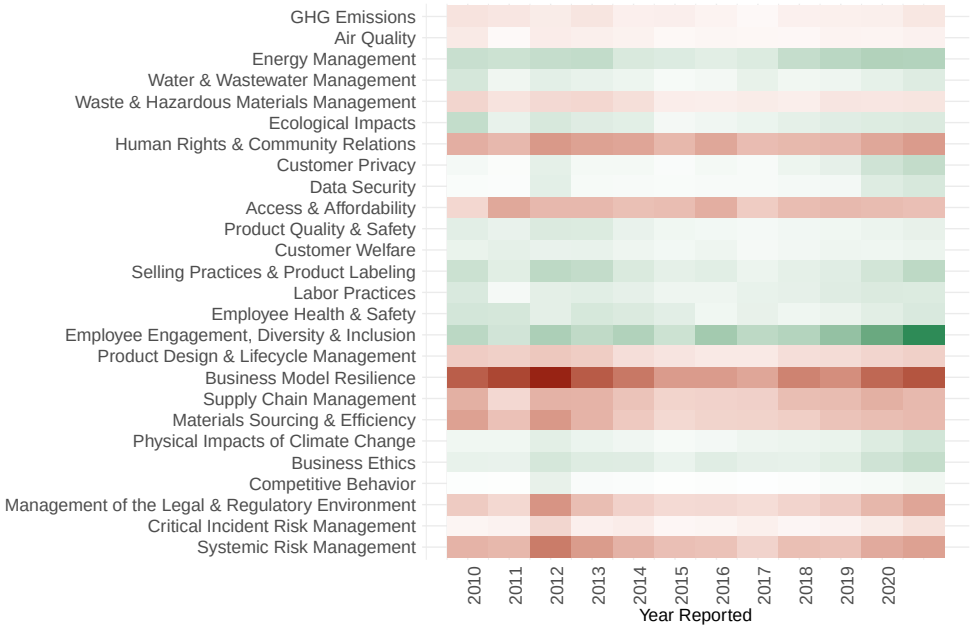
Infrastructure



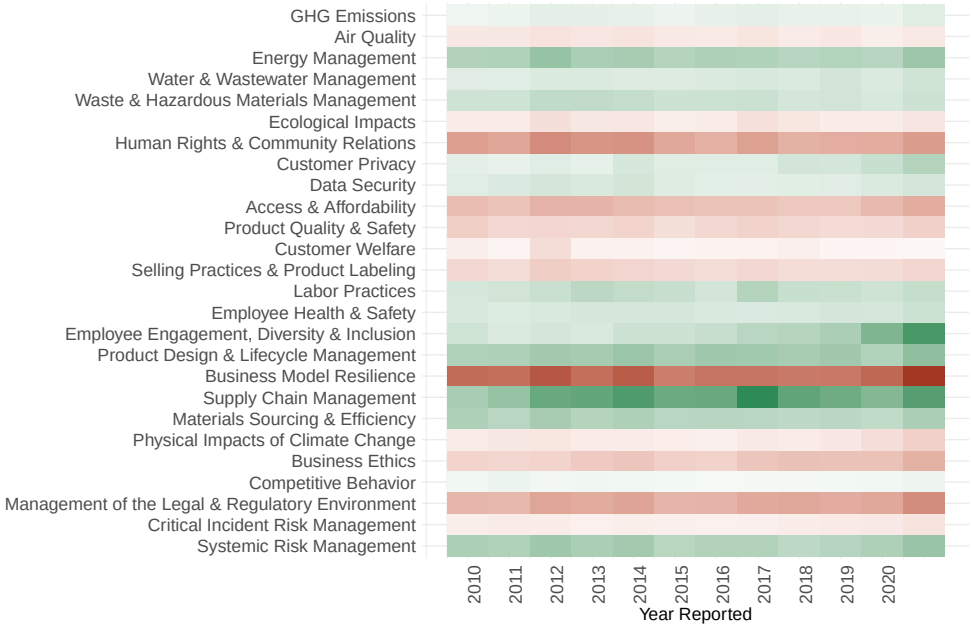
Renewable Resources & Alternative Energy



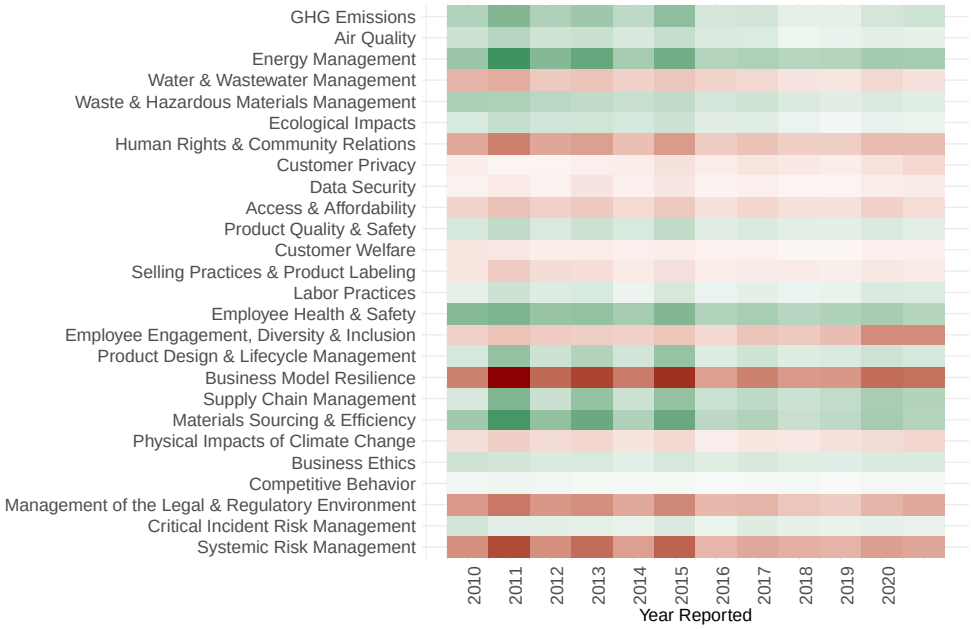
Resource Transformation



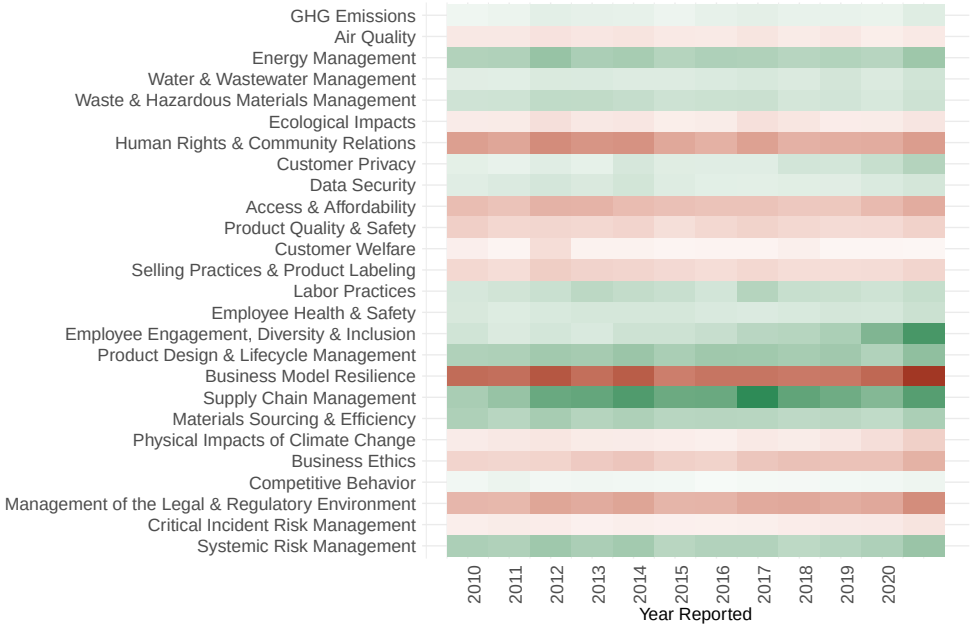
Services



Technology & Communication



Transportation



Technology & Communication

TABLE IA.1: STANDARDS TIMELINE

This table presents the provisional standard release dates for the 11 SASB sectors.

Sector	Provisional Standard Release Date
Health Care	6/1/2013
Financials	2/1/2014
Technology & Communications	4/1/2014
Extractives & Minerals Processing	6/1/2014
Transportation	9/1/2014
Services	12/1/2014
Resource Transformation	3/1/2015
Consumer Goods	6/1/2015
Food & Beverage	6/1/2015
Renewable Resources & Alternative Energy	12/1/2015
Infrastructure	3/1/2016

TABLE IA.2: TOP IMPORTANT WORDS PER TOPIC

Table IA.2 lists the 50 most important words, as identified by the topic modeling algorithm, in order by descending importance for each of the 26 ESG topics defined by SASB.

Topics	Top Words
Environment	
GHG Emissions	methane, gas, carbon_dioxide, c02, natural_gas, potent_greenhouse_gas, nitrous, co2, methane_release, release_atmosphere, carbon_dioxide_c02, methane_gas, sulfur, methane_emission, fugitive, sulfur_hexafluoride_sf6, coz, landfill_gas, hydrocarbon, nitrous_oxide, waste_gas, fugitive_emission, oxygen, gas_produce, sf6, combustion_emission, carbon_dioxide_methane, sulfur_hexafluoride, escape_atmosphere, flare, nitrogen_oxide, natural_gas_produce, carbon_dioxide_coz, emit_atmosphere, waste_heat, nitrous_oxide_nox, flue_gas, refrigerant_gas, condensate, oxide, vent, air_pollutant, burn_natural_gas, gas_flare, capture_methane, combustion_fuel, heat_energy, sulfur_dioxide, process_emission, emit
Air Quality	particulate, pollution, pollutant, ppm, contaminant, particulate_matter, ozone, ground_level_ozone, smog, phosphorus, particulate_matter_pm, odor, toxin, vibration, oxide_nitrogen_nox, sulfur_dioxide_nitrogen_oxide, algal_bloom, soil_water, runoff, nitrogen_dioxide, particulate_emission, heavy_metal, pm2, soot, dust, trace_amount, emission_volatile_organic, volatile_organic_compound, aerosol, nitrate, hydrogen_sulfide, nitrogen_oxide_nox, carbon_monoxide, chloride, ozone_layer, chlorine, fume, mercury, hexavalent_chromium, dioxin, moisture, toxic_substance, receive_stream, water_temperature, acidification, volatile_organic_compound_voc, compound_voc, oxygen_level, total_suspend_solid, fine_particulate
Energy Management	electricity, grid, energy_use, energy_consumption, utility, energy, energy_efficiency, electricity_grid, power_grid, electric_grid, electricity_use, emission, energy_usage, renewable_energy_source, energy_source, electrical_grid, generate_electricity, energy_consume, electricity_supply, renewable_energy_generation, renewable_energy_resource, power_generation, fossil_fuel, energy_generation, electric_power, power_supply, generation_resource, cogeneration, renewable_electricity, renewable_source_energy, alternative_energy_source, heating_cool, energy_grid, electricity_consumption, renewable_power, combustion, power_source, energy_supply, electrical_power, reduce_energy_demand, electricity_generation, fuel_cell, increase_energy_efficiency, cooling, renewable_generation, ghg_emission, electricity_purchase, cleaner_burning_natural_gas, such_as_wind_solar, electricity_natural_gas
Water & Wastewater Management	freshwater, wastewater, fresh_water, water, freshwater_source, groundwater, recycled_water, non_potable_water, brackish, process_water, water_source, seawater, freshwater_use, rainwater, water_supplies, river_water, potable_water, water_supply, amount_water, treat_discharge, water_reuse, ground_water, water_use, brackish_water, irrigation, surface_water, purify_water, reclaim_water, grey_water, waste_water, cool_water, produce_water, evaporation, treated_water, potable, greywater, drinking_water, groundwater_source, fresh, surface_water_source, water_stream, reverse_osmosis, water_volume, treat_wastewater, fresh_water_use, surface_groundwater, nonpotable, cooling_water, reuse_water, reused_water
Waste & Hazardous Materials Management	hazardous, recycling, hazardous_waste, nonhazardous, non_hazardous, waste, disposal, waste_stream, incineration, waste_material, reuse_recycle, treatment_disposal, waste_management, reuse_recycling, hazardous_nonhazardous_waste, universal_waste, waste_treatment, regulate_waste, chemical_waste, recycling_reuse, hazardous_material, electronic_waste, waste_oil, recycle_reuse, disposal_waste, recycle_composting, waste_liquid, responsible_disposal, sludge, discard_material, hazardous_non_hazardous, ewaste, waste_collection, disposal_hazardous, hazardous_waste_generate, hazardous_waste_disposal, paper_plastic, waste_item, process_waste, waste_disposal, increase_recycling, safe_disposal, solid_waste, waste_recycling, scrap_metal, disposal_recycling, properly_dispose, recyclable_material, proper_disposal, municipal_solid_waste
Ecological Impacts	vegetation, biodiversity, topsoil, natural_habitat, wetland_habitat, wetland, plant_species, soil, native_species, wildlife, wildlife_habitat, habitat, critical_habitat, flora_fauna, native_grass, buffer_zone, riparian, native_plant_species, native_plant, invasive_species, erosion, reduce_erosion, native_vegetation, erosion_control, aquatic, grassland, wetland_area, habitat_species, ground_cover, habitat_wildlife, invasive_plant, shoreline, species, wildflower, sedimentation, vegetative, grass, floodplain, flora, vegetation_growth, grazing, pollinator_habitat, endanger_species, rangeland, fish_habitat, sensitive_habitat, plant_growth, protect_biodiversity, mangrove, migratory
Social Capital	
Human Rights & Community Relations	human_rights, political, rights, social, labor, law, public_policy, protect_human_rights, human, respect_human_rights, fundamental_human_rights, protection_human_rights, woman_s_rights, criminal_justice_reform, citizen, indigenous_people, rule_law, respect, democracy, welfare, government, association, issue, social_cultural, force, freedom, human_rights_defender, discourse, civil, protect_rights, public, immigration, congress, economic, indigenous, public_policy_position, principle, employment, political_process, advocacy_effort, freedom_expression, public_policy_effort, mining_community, respect_human, politics, society, responsible_gaming, human_rights_statement, human_right, immigration_reform

Topics	Top Words
Social Capital (continued...)	
Customer Privacy	cybersecurity, cyber_security, datum_security, information_security, datum_privacy, datum_protection, privacy, cyber, product_security, security, cybersecurity_risk, fraud_prevention, security_privacy, cyber_threat, cybersecurity_datum_privacy, information_security_cybersecurity, datum_privacy_cybersecurity, identifiable, consumer_privacy, datum_governance, risk_management, cyber_physical, cybersecurity_strategy, physical_security, vulnerability_management, privacy_security, business_continuity_cybersecurity, information_security_datum_privacy, cybersecurity_privacy, privacy_datum_security, datum_security_privacy, network_security, privacy_risk, ransomware, information_security_risk, cybersecurity_threat, information_security_privacy, customer_privacy, datum_privacy_security, security_product, customer_privacy_datum_security, datum_breach, sensitive_information, security_control, privacy_cybersecurity, cyber_attack, identity_access_management, protect_confidential_information, personally_identifiable_information, information_protection
Data Security	pii, it_infrastructure, phishing, encrypted, access_management, cyber_intrusion, attack, hacktivist, malicious, involve_personally_identifiable, denial_service_attack, hard_disk, radio_frequency_transceiver, baseband_processor, messenger, intrusion_prevention_system, security_testing, defend_against, firewall, phone_mail, detect_respond, patch_management, attacker, member_datum, confidential_personal_information, healthcare_datum, personal_sensitive_information, computing_device, facial_recognition, unwanted_party, anacy_pitney_bowe, lock_unlock, door_hanger, access_restriction, identity_theft, connected_device, mobile_device, mydavita, password, onstar, botnet, onedrive, identifiable_information, become_susceptible, video_surveillance_system, software_update, patient_datum, fraud_alert, scam, voltage_regulation
Access & Affordability	underserved, affordability, underserve, underserved_community, affordable, low_moderate_income_community, underserved_population, low_income, access_affordable, expand_access, equitable_access, low_income_community, affordable_housing, access_health_care, access_high_quality, reduce_health_disparity, improve_health_outcome, underserved_area, vulnerable_population, community_color, health_care_access, healthcare_access, quality_health_care, minority_community, safe_affordable, homeownership, low_moderate, health_care, underbanked, health_care_education, underserved_market, low_income_family, improve_access_healthcare, low_moderate_income, financial_inclusion, quality_healthcare, uninsured, increase_access_healthcare, low_income_population, healthcare, increase_access, rural, affordable_rental, rural_community, address_health_need, access_safe_affordable, access_quality_care, access_care, access_affordability, access_healthy_food
Product Quality & Safety	product_quality, quality, safety_quality, quality_safety, product_performance, product_testing, quality_control, quality_reliability, time_delivery, food_safety, patient_safety, food_safety_quality, product_safety, product_integrity, product_quality_safety, reliability, delivery_performance, purity, quality_standard, product_safety_quality, quality_service, service_delivery, service_level, quality_system, quality_delivery, cost_efficiency, product_development, service_quality, throughout_product_lifecycle, food_quality_safety, safety_efficacy, quality_integrity, product_reliability, quality_product, safety_reliability, quality_assurance, robustness, food_safety_program, meet_customer_requirement, improve_product_quality, quality_management_system, high_quality_product, product_availability, quality_assurance_program, safety_standard, operational_excellence, supply_chain_resilience, product_development_manufacturing, product_compliance, operational_reliability
Customer Welfare	antibiotic, nutrition, counterfeit, antimicrobial, medicine, pathogen, poultry, drug, medically_important_antibiotic, bovine, animal_health, parasite, hepatitis_c, counterfeit_medicine, cough, pesticide, anti_inflammatory, pharmaceutical_product, disease, animal, dietary_supplement, medication, prevent_disease, protein, preservative, prevention_treatment, treatment_cancer, treat_disease, animal_welfare, digestive_health, infectious_disease, farm_animal, hormone, gastrointestinal, tuberculosis, biotechnology, ingredient, veterinarian, cattle, health_nutrition, infection, allergy, infection_prevention, respiratory_disease, drug_abuse, analgesic, counterfeit_product, malnutrition, ailment, treatment_patient
Selling Practices & Product Labeling	advertising, labeling, marketing, transparency, marketing_practice, marketing_communication, responsible_marketing, labelling, product_marketing, marketing_material, accuracy, product_label, marketing_advertising, marketing_regulation, truthful, authentication, product_transparency, consumer_insight, communication, advertising_marketing, mislead, consumer_information, datum_integrity, include_advertising_promotion, digital_marketing, product_information, timeliness, transparent, product_labeling, responsible_marketing_practice, voluntary_code_relate_marketing_communication, increase_transparency, credibility, ethical_marketing, customer_education, promotional, transparency_reporting, product_description, greater_transparency, privacy_choice, label_use, consumer_engagement, promotional_material, label, fairness, accurate, clinical_research, protection_intellectual_property, advertising_material, customer_communication

Topics	Top Words
Human Capital	
Labor Practices	child_labor, freedom_association, child_force_labor, force_child_labor, force_labor, labor_child_labor, child_labor_force_labor, child_labor_force, force_labor_child_labor, child_force, underage_labor, freely_choose_employment, bond_labor, non_discrimination_freedom_association, freedom_association_collective_bargaining, force_labor_human_trafficking, force_bond, collective_bargaining, bargaining, union, labor_discrimination, force_compulsory_labor, involuntary_labor, labor_condition, prohibition_force, rights_worker, employee_rights, prohibit_force, harassment, labor_practice, harassment_abuse, international_labor_standard, young_worker, compulsory_labor, payment_wages, overtime, force_involuntary, labor_law, labor_human_trafficking, labor_standard, slavery, exploitative, labor_rights, non_discrimination, child_labor_discrimination, foa, child_labor_force_labor_human, human_trafficking, human_trafficking_slavery, compulsory
Employee Health & Safety	injury, accident, safety, safety_incident, incident, injury_illness, serious_injury, fatality, serious_incident, workplace_injury, injury_incident, accident_injury, employee_injury, injury_accident, collision, slip_trip_fall, incident_injury, hazard, workplace_incident, near_miss, prevent_injury, process_safety, vehicle_accident, improve_safety_performance, injury_occur, safety_risk, vehicle_incident, workplace_injury_illness, prevent_incident, workplace_accident, injury_prevention, ergonomics, safety_performance, recordable_injury, safety_program, safe_behavior, occupational_injury, injury_fatality, train_accident, illness, injury_rate, drop_object, safety_effort, process_safety_event, sprain_strain, human_error, slip_trip, work_height, safety_culture, related_incident
Employee Engagement, Diversity & Inclusion	inclusion, diversity, gender, racial, inclusion_diversity, discrimination, diversity_inclusion, equality, diversity_equity, dei, diversity_equity_inclusion, gender_racial, race_ethnicity, gender_race, equity_inclusion, racial_ethnic, inclusiveness, intersectionality, allyship, inclusion_effort, race_gender, racial_equality, workforce_diversity, racism, gender_equity, gender_diversity, equal_opportunity, inclusion_belong, dimension_diversity, commitment_diversity_inclusion, pay_equity, create_inclusive_environment, ethnicity, employee_diversity, racial_equity, equality_inclusion, inclusive_behavior, inclusive_culture, unconscious_bias, equity, gender_identity, inclusive_leadership, lgbtq, culture_inclusion, workplace_diversity, sexual_orientation, gender_equality, inclusive, underrepresented_group, inclusivity
Business Model & Innovation	
Product Design & Lifecycle Management	packaging, innovation, product_packaging, packaging_design, packaging_solution, product, sustainable_packaging, innovate, recyclability, packaging_innovation, product_design, packaging_material, product_innovation, circular_economy, material_innovation, circular_design, circularity, circular, technology_innovation, innovative_solution, solution, new_business_model, plastic_packaging, breakthrough_innovation, innovative, drive_innovation, material_selection, technology, sustainable, design_process, material_use, product_solution, manufacturing, material_choice, remanufacturing, contribute_circular_economy, green_chemistry, throughout_product_life_cycle, compostability, innovation_process, sustainable_solution, eliminate_waste, process_innovation, manufacturing_process, innovative_technology, end_life, product_service, material, circular_solution, breakthrough
Business Model Resilience	business_model, responsiveness, planning, execution, approach, strategy, agility, business_strategy, business_process, decision_make, decision_make_process, decision_making, capital_allocation, resiliency, resilience, value_creation, investment_process, integration, ability_adapt, flexibility, scalability, operating_model, capability, disciplined, long_term, focus, model, drive, strategic_planning, adaptation, ability, thoughtful, engagement, performance, structure, diversification, risk_mitigation, coordination, client, business_practice, in_order, allocation_capital, predictability, evolve, understanding, growth_strategy, profitability, strategic, collaboration, risk_management_approach
Supply Chain Management	supply_chain, supplier, supply_base, throughout_supply_chain, supply_chain_partner, chain, within_supply_chain, tier_supplier, value_chain, supplier_base, vendor, supplier_relationship, responsible_source, traceability, agricultural_supply_chain, factory, palm_oil_supply_chain, contract_manufacturer, across_value_chain, extend_supply_chain, product_supply_chain, raw_material_supplier, work_closely_supplier, throughout_value_chain, supplier_engagement, supplier_code_conduct, electronics_supply_chain, supplier_network, responsible_source_practice, material_supplier, supplier_partner, supplier_partnership, industry_wide_effort, business_partner, source, procurement_process, contract_factory, packaging_supplier, supplier_performance, business, supplier_sustainability, entire_supply_chain, electronics_industry, supply_chain_practice, business_relationship, conflict_mineral, manufacturing_supply_chain, along_supply_chain, palm_oil_supplier, component_supplier
Materials Sourcing & Efficiency	renewable, recycle, efficiency, renewable_energy, recycled, renewable_resource, recycled_material, reduce_emission, reduce_waste, reuse, renewable_natural_gas, energy_and_water_efficiency, recycled_content, waste_reduction, high_efficiency, improve_energy_efficiency, reduce_carbon_footprint, recycled_plastic, efficient, improve_efficiency, reduce_carbon_emission, renewable_material, carbon_reduction, reuse_material, renewable_fuel, lightweighting, energy_savings, cleaner, lower_carbon_footprint, maximize_energy_efficiency, increase_efficiency, steel_product, reduce_consumption, renewable_power_source, electric_vehicle, cullet, use_recycled_material, recyclable, alternative_fuel, recycle_infrastructure, plastic, efciency, reduce, reduce_greenhouse_gas_emission, packaging_efficiency, energy_efficient_lighting, lower_emission, lightweight_material, energy_optimization, material_efficiency

Topics	Top Words
Business Model & Innovation (continued...)	
Physical Impacts of Climate Change	extreme_weather, climate_change, rise_sea_level, change_climate, sea_level_rise, extreme_weather_event, drought_flood, change_weather_pattern, weather_event, flood, impact_climate_change, effect_climate_change, flooding, rise_temperature, water_shortage, severe_weather, water_scarcity, drought, flood_drought, severe_weather_event, severe_storm, climate_event, flooding_drought, natural_catastrophe, resource_scarcity, heat_wave, climate_change_resource_scarcity, increase_frequency_severity, biodiversity_loss, storm_surge, wildre, climate_related_event, natural_disaster, exacerbate, wildfire, change_precipitation_pattern, extreme_weather_condition, climate_change_water_scarcity, weather_related_event, natural_hazard, climate_pattern, frequent_extreme_weather_event, storm_flood, climatic_event, cyclone, weather_pattern, heatwave, catastrophic_weather_event, wildfire_flood, drought_wildfire
Leadership & Governance	
Business Ethics	corruption, bribery, bribery_corruption, corruption_bribery, money_laundering, financial_crime, conduct, corrupt_practice, bias, corrupt, facilitation_payment, competition_law, sexual_harassment, anti_corruption_law, anti_bribery, prohibit_bribery, foreign_corrupt_practice_act, conflict_interest, antitrust_competition, improper_payment, corruption_risk, terrorist_financing, unlawful_discrimination, economic_sanction, antitrust, corrupt_activity, anti_corruption_policy, corruption_anti, applicable_anti, trafficking_person, discriminatory, violation_law, anti_money_laundering, anti, antitrust_competition_law, discrimination_harassment, workplace_violence, human_rights_violation, misconduct, trade_sanction, improper_behavior, unethical, fraud, bribe, tolerate_form, corruption_form, corrupt_behavior, workplace_harassment, fcpa, discrimination_workplace
Competitive Behavior	collusion, bargaining_power, price_fix, embezzlement, corruption_extortion, payment_kickback, give_appearance, facilitation_tax_evasion, even_appearance, deceptive, extortion, kickback, unlawful_activity, falsification_document, discrimination_sexual_harassment, unethical_conduct, conspire, retaliation_harassment, any_and_all_form, client_s_confidential_information, collusive, violence_threat_violence_workplace, unfair_treatment, unfair_business_practice, forgery, criminal_civil, unethical_business_practice, other_intellectual_property_rights, debt_bondage, bribe_kickback, insider_trading_antitrust, labor_involuntary, coercive, illegal_payment, illegal_business_practice, unfair_deceptive_abusive_act, bribery_/_corruption, peabody_pac_solicitation, origin_age_disability_veteran_status, regard_working_condition_without_fear, acceptance_inappropriate, auditing_matter, antitrust_anti_competitive_behavior, infringement_indigenous, form_force_bond_indenture, financial_industry_law_regulation_description, prohibit_payment, unfair_trade_practice, associate_fraud_insider_trading, coercion
Management of the Legal & Regulatory Environment	regulatory, compliance, regulatory_environment, regulation, law_regulation, legislative_regulatory, regulatory_requirement, regulatory_compliance, legal_regulatory, regulatory_change, regulation_legislation, legislation, policy, legislation_regulation, government_regulation, legal_regulatory_requirement, regulator, regulatory_oversight, lobbying, compliance_law_regulation, compliance_requirement, comply_regulation, compliance_applicable, tax_law_regulation, regulatory_legal, governmental_regulation, requirement, regulatory_framework, legislative, compliance_activity, tax_law, regulatory_landscape, compliance_applicable_law_regulation, regulatory_matter, rule_regulation, compliance_obligation, legal_requirement, privacy_law, public_policy_activity, public_policy_issue, emerge_regulation, regulatory_legislative, comply, regulatory_obligation, governmental, applicable_regulation, public_policy_matter, regulatory_body, legal, federal_state_law
Critical Incident Risk Management	leakage, catastrophe, storm_damage, such_as_hurricane, landslide, natural_disaster_such_as_hurricane, system_failure, hurricane_tornado, hurricane_flood, terrorist_attack, cause_damage, business_interruption, damage, ice_storm, extreme_event, catastrophic, earthquake, damage_cause, flood_earthquake, catastrophic_event, fire_flood, power_loss, operator_error, earthquake_flood, hurricane_earthquake, equipment_failure, terrorist_act, high_wind, outage_cause, mechanical_failure, cyberattack, unexpected_event, flooding_wildfire, earthquake_fire, business_disruption, labor_strike, ooding, design_withstand, extreme_weather_event_natural_disaster, supply_chain_interruption, flood_fire, gas_leak, shock, service_interruption, fire_explosion, tornado_flood, tornado_hurricane, weather_event_such_as, earthquake_hurricane, chemical_spill
Systemic Risk Management	complexity, complex, complicated, highly_complex, uncertainty, scale_complexity, size_complexity, constraint, complex_nature, breadth, complicate, difficulty, inherent_uncertainty, increasingly_complex, interdependency, size, obstacle, variability, add_complexity, interconnectedness, ambiguity, challenge, sophistication, fragmented, specificity, logistical_challenge, complex_issue, cost_complexity, confusion, interplay, variation, urgency, inherent, logistical, reality, challenge_employee, vulnerability, face, increase_sophistication, size_scope, demand, challenge_face, change, breadth_depth, difficult, geography, context, market_dynamics, challenge_pose, bureaucracy

TABLE IA.3: ESG SCORES AND ESG DISCLOSURES

This table presents the relation between commercial ESG Scores and TF-IDF, which is the term frequency-inverse document frequency of all words within a topic. All Material Topics equals one if the topic measured by TF-IDF is material, and 0 otherwise. Standard errors are robust to heteroskedasticity and clustered at the firm and year levels.

	ESG Score (Standardized)	
	Refinitiv	MSCI
	(1)	(2)
TF-IDF	0.033***	0.026*
	(0.009)	(0.014)
TF-IDF × All Material Topics	−0.009	−0.007
	(0.005)	(0.006)
Topic × Year FE	Yes	Yes
Topic × Sector FE	Yes	Yes
Firm FE	Yes	Yes
Observations	95,160	88,218
Adjusted R^2	0.695	0.802

TABLE IA.4: ALTERNATIVE MEASURE OF TOPICS (WF-IDF)

This table examines the impact of SASB guidance on the level of material and immaterial disclosure using an alternative topic measure, WF-IDF, calculated as the weighted frequency-inverse document frequency. All Material Topics equals one if the topic is material for a given firm-year, as defined by SASB. Post indicates the years after SASB issued provisional standards for the focal sector. Control variables are those reported in Table 2. Column 6 reports results for a subsample from four years before to four years after the release of the SASB standards. Standard errors are robust to heteroskedasticity and clustered at the firm as well as year level.

	WF-IDF					
	Full	Full	Full	Full	Full	Narrow Window
	(1)	(2)	(3)	(4)	(5)	(6)
All Material Topics x Post	2.984**	3.282***	3.428***	3.428***	3.842***	2.067***
	(1.461)	(1.151)	(1.153)	(1.153)	(1.069)	(0.747)
Topic × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Topic × Sector FE	No	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	No	No
Year FE	Yes	Yes	Yes	Yes	No	No
Firm × Year FE	No	No	No	No	Yes	Yes
Controls	No	No	Financial	Financial and Text	Subsumed by FE	Subsumed by FE
Observations	87,048	87,048	87,048	87,048	87,048	61,256
Adjusted R^2	0.426	0.478	0.480	0.480	0.597	0.592

TABLE IA.5: CHANGES IN MATERIAL DISCLOSURES (SUBSAMPLE ANALYSIS)

Table IA.5 documents changes in the relative amount of material information disclosed in ESG reports around the release of the SASB sector-level standards that defined material topics. The table replicates Table 7 but randomly drops 20% of the sample from each year. TF-IDF is the term frequency-inverse document frequency of all words within a topic. All Material Topics equals one if the topic measured by TF-IDF is material, and 0 otherwise. Post is an indicator equal to 1 in the years after SASB published its standards for the firm's sector, and 0 otherwise. Control variables are those from Table 2. Column 6 reports the results for a subsample from four years before to four years after the release of the SASB standards. Standard errors are robust to heteroskedasticity and clustered at the firm and year levels.

	TF-IDF					
	Full	Full	Full	Full	Full	Narrow Window
	(1)	(2)	(3)	(4)	(5)	(6)
Material Topics × Post	7.696**	6.742**	6.974**	6.478*	8.162***	4.167*
	(3.809)	(3.252)	(3.264)	(3.423)	(2.945)	(2.065)
Material Topics	24.084***					
	(2.704)					
Topic × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Topic × Sector FE	No	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	No	No
Year FE	Yes	Yes	Yes	Yes	No	No
Firm × Year FE	No	No	No	No	Yes	Yes
Controls	No	No	Financial	Financial and Text	Subsumed by FE	Subsumed by FE
Observations	76,076	76,076	76,076	73,372	76,076	49,790
Adjusted R^2	0.372	0.412	0.413	0.414	0.507	0.496

Appendix A.1 SASB Standard Setting Process

The development of standards across all industries within sectors adhered to the following timeline. In the initial research phase, SASB collected evidence for each industry on the financial impact of sustainability issues to identify the industry-specific materiality of sustainability activities and related metrics. Unlike most industry classification systems that use sources of revenue to group companies into different sectors and industries, SASB uses a Sustainable Industry Classification System (SICS) to group similar companies based on their sustainability-related risks and opportunities. After the research phase, IWGs of stakeholders were organized to provide feedback on the identified issues and metrics. These groups had balanced representation from corporations, market participants, and public interest intermediaries. Their feedback on the materiality of topics and the usefulness of metrics was incorporated into the exposure draft standard.

In the final phase, the exposure draft standard was released for a 90-day comment period for any member of the public to provide feedback, a process that concluded in January 2018. Feedback was then analyzed and incorporated into the provisional draft standard. The standards were considered final when the complete set for all industries was reviewed and approved by the American Standards Institute, an independent Standards Council. SASB released a final set of codified standards for 77 industries across 11 sectors in November 2018.

Given that the final standards were adopted across sectors at the same time and that many firms cited and relied upon the provisional standards prior to full adoption, in our analysis, we focus on changes in ESG disclosures around the publication of the provisional standards. Since the standards did not undergo material revisions from when they were proposed to when they were finalized, this choice allows us to study how the staggered introduction of voluntary standards relates to disclosure choices.