Legislation vs Guidance: The ESG disclosure dilemma*

Luke Johnson
Bachelor of Business (Honours)

November, 2021

Abstract

I develop a language model to identify environmental, social, and governance (ESG) related sentences in corporate filings. Using a difference-in-differences design, I assess the effectiveness of the UK mandate on ESG disclosure. With Australian companies as the control group, I find that the UK mandate does not increase the ESG content within corporate disclosure and does not improve ESG performance. However, the ESG reporting mandate increases the comparability of corporate disclosure. Further, after the mandate, UK companies use language that is more impactful on investor perception. This thesis highlights that ESG reporting mandates act as an intermediary, providing certainty for both companies and investors.

`Keywords:` ESG reporting, mandatory disclosure, natural language processing
`JEL classification:` C8, M14, M48, G28

*I would like to thank my supervisors—Kenny Phua and Vitali Alexeev—for their ongoing advice and feedback throughout the completion of this research project. I am also grateful for the feedback of my industry mentor, Desiree Lucchese (U Ethical).

Certificate of Original Authorship

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text. I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Name of Student: Luke Johnson

Signature of Student: LJohnson

Date: November 22, 2021

Contents

1 Introduction

Global environmental, social, and governance (ESG) assets are set to reach one third of assets under management by 2025, over $50 trillion[1] (Bloomberg Intelligence, 2021). Regardless, fund managers face significant challenges with the quality of ESG data. In a survey of asset managers, several cite a "lack of transparency or insufficient corporate disclosure in relation to firms' ESG activities" (Index Industry Association, 2021, p. 5). This is because the principal source of ESG data comes from corporate disclosure.

There is serious debate surrounding ESG disclosure mandates, but are they effective? The need for ESG disclosure comes from investors. Regulators view mandatory ESG reporting as a tool to mitigate the challenges investors face with ESG data quality. On one hand, mandates help inform investors so they can divest from companies that may not align with their personal views. Further, if poor ESG performance results in divestment, a company may improve their ESG performance to save shareholder value. However, is regulation necessarily better than the free market at enhancing corporate disclosures?

In this thesis, I focus on three hypotheses. Do ESG reporting mandates (i) increase the ESG content within corporate filings, (ii) improve ESG performance, and (iii) change the content tonality in firms' ESG disclosure. I employ a difference-in-differences design to contrast the effects of mandatory ESG reporting in the United Kingdom (UK) to the investor driven ESG reporting in Australia. I develop a Bidirectional Encoder Representations from Transformers (BERT) model to identify the ESG content in corporate filings. BERT is a state-of-the-art language model that offers superior performance on text classification tasks. BERTs performance is driven by the model's ability to understand context. Context is important because words can have dual meanings. Consider the sentences "At the bank of the river" and "I need to deposit at the bank". The word bank holds different meanings in both sentences, where the meaning of the word depends on its context. BERT captures context by embedding each word in a sentence, with some part of every other word in the sentence. The word for bank would contain some part of the words for deposit or river, depending on the context. I train a BERT model to identify the ESG sentences in corporate filings to examine the effectiveness of ESG reporting mandates.

I find that, on average, ESG reporting mandates do not increase the level of ESG content in reports. By examining the ratio of sentences that relate to ESG in corporate fillings, I observe the ESG content in corporate filings grows over time, pointing to the increasing

[1] Dollar amount in USD

prominence of ESG topics in corporate filings. However, my results suggest that ESG mandates have reduced the variance in the proportion of ESG content across companies. While mandates are not effective at adding extra ESG content into corporate filings, they make reports more homogeneous. Firms are unique and require different levels of reporting, in which homogeneous reporting is more standardised, and standardised reporting allows for the comparability of unique firms.

Furthermore, the UK ESG reporting mandate does not improve company ESG performance. Using ESG scores from MSCI and Refinitiv, I find a strong relation between ESG content and ESG performance. This is consistent with the view that more ESG information allows ratings agencies and investors to better assess a company's ESG performance. The increase in ESG performance depends on the level of ESG content in corporate filings. The UK ESG reporting mandate was not effective at increasing the ESG content in corporate filings, and as a consequence, ESG performance.

Finally, I perform sentiment analysis to examine if firms become more positive in filings because of the ESG reporting mandate. I find that ESG reporting mandates increase the positivity of company ESG statements relative to general disclosure. Sentiment analysis allows me to assess the perception of the company portrayed image on ESG. Taking the difference between the ESG sentiment and the sentiment of the general content in corporate disclosure, I control for any effects that are driven by changes in the overall language used in reports. The sentiment of ESG sentences is generally more positive, and I conclude that mandated ESG disclosure has more impact on investor perception.

The findings of this thesis have implications for regulators exploring ESG reporting mandates. The ESG space is rapidly growing and with it the need for accurate information. By assessing the effectiveness of the UK ESG reporting mandate, I believe this research will facilitate more informed policy decision making. My findings suggest that mandates on ESG disclosure do not increase ESG content in corporate filings, however mandates act as an important intermediary, providing certainty for both companies and investors.

This thesis extends the literature on corporate ESG disclosure and corporate social responsibility. I present a novel way to assess and analyse the ESG content within corporate filings, using state-of-the-art textual analysis techniques. I highlight that ESG reporting mandates may not be as effective as previously thought. Prior attempts at assessing the UK ESG reporting mandate find that key performance indicators and narrative disclosure increase after the UK mandate was introduced (Hummel and Rötzel, 2019). These attempts employ weak textual analysis techniques using a keyword search. This methodology creates substantial levels of noise as firms can refer to their business environment, along with their businesses' impact on the environment. By implementing BERT, I can capture context and reduce the noise that would otherwise affect results by using keyword search methods.

Moreover, Ioannou and Serafeim (2017) find evidence to suggest that sustainability disclosure increases with disclosure mandates. Although my findings differ, their study assesses smaller countries within emerging markets like (China and South Africa) between 2005 and 2012. My thesis examines the UK and Australia between 2010 and 2019, two developed markets with relatively more mature reporting standards comparatively. The ESG space is booming, and with it, the investor requirements for information. Examining recent ESG disclosure mandates highlights that ESG disclosure levels have increased but not directly because of the ESG reporting mandate. The effectiveness of ESG reporting mandates may depend on the general market environment (i.e., emerging markets, developed markets).

The remainder of this thesis is structured as follows. I provide a brief background on reporting mandates in Section 1.1. Section 2 gives an overview of existing literature and develops hypotheses. Section 3 describes the data creation process. Section 4 outlines the research design. Section 5 presents my results and discusses my findings, and Section 6 concludes.

1.1 Background on Reporting Mandates

There has been a push in Europe, New Zealand, and the United Kingdom to mandate the disclosure of non-financial information focusing on climate and stakeholder related disclosures. Krueger et al. (2021) outlines the mandatory ESG reporting framework globally.

In the United Kingdom, section 414C(7) of the Companies Act 2006 (UK) was amended in 2013 on a "comply or explain" basis. This amendment mandates that large companies disclose information surrounding the impact that a company's business has on the environment, the company's employees, and any social/community and human rights issues that may affect the company. If the report does not include information surrounding these issues, the report must state which information it does not contain. This section has been further amended in 2018 and 2019 to include the section 172 statement. This statement requires directors to explain how they consider the interests of stakeholders and is aimed at encouraging improved corporate governance practices with investors still pushing for ESG issues to be considered in business activities (The Companies (Miscellaneous Reporting) Regulations 2018, UK). The later amendments have made ESG issues a responsibility of directors. The focus of my thesis is to examine the effectiveness of the 2013 disclosure mandate.

Krueger et al. (2021) state that Australian ESG reporting has been mandatory since 2003. Based on the ASX listing rule 4.10.3 (ASX, 2019). This rule states that a company must disclose a corporate governance statement following the recommendations of the ASX Corporate Governance Council. The focus of this listing rule was to guide companies to report on governance issues with an, if not, why not approach. This approach allows for a lot of freedom within reporting practices. There have been regular updates to the recommendations since 2003 in 2007, 2010, 2014, and 2019 (ASX Corporate Governance Council, 2019).

Regarding the ESG theme, the most relevant principles from these recommendations are managing risk and acting ethically and responsibly. Although classed as mandatory, about 8% of companies are deemed non-reporting in ESG by the Australian Council of Superannuation Investors (2015). Australia's mandatory reporting surrounds governance specifically and not the ESG theme. With ESG, an Australian "listed entity should disclose whether it has any material exposure to economic, environmental and social sustainability risks and, if it does, how it manages or intends to manage those risks." (Australian Council of Superannuation Investors, 2015, p.4). Although highly recommended by the ASX, the critical term is material, where a company must draw a link to the economic downside from their ESG risks to warrant reporting. Materiality allows for judgment, in which different parties may have different interpretations of materiality. Because of this judgment, companies take different approaches to the way they report. Where a company may err on the side of caution—or report with neglect—in their ESG disclosure. The ASX listing rule ultimately relates to governance, where companies are allowed significant freedom to report on ESG themed issues.

2 Literature Review and Hypothesis Development

One of the main drivers of ESG investing is the apparent link between corporate social performance (CSP) and corporate financial performance (CFP); criticised by Ullmann (1985) as "data in search of theory", noting the inconsistent findings of studies (Aupperle, Carroll, and Hatfield, 1985; Cochran and Wood, 1984) and further criticised by Griffin and Mahon (1997) for issues in methodology. The idea that a firm can do good while doing well counters the early argument of Friedman (1970) that a company must sacrifice profits to ensure CSP, where profit is the only goal of a company. Later studies find a positive relation between CSP and CFP (Waddock and Graves, 1997; Orlitzky, Schmidt, and Rynes, 2003; Russo and Fouts, 1997). Further studies into CSP and CFP stress the unique nature of CSP, focusing heavily on long term future performance (Eccles, Ioannou, and Serafeim, 2014; Hartzmark and Sussman, 2019). Material information is crucial in drawing the link between CSP and CFP (Eccles and Serafeim, 2013; Khan, Serafeim, and Yoon, 2016). Renneboog, Ter Horst, and Zhang (2008) highlight that investors are happy to accept the Friedman view, sacrificing financial gain for an ethical objective. Studies that explore CSP find it is mainly driven by the engagement of institutional investors (Dyck et al., 2019; Kim et al., 2019). ESG investors try to justify the link between CSP and CFP, enforcing the idea that they can do good whilst doing well. In a meta-study of 2,200 studies by Friede, Busch, and Bassen (2015), many reports highlight a positive relation between CSP and CFP that is stable over time.

Mandatory ESG disclosure is driven by the idea that financial markets will reward the proper behaviour in companies if they are transparent about their environmental and social impact (Konar and Cohen, 1997). Evidence suggests that increased regulatory-driven sustainability disclosure leads to an increase in firm valuations (Ioannou and Serafeim, 2017). When the United Kingdom introduced their mandatory reporting on ESG issues, Jouvenot and Krueger (2019) find a reduction in firm greenhouse gas emissions. The reduction in emissions was attributed to the costs and increased comparability of a firm disclosing high emissions. In the EU, mandating ESG reporting increased reporting quality (Mion and Adaui, 2020). Additionally, the market reacted negatively overall, with an exacerbated reaction to firms that did not have adequate sustainability disclosure before the mandate (Grewal, Riedl, and Serafeim, 2019). Palmiter (2017) reviews the SEC's attempt to mandate climate disclosure, concluding that mandates alone are insufficient, but may be a necessary part of managing corporate climate risk.

The informativeness of corporate disclosure is a focus for some academics. Eccles, Serafeim, and Krzus (2011) emphasise the need for a company to provide information specific to the needs of market users. Matisoff (2013) highlights the limitations of disclosure and transparency about the Carbon disclosure project. Regardless, they play an essential role in policy surrounding disclosure. Clarkson et al. (2013) show that transparent environmental disclosure, combined with proactive environmental strategy, can enhance a firm's share price. Companies that disclose material sustainability information exhibit greater price informativeness and allow inferences about the credibility of a firm's disclosure (Grewal, Hauptmann, and Serafeim, 2021; Matsumura, Prakash, and Vera-Muñoz, 2020). The biggest issue within ESG disclosure is the lack of consistent disclosure and reporting standards (Amel-Zadeh and Serafeim, 2018).

The voluntary nature of ESG reporting and disclosure in Australia and much of the world are typically uninformative and boilerplate (Palmiter, 2017). Uninformative reporting, combined with the growing length, redundancy, and decreased readability of current reports, highlights crucial issues pushing for more ESG disclosure (Dyer, Lang, and Stice-Lawrence, 2017; Loughran and McDonald, 2014). Increased disclosure in ESG related topics through mandatory statements may add further complexity to annual reports exacerbating an existing issue of narrative disclosure (Davies and Brennan, 2007; Miihkinen, 2012).

A significant criticism of ESG reporting mandates is the idea that they will increase complexity in reports, exacerbating boilerplate and uninformative reporting. But ESG conscious investors require this information to inform decision making. With the goal of ESG reporting mandates to increase ESG themed information in reports, I formulate the following hypothesis.

Hypothesis 1: *ESG reporting mandates increase the ESG content within corporate filings.*

Papoutsi and Sodhi (2020) find evidence to suggest that sustainability reports show sustainability performance. There also appears to be growing awareness of the climate issue (Hamilton, 2016). From a climate science perspective, there is no accurate measure at this point, and models are limited (Fiedler et al., 2021). Ilinitch, Soderstrom, and Thomas (1998) early notion holds the need for explicit measures for environmental performance to guide decision making.

Several studies emphasise the issues that come with environmental reporting. With a focus on chemical companies, Delmas and Blass (2010) find that firms with more advanced reporting and environmental management also had lower levels of environmental compliance and more toxic releases. Companies seeking to enhance the credibility of their sustainability reports are more likely to get them assured. However, it does not matter if the assurance comes from the auditing profession or not (Simnett, Vanstraelen, and Chua, 2009).

Evidence suggests that ESG ratings are not a suitable forward-looking predictor for CSP and are not very useful for investors (Chatterji, Levine, and Toffel, 2009; Daines, Gow, and Larcker, 2010). With the major concerns of ESG ratings being heavily driven by the difference and lack of convergence in the ratings, further studies have aimed to explore the validity of these ratings. An early study by Sharfman (1996) highlight that KLD social ratings were a weak indicator of CSP (MSCI purchased KLD in 2009). Moreover, Szwajkowski and Figlewicz (1999) find that KLD's social ratings were not influenced by firm value. A more recent study by Semenova and Hassel (2015) show that the environmental performance metrics used by the leading rating agencies—including the former versions of MSCI and Refinitiv—reflected environmental performance, even with the lack of convergence.

There is a clear link that an increase in reporting on ESG themed issues affects ESG performance. With the goal of ESG reporting mandates to increase ESG themed content in corporate disclosure, I develop the following hypothesis:

Hypothesis 2: *ESG reporting mandates improve ESG performance.*

There have been several studies criticising sustainability ratings. Christensen, Hail, and Leuz (2021) and Kotsantonis and Serafeim (2019) find that more disclosure on ESG issues leads to a greater disagreement in ratings. The most common focus is on the difference in ratings between different providers (Berg, Koelbel, and Rigobon, 2020; Delmas, Etzion, and Nairn-Birch, 2013; Gibson, Krueger, and Schmidt, 2019; Semenova and Hassel, 2015). The root cause for the difference in ratings is compound and complex. It may be because of different definitions of sustainability (Gray, 2010; Chatterji et al., 2016; Berg, Koelbel, and Rigobon, 2020), definitions of materiality (Eccles and Stroehle, 2018), or the scope and measurement of ESG issues (Widyawati, 2021; Berg, Koelbel, and Rigobon, 2020). Transparency was highlighted as a significant concern by European Union led studies (Boffo and Patalano, 2020; ERM, 2021) whereas rating agencies that rely on publicly available information transparency may have a non-trivial impact on the ESG ratings.

Additionally, there have been concerning issues within the ESG ratings. Tang, Yan, and Yao (2021) emphasise rater ownership and sister firms, with firms that are held by the same owner as the rating agency receiving higher ESG scores. These sister firms also have more future negative ESG incidents. Referring to Refinitiv ESG ratings specifically, Berg, Fabisik, and Sautner (2021) find that historic ratings change in consecutive annual downloads of the dataset, providing different ESG scores for the same points in time. The change was clarified as an update to their scoring methodology. Refinitiv ratings are recalculated weekly at an annual period based on new data available, which is detrimental to the replicability of future studies.

I use ESG ratings as a measure of ESG performance. ESG performance is hard to quantify, in which the ESG ratings are a quantified metric that is easily accessible. The ratings are typically a relative measure of performance based on a comparison to peers, where ratings may be a fuzzy measure of ESG performance. Further, with the difference in ratings, it is important to consider more than one definition of ESG performance. As such I use both MSCI and Refinitiv ESG ratings. Appendix I contains the different scoring methodologies used by both MSCI and Refinitiv.

Many studies investigate the effect voluntary disclosure has on ESG reporting. Davies and Brennan (2007) find opportunistic behaviour like impression management—where management frames the disclosure to mislead stakeholders—a significant driver for narrative disclosure. This behaviour leads to the obfuscation of bad news, which is also prevalent in declining readability of company reports (Dyer, Lang, and Stice-Lawrence, 2017; Fabrizio and Kim, 2019; Loughran and McDonald, 2014). Similarly, Kim and Lyon (2011) find selective reporting where firms report a reduction in emissions, but they are realistically increasing. Evidence suggests that an increase in the quality of voluntary environmental disclosure results in greater firm value; This can be driven by shareholder activism or company self-improvement (Flammer, Toffel, and Viswanathan, 2021; Plumlee et al., 2015).

Various studies explore the effect of greenwashing by companies. Delmas and Burbano (2011) express how greenwashing is hard to mitigate within the limited and uncertain regulations. They find a significant negative effect on the confidence in green products. Another way firms greenwash is through selectively disclosing their benign impacts to appear transparent whilst hiding their actual performance (Marquis, Toffel, and Zhou, 2016). By directly comparing green claims against actual performance, Kim and Lyon (2015) find that corporate output growth and deregulation affect the choice to greenwash. Cho, Roberts, and Patten (2010) emphasise that the language and verbal tone used in environmental disclosure should be considered.

Similarly, Naumer and Yurtoglu (2020) highlight that it is not what you say, but how you say it with ESG news impacting CDS spreads. Companies use disclosure as a legitimising tool, where companies with poor environmental performance provide offsetting or positive disclosure in their financial reports (Cho and Patten, 2007; Patten, 1992). Because disclosure surrounding ESG is open to managerial discretion, investors are forced to analyse the narrative being portrayed critically. It is important to consider the nature of language used in ESG themed reporting, as such I formulate the following hypothesis:

Hypothesis 3: *ESG reporting mandates change the content tonality (sentiment) in firms' ESG disclosure.*

Textual analysis in finance has many applications, from detecting fraud in financial statements to inferring the information in company reports (Craja, Kim, and Lessmann, 2020). Early studies explored the effect that news media sentiment has on stock returns (Tetlock, 2007; Tetlock, Saar-Tsechansky, and Macskassy, 2008). Further studies add to this, exploring the effects that news media has on momentum, investor relations, and even exploring how local news media outlets receive biased coverage (Gurun and Butler, 2012; Hillert, Jacobs, and Müller, 2014; Sinha, 2016; Solomon, 2012). Multiple studies explore the content of 10-k reports, finding that tonality, sentiment, ambiguity, and constraining words all contain meaningful financial information (Bodnaruk, Loughran, and McDonald, 2015; Loughran and McDonald, 2011; Feldman et al., 2008; Friberg and Seiler, 2017). Additional studies find that the content of 10-k's help quantifies company risk (Bao and Datta, 2014; Jegadeesh and Wu, 2013). Textual analysis can help identify complexity and obfuscation in financial reports (Courtis, 2004).

More recent studies use textual analysis and natural language processing (NLP) to identify ESG related information in financial reports. Moniz (2016) uses latent Dirichlet allocation (LDA), a generative topic model, to identify the material, social issues in news media. Findings suggest that accurate linguistic measures can infer information to predict firm earnings. Recent advancements in NLP models, such as BERT have enabled studies that can accurately identify distinct complex topics within textual data. Kölbel et al. (2021) use BERT to show that climate risk is interpreted as financial risk. Firms are dis-incentivized to disclose transition risk, requiring a greater need for accurate regulation. In another application for BERT, Bingler, Kraus, and Leippold (2021) focus on voluntary climate disclosure, highlighting that most of this disclosure is heavily cherry-picked and non-material, concluding with a push for mandatory reporting. Sentiment also has a prominent role surrounding ESG issues. Many studies find significant evidence that ESG related news sentiment not only reflects in stock returns but also can be used to hedge risk (Bessec and Fouquau, 2020; Engle et al., 2020; Serafeim, 2020).

In summary, to assess the effectiveness of the UK ESG reporting mandate, I focus on 3 main hypotheses:
ESG reporting mandates:

1. **Increase** the ESG content within corporate filings

2. **Improve** ESG performance

3. **Change** the content tonality (sentiment) in firms' ESG disclosure

3 Data

Corporate filings data for UK and Australian companies are sourced from Refinitiv. These filings are used to assess the thematic ESG content and sentiment for the period between 2010 and 2019. Only the constituents of the ASX 300 and FTSE 350 indices are considered. The Refinitiv database can be screened for relevant filings like annual reports or for specific filings like 10-k's. I screen filings by annual and ESG related disclosure based on Refinitiv's classification filter. The annual filter contains filings such as annual reports and 10-k's. The ESG disclosure filter contains bylaws, web-based ESG disclosure, audit/remuneration reports, code of conducts, and other general ESG related filings. Based on availability, I only consider publicly listed companies in this study. The UK mandate for ESG disclosure covers all large companies. But the Refinitiv corporate filings database only considers global listed equity and access to private company filings is limited.

For robustness, I consider both MSCI and Refinitiv ESG ratings for the period from 2010 to 2019[2]. For added comparability, ratings from both sources are scaled to a [0,1] range.

Figure 1 shows the reduction in the sample size from 650 companies (the combined number of constituents in the ASX 300 and FTSE 350). I first remove 208 Australian and 199 UK companies from the sample because of the filing's availability for the entire sample period. A further 26 Australian and 50 UK companies are removed from the sample because of ESG ratings' availability. The final sample size is 101 UK companies and 66 Australian companies. The sample comprises larger companies that existed throughout the sample period. There are a total 17,625 filings, that come in a single file PDF format. The textual data is extracted from the filings to be used in a language classification model. The textual data cleaning process can be found in Appendix A.

The final sample group is economically meaningful. Using 2019 market capitalisation, in Australia, the final sample makes up 50.92% of the size of the ASX 300 starting point. Similarly, in the UK, the final sample makes up 53.01% of the size of the FTSE 350 starting point.

[2] I downloaded the Refinitiv ratings on the 20/08/2021

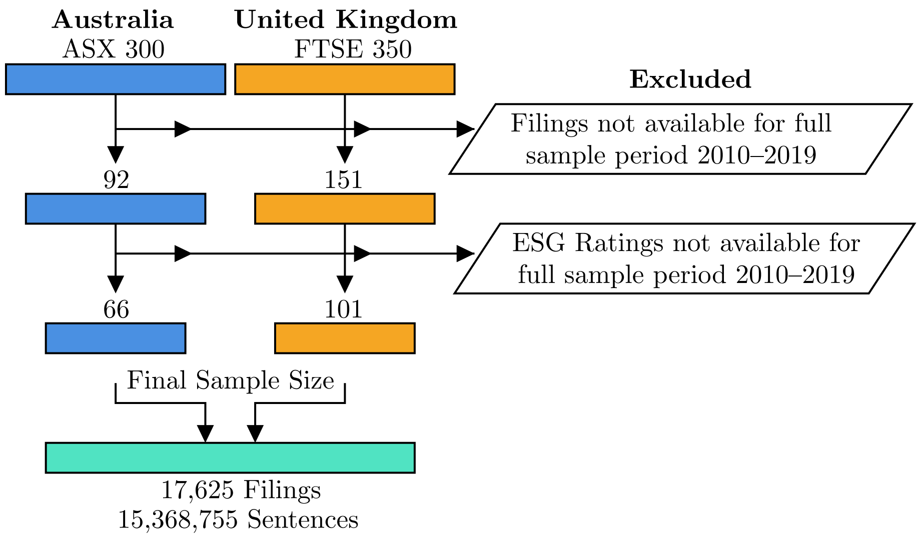


Figure 1. Sample Reduction Process.
This figure outlines the sample reduction process starting from the ASX 300 and FTSE 350. The final sample consists of 66 Australian and 101 UK companies. There are a total of 17,625 filings and 15,368,755 Sentences.

3.1 Language Classification

To examine the ESG content in reports, I develop a language model to identify sentences that relate to ESG. There are many methods available with ranging performance. Table 1, summarises the methodologies available for text classification.

Table 1. Approaches for text analysis.

This table outlines the different methods available for text analysis. Lexicon-based methods are only appropriate for sentiment analysis. Supervised machine learning methods can manage all text classification tasks. Unsupervised machine learning methods are only appropriate for topic modelling. Only supervised machine learning models like BERT are appropriate for text classification.

		Machine Learning Methods	
Task	Lexicon-based methods	Supervised methods	Unsupervised Methods
Text Classification			
Sentiment Analysis	✓	✓	
Text Classification		✓	
Topic Modelling			✓
Examples	VADER	BERT, SVM	LDA

Lexicon based methods are limited to classification of sentiment as they are reliant on dictionaries. This reliance on a dictionary also means words with dual meanings are lost. The dual meaning is important to capture as it limits the performance of Lexicon based methods. Consider the two short phrases "wait on second" and "he ran second place". The word "second" has drastically different meanings in each example. One refers to time and the other position. A lexicon-based model cannot classify words with dual meanings. Loughran and McDonald (2011) use a lexicon based measure for sentiment analysis.

Unsupervised machine learning methods for textual analysis, such as LDA have been misused for text classification. LDA is a topic model which can digest a large corpus of text (Blei, Ng, and Jordan, 2003). There have been some uses in finance to classify text using LDA but in textual analysis, it is a model only effective at topic modeling (Hansen and McMahon, 2016).

Supervised machine learning methods based on the attention architecture offer significant performance increases compared to existing methods (Vaswani et al., 2017). This can be attributed to the model's ability to understand context. I use a BERT model in this thesis to classify whether a sentence from corporate filings relates to ESG (Devlin et al., 2018). BERT is used in the papers of Kölbel et al. (2021) and Bingler, Kraus, and Leippold (2021) to identify climate related disclosure.

3.2 Bidirectional Encoder Representations from Transformers

I use a BERT model to measure the ESG content and sentiment of sentences in corporate filings. The model is trained to classify sentences into two classes. If they relate to ESG or are a part of the general content in corporate filings. An explanation of the inner workings of BERT can be found in Appendix B.

BERT is a pretrained language model that is said to have a base semantic understanding of language. The model can be further fine-tuned for task specific applications. BERT achieves superior performance on natural language processing (NLP) tasks by leveraging transfer learning. While there are newly emerging robust alternatives to BERT like RoBERTa[3], ALBERT[4] and ERNIE[5], these are merely extensions to the BERT framework (Liu et al., 2019; Lan et al., 2019; Zhang et al., 2019). In this project, developing a ESG

[3] RoBERTa: A Robustly Optimized BERT Pretraining Approach
[4] ALBERT: A Lite BERT for Self-supervised Learning of Language Representations
[5] ERNIE: Enhanced Language Representation with Informative Entities

classifier allows for consistent, comparable, and replicable results. For sentence classification, the model can undergo specific supervised training on a labelled dataset. The task specific training leverages the previous understanding of language and applies it to a specific task, such as ESG classification.

3.2.1 Context and Attention

In machine learning NLP applications, words require numeric representations. For BERT, this comes from the word embeddings. Word embeddings are vector representations of words (word vectors). The fundamental idea with word vectors is that words with similar meaning should have similar vectors. The famous quote from Firth (1957), "You shall know a word by the company it keeps", emphasises that words of similar meanings should have similar vectors (think big and large).

But similar vectors dont account for words with dual meaning. This is why context is important. Consider the bigram "climate change". The word "change" on its own can have many meanings. But together with "climate", the words each make up a different meaning than they would individually. The word for "change" is known by the company of "climate" and vice versa. Simply mapping a single vector representation for a word is ineffective when words have dual meanings. BERT can capture a level of context through the attention mechanism. This mechanism assigns a piece of every word in the sentence to every other word (meaning is assigned based on a words company). BERT outputs a contextualised word embedding where the token vectors of the words in a sentence contain some part of every other word in that sentence. For further detail on the attention mechanism and how BERT captures context, see Appendix B.10.

3.2.2 ESG Definition

Contextualised embedding can be gainfully applied in a classification task - the similarity of sentence vectors can help classify whether a sentence relates to an ESG-specific theme or if it is a part of general content. Before the model can determine that a sentence relates to ESG, I need to establish a clear definition of what an ESG sentence is to label examples. In this application, I label sentences as ESG-related if they relate to the topics outlined in Table 2.

Table 2. Thematic ESG Definition.

This table outlines the main themes I base the definition of ESG on for labelling sentences. The main themes include relevant topics and subtopics. Sentences are labelled as ESG if they relate to one of these themes.

Environment	Social	Governance
Emissions	Workforce Development	Remuneration
Biodiversity	Health and Safety	Board Composition
Waste Management	Data Security	Board Oversight
Water Management	Human Rights	Ethical Behaviour
Renewable Energy	Workforce Diversity	
	Community Contribution	
	Stakeholder Engagement	

3.2.3 Performance Metrics

Before I outline the process for training a BERT model, I will summarise the metrics to assess model performance. Model performance is based on three key measures, precision, recall, and an F1-score. Precision is the ratio of correctly predicted positive examples to the total predicted positive examples. The model's ability to not label a positive example as negative. Recall is the ratio of correctly predicted positive examples to all examples in the class. The model's ability to find all positive examples. The F1-score is the weighted average of the precision and recall scores. Using an F1 score to assess model performance assumes that false negatives and false positive are equally costly. Model performance is task specific, but generally a model with an F1-Score greater than 0.9 would be excellent performance, where the closer the F1-score is to 1, the better.

3.2.4 ESG Classifier

Figure 2 outlines the two-stage approach for training the ESG Classifier with limited training data. This process is based on the method used by Kölbel et al. (2021) to improve performance with limited training data.

In Stage 1, I train the first iteration of the ESG classifier on a hand labelled dataset of 3,263 sentences, with 500 sentences maintained as a test set. Following Kölbel et al. (2021), performance of the model can be improved by running the ESG classifier on a sub-sample of the dataset. Confusing examples are hand labelled to improve the overall accuracy of the model. Confusing examples surround the classification probability of 0.5. Outlined in

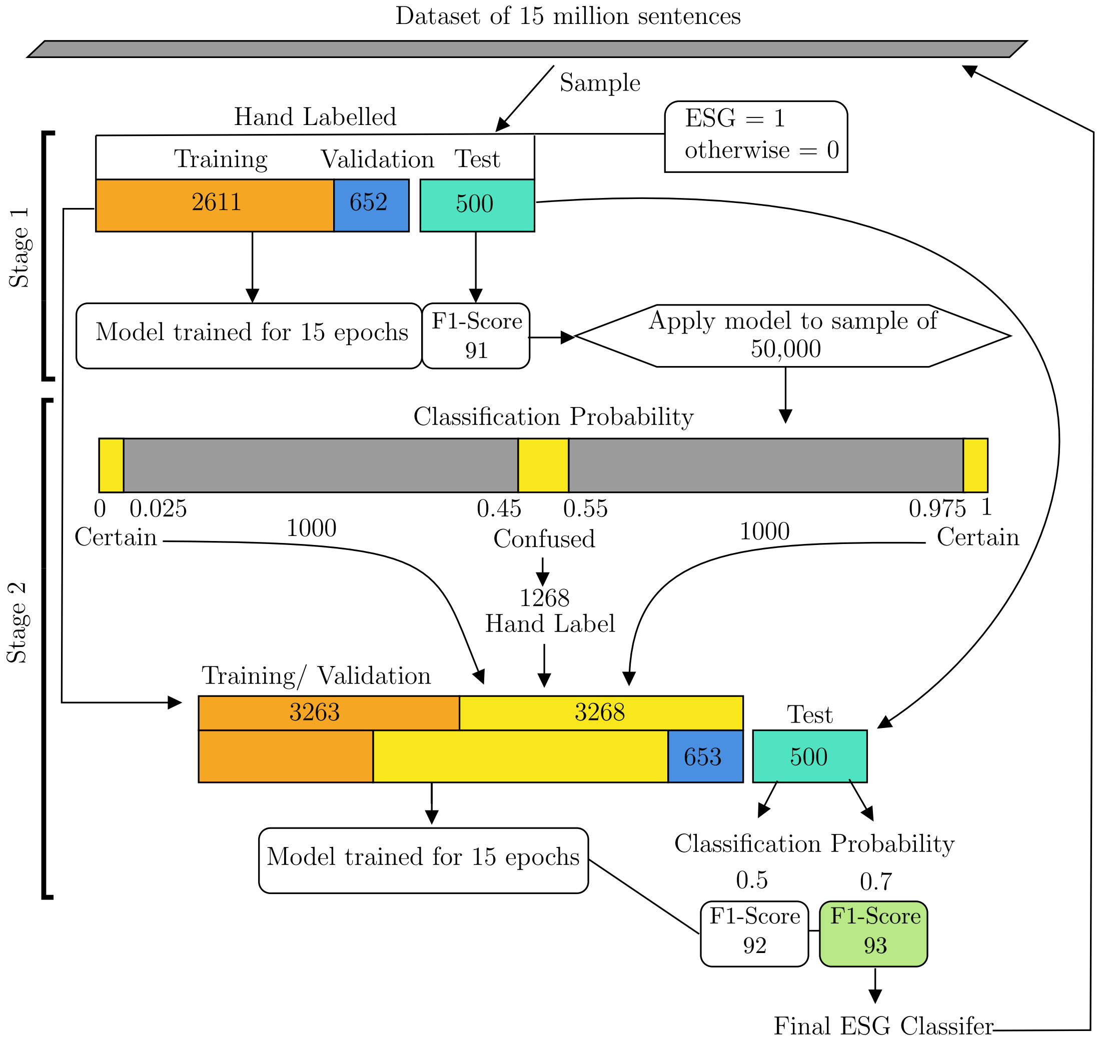


Figure 2. Two stage training procedure.
This figure outlines two stage process implemented to train the ESG classifier. Following Kölbel et al. (2021), the models performance is improved by labelling confusing examples. In Stage 1, the BERT model is trained on a dataset of 3,263 sentences labelled if they are ESG related based on the definition in Section 3.2.2. This Stage 1 model is trained for 15 epochs and has an F1-Score of 0.91. The Stage 1 model is then applied to a sample of 50,000 sentences to classify them into ESG or general content classes. Following in Stage 2, the probability of these examples are examined to highlight confusion and certainty in the model. A new training dataset is created in which contains the Stage 1 training dataset, 1000 certain examples from each class, and 1268 confused examples that were further hand labelled. This new dataset is used to train the Stage 2 ESG classifier for 15 epochs. Based on a 0.7 classification probability the final Stage 2 ESG classifier has an F1-Score of 0.93 and is applied to the full dataset of 15 million sentences.

stage 2, the model classifies a random sub-sample of 50,000 sentences; I hand label 1268 sentences with a classification probability between 0.45 and 0.55. Retraining the model with only the hand labelled confusing examples creates sample bias. Padding the newly labelled data with previous and confident examples reduces this bias. Confident examples are based on a 0.975 (0.025) classification probability, where examples are capped at 1000 to reduce sample bias. I retrain the ESG classifier on the new dataset of 7030 examples. Appendix C contains the hyper-parameters used in the model's training. I use a classification probability of 0.7 to maximise the F1-score on the test dataset. Table 3 contains the results of the final ESG classifier on the test dataset, the results from the Stage 1 iteration and Stage 2 0.5 classification probability are in Appendix D.

Table 3. ESG Classifier Performance.

This table presents the results of the Stage 2 ESG classifier with a 0.7 classification probability. Precision is the ratio of correctly predicted positive examples to the total predicted positive examples. Recall is the ratio of correctly predicted positive examples to all examples in the class. The F1-score is the weighted average of the precision and recall scores. Support is the total number of sentences in each sample class.

		Precision	Recall	F1-Score	Support
General Content	0	0.96	0.96	0.96	407
ESG	1	0.81	0.82	0.81	93
Accuracy				0.93	500
Macro avg		0.88	0.89	0.88	500
Weighted avg		0.93	0.93	0.93	500

The ESG classifier presents an F1-score of 0.93. The 0.7 classification probability balances the errors across each class. For sentences related to general context, the precision and recall scores are both 0.96. While the sentences related to ESG are classified with precision and recall scores of 0.81 and 0.82, respectively. The balanced scores reflect balanced errors within the classification model.

The F1-score is in the middle ground compared to the most comparable papers of Kölbel et al. (2021) and Bingler, Kraus, and Leippold (2021). Each of these papers have their own downfalls with the application of BERT for sentence classification. Kölbel et al. (2021) present an astonishing F1-score of 0.995 and a more realistic score of 0.9027 for their two different tasks. Typically, extremely high F1-scores reflect a model that has seen and memorised the test dataset. Models that memorize the test dataset do not adapt well to the task outside the training environment. With Kölbel et al. (2021) the F1-score of 0.995 presents concerns around the implementation of the method to pad the dataset. This process could have resulted in a model with previous biases to the test dataset. This would occur if Kölbel et al. (2021) were not careful in separating the test set from the padding of data. This is only hypothetical, but the remarkable results raise concerns surrounding the application of the model. Moreover, Bingler, Kraus, and Leippold (2021) has an overall F1-score of 0.81. Again, the key issue comes from the training data. The authors use a significantly large dataset upwards of 317,000 labelled sentences to train their large language model. Their dataset comprised 17,000 human labelled sentences, then they supplemented the remaining data with 300,000 general language sentences from annual reports. This supplementation assumes that annual reports mention nothing related to climate change. This thesis shows that firms use their annual reports to disclose some level of information on climate related issues. Bingler, Kraus, and Leippold (2021) have created a significant amount of noise within their training dataset, which would impede accuracy.

I implemented the Kölbel et al. (2021) method of improving performance with limited training data due to time constraints. The restrictions of this project in terms of time and resources mean some concessions had to be made. As a result, I adopted the method of padding the dataset with confusing examples. This method allows for an effective training dataset without the significant investment of resources. According to best practice, the full dataset would be labelled by different annotators to achieve a level of agreement for the classification of sentences.

3.2.5 ESG Overview

By applying the ESG classifier to the full dataset containing over 15 million sentences, I find that 20.87% of sentences relate to ESG. The following Wordclouds show the most frequent words in each class, excluding common stop words.[6]

The Wordclouds highlight the difference between the ESG and general content classes. The general content classification Wordcloud revolves heavily around accounting terminology, with some of the most common bigrams being financial assets, fair value, and interest rate. Notably, the word director appears in both word clouds, with the ESG sentences having reference to directors on a remuneration and responsibility aspect. The ESG Wordcloud is heavily focused on governance and social aspects. Environmental aspects like climate change make sparse appearances because of their rarity in corporate filings. Appendix G contains further examples of sentences classified as ESG by the model.

[6] Stop words are words that occur frequently and add no value to the interpretation an example list is as follows: "a, an, the, and, but, if, or, because, further, then, once, here, there, when, where, why, how, all, any, . . . "

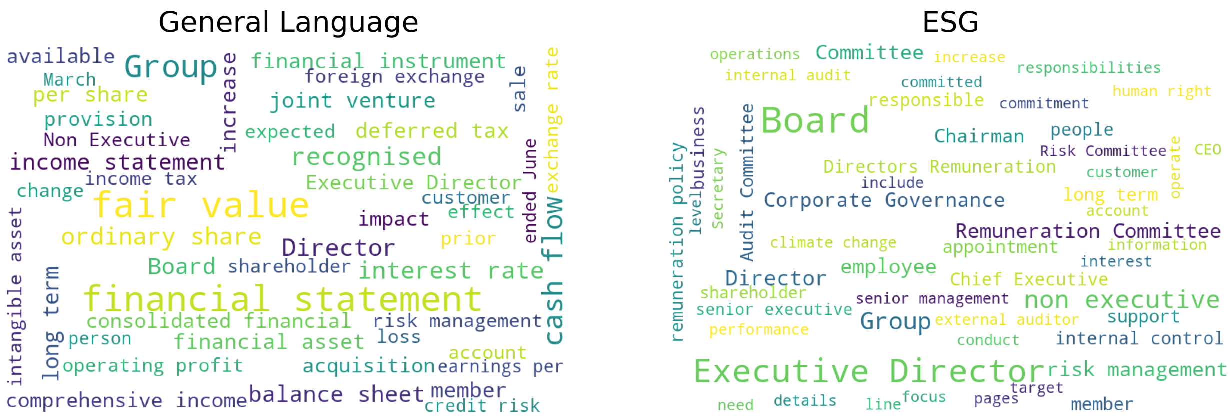


Figure 3. Wordcloud for ESG and general content classes.
This figure presents Wordclouds on the general content and ESG classes from the ESG classifier. The Wordclouds show the most frequent words in each class, excluding common stop words. Stop words include words like "a, an, the, and, but, if, or, because, further, then, once, here, there, when, where, why, how, all, any, ... ". The general content Wordcloud revolves heavily around accounting terminology. The ESG Wordcloud is heavily focused on governance and social aspects.

3.3 Sentiment Analysis

I use FinBERT for sentiment analysis (Araci, 2019). This is a BERT model that is trained to analyse the sentiment in financial texts. Loughran and McDonald (2011) highlight the issue that financial text is unique, and typical NLP models struggle to assign accurate sentiment values. Their approach uses a lexicon-based method, while FinBERT is a BERT model that uses the attention mechanism to capture context. FinBERT scores a very high classification F1-score of 0.97, with data that has a 100% annotator agreement. FinBERT outputs a probability for whether a sentence is negative, neutral, or positive. I calculate the overall sentiment score by taking the difference between the positive and negative weights. The sentiment score is a number between -1 and 1 (negative to positive), with 0 being neutral. I apply FinBERT to the full 15 million sentence dataset. Table 4 highlights some examples of sentiment scores for both ESG and general content sentences.

FinBERT is not a perfect solution for sentiment classification of ESG related sentences. As a result, when companies mention unique internal metrics, it is hard to identify if it is a positive or a negative. Araci (2019) highlights similar limitations in their original paper. The most common example of this is from injury loss time metrics. Ultimately, a reduction in injuries is positive. But from a sentence level perspective, it is impossible to understand the measurement level of injury and loss time metrics and whether a decline is positive or negative. FinBERT assigns these instances negatively, but on review, I would say these sentences are positive. Examples of this misclassification can be found in Appendix H. These limitations are rare and mainly relate to negative classification. The overall sentiment score is the average sentence sentiment for a company in a year. This averaging would reduce nearly all the noise that could affect further analysis using this sentiment analysis. Overall, the true sentiment may be slightly more positive, with negligible impact on results. For this project, FinBERT is satisfactory for sentiment analysis.

Table 4. FinBERT Sentiment Classification Examples.

This table outlines examples of sentiment scores on a small sample of sentences. The sentiment score ranges from -1 to 1, negative to positive, with 0 being neutral. Further examples are in Appendix G.

	ESG Examples
0.93	Environment - The launch of our latest innovation, Caroma Smart Command, an intelligent bathroom system to monitor and manage water in the built environment, further enhances Caromas reputation and commitment to reducing water usage in the built environment.
0.00	The process comprised the Company Secretary issuing a detailed questionnaire covering the Board and its Committees to Board members.
−0.93	Failure to manage these environmental risks properly could result in litigation, regulatory action and additional remedial costs that may materially and adversely affect our financial results.
	General Content Examples
0.94	Our focus on productivity has improved operating performance at each of our Businesses.
0.00	The final dividend for 2017/18 is subject to approval by shareholders at the AGM on 19 July 2018 and will be paid on 15 August 2018 to shareholders on the register at 13 July 2018.
−0.93	and for a portfolio with an exposure of more than $550 billion, the losses in 2018 were $86 million1.

3.4 Variables of Interest

This section outlines the key variables required for this project. Table 5 contains an overview of the required variables organised by data source. This table includes variables constructed from the analysis of textual data, such as sentiment. Further description of the variables can be found below.

Table 5. Variable by data source.

This table outlines the variables used in this project by data source. A detailed description of the variables can be found in Sections 3.4.1 and 3.4.2.

Source	Variable	Description
Corporate Filings	**ESG Content**	Average ratio of ESG sentences to all sentences in a filing for the full year ending December 31.
	ESG Sentiment	Average sentence level sentiment for all ESG sentences in all filings for the full year ending December 31.
	Document Sentiment	Average sentence level sentiment for all sentences in all filings for the full year ending December 31.
	Excess Sentiment	The difference between the average sentence level sentiment for all ESG sentences and the average sentence level sentiment for all general content sentences in all filings for the full year ending December 31.
ESG Scores	**MSCI ESG Score**	A relative measure of a company's resilience to long-term material ESG risks.
	Refinitiv ESG Score	A relative measure of a company's ESG performance focusing on material issues
	ESG Leader	Is a dummy variable where *MSCI* (*Refinitiv*) *Leader* equals 1 if the company is in the top quartile of the ESG score
	ESG Laggard	Is a dummy variable where *MSCI* (*Refinitiv*) *Laggard* equals 1 if the company is in the bottom quartile of the ESG score
Financial Fundamentals	*Large Company*	Is a dummy variable if the company is part of the ASX 50 or FTSE 100 on December 31, 2019.
	Excess Returns	Are the log share price returns less the log market returns for a company year ending December 31.

Note: **Main variable** in bold, *Control variable* in italics

3.4.1 Main Variables

ESG Content is the average ratio between ESG sentences to all sentences in all filings for the year ending December 31. The *ESG Content* is a novel measure of relative ESG content in company filings. *ESG Content* is a ratio between [0,1]. Measuring the *ESG Content* allows me to assess the effectiveness of the UK mandate for ESG reporting with a focus on any increase in content overtime. Using a relative measure over a count for the *ESG Content* was based on comparability. Companies are unique and require different levels of reporting. A relative measure accounts for the inherent differences in the reporting styles. Simply taking a count of ESG sentences would not remove the noise created from companies that release significantly long and boilerplate filings. An overall increase in the *ESG Content* because of the introduction of the mandate would show that mandates are effective at increasing the ESG content in reports.

The *MSCI ESG Score* is a relative measure of a company's resilience to long-term material ESG risks. I scaled the score to a [0,1] range, dividing the score by 10. The *MSCI ESG Score* will be used as a measure to assess the effect of the UK ESG reporting mandate on ESG performance.

The *Refinitiv ESG Score* is a relative measure of a company's ESG performance focusing on material issues. I scaled the score to a [0,1] range, dividing the score by 100. The *Refinitiv ESG Score* will be used as a measure to assess the effect of the UK ESG reporting mandate on ESG performance.

For robustness, examining both MSCI and Refinitiv scores captures the difference in methodologies/ESG definitions that each ratings agency has. Appendix I outlines the differences in the rating methodologies. An increase in the ESG performance in the policy period would show that ESG reporting mandates are effective at increasing ESG performance.

ESG Sentiment is the average sentence level sentiment for all ESG sentences in all filings for the year ending December 31. Sentiment is a measure of the content tonality which quantifies the company portrayed image on ESG. Using *ESG Sentiment*, I can assess if the UK ESG reporting mandate influences the nature of language used in reports. The value ranges between [-1,1] (negative to positive), where ESG sentiment is an indicator of how positive companies are surrounding ESG issues. Evidence suggests companies use disclosure as a legitimizing tool, where companies with poor environmental performance provide offsetting or positive disclosure in their financial reports (Cho and Patten, 2007; Patten, 1992). Companies will frame the language and metrics they use positively to influence investor perception; the sentiment score helps capture this perception. Further, the more positive companies are within their ESG related sentences, the more likely they are using ESG disclosure to better their image in these themes. A company that is more honest and truthful in its ESG disclosure might be more cautionary and conservative, limiting excessively positive statements. Ultimately, *ESG Sentiment* is a measure of company portrayed image for ESG. Assessing whether ESG reporting mandates impact *ESG Sentiment* highlights how effective these mandates are at changing the nature of language used in ESG reporting.

3.4.2 Control Variables

Document Sentiment is the average sentence level sentiment for all sentences in all filings for the year ending December 31. Sentiment is a measure of the content tonality which quantifies the company portrayed image. Using *Document Sentiment*, I can control for any effects of general document tonality that may drive *ESG Sentiment*.

Excess Sentiment is the difference between the average sentence level sentiment for ESG sentences and general content sentences in all filings for the year ending December 31. *Excess Sentiment* quantifies the difference in a companies reporting style between ESG disclosure and general disclosure. I can examine any effects of the UK mandate without the noise from the overall *Document Sentiment*.

Large Company is a dummy variable which equals 1 if the company is part of the ASX 50 or FTSE 100 on December 31, 2019. Index constituents are accessed through from Refinitiv. This method is a simplistic approach, as no historic index constituents where readily available. The *Large Company* variable captures a measure of size through a company's inclusion in a major index relative to their country. Typically, larger companies are more capable of dealing with an ESG risks. The Australian Council of Superannuation Investors (2015) report highlights that the ASX50 leads the way in terms of ESG disclosure compared to their counterparts. I will use this dummy variable to control for any effects that company size may have on my analysis.

Excess Returns are the log share price returns less the log market returns for a company year ending December 31. I base market returns on the ASX 300 and FTSE 350 returns for Australian and UK companies, respectively. This variable will control for any aspect of financial performance, excess of the market. I downloaded the price data from Refinitiv.

ESG Leader is a dummy variable where *MSCI* (*Refinitiv*) *Leader* equals 1 if the company is in the top quartile of the ESG ratings (ESG score greater than 0.75 on the comparable scale). With both MSCI and Refinitiv ratings being used as measures of ESG performance, this dummy with be used independently for both issuers. The ESG leader dummy variable controls for high levels of ESG performance.

ESG Laggard is a dummy variable where *MSCI* (*Refinitiv*) *Laggard* equals 1 if the company is in the bottom quartile of the ESG ratings (ESG score less than 0.25 on the comparable scale). With both MSCI and Refinitiv ratings being used as measures of ESG performance, this dummy with be used independently for both issuers. The ESG leader dummy variable controls for low levels of ESG performance.

3.5 Descriptive Statistics

Table 6 presents the summary statistics and correlation of the variables mentioned in Section 3.4. Panel A reports the summary statistics of ESG disclosure characteristics. Panel B presents Pearson pairwise correlations between variables.

Table 6. Descriptive statistics.

This table presents descriptive statistics of ESG disclosure variables. Panel A reports the summary statistics of ESG disclosure characteristics. Panel B presents Pearson pairwise correlations between variables. Section 3.4 contain definitions of the variables presented in this table.

Panel A. Summary statistics.

	N	μ	σ	Min	10%	50%	90%	Max
ESG Content	1,670	0.21	0.07	0.06	0.13	0.21	0.30	0.49
MSCI ESG Score	1,670	0.64	0.22	0.00	0.36	0.65	0.94	1.00
Refinitiv ESG Score	1,670	0.57	0.18	0.02	0.32	0.58	0.82	0.94
ESG Sentiment	1,670	0.11	0.09	−0.28	0.02	0.11	0.23	0.42
Document Sentiment	1,670	0.11	0.06	−0.11	0.03	0.10	0.19	0.35
Excess Sentiment	1,670	0.01	0.07	−0.38	−0.06	0.02	0.09	0.35
Excess Returns	1,670	0.02	0.29	−1.86	−0.29	0.04	0.31	2.63
Large Company	1,670	0.50	0.50	0.00	0.00	0.00	1.00	1.00
MSCI Leader	1,670	0.35	0.48	0.00	0.00	0.00	1.00	1.00
MSCI MSCI Laggard	1,670	0.05	0.21	0.00	0.00	0.00	0.00	1.00
Refinitiv Leader	1,670	0.19	0.39	0.00	0.00	0.00	1.00	1.00
Refinitiv Laggard	1,670	0.04	0.18	0.00	0.00	0.00	0.00	1.00

Panel B. Pairwise correlations.

		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
ESG Content	(a)											
MSCI ESG Score	(b)	0.21										
Refinitiv ESG Score	(c)	0.22	0.37									
ESG Sentiment	(d)	0.13	0.11	0.12								
Document Sentiment	(e)	0.17	0.10	0.09	0.67							
Excess Sentiment	(f)	0.03	0.05	0.08	0.69	−0.05						
Excess Returns	(g)	0.04	0.06	−0.02	0.06	0.10	−0.02					
Large Company	(h)	−0.02	0.01	0.06	0.02	0.00	0.03	0.02				
MSCI Leader	(i)	0.08	0.15	0.04	0.07	0.07	0.03	0.03	0.00			
MSCI Laggard	(j)	−0.11	−0.08	−0.07	−0.04	−0.04	−0.02	0.03	−0.09	−0.16		
Refinitiv Leader	(k)	0.02	0.00	0.16	0.03	0.01	0.04	−0.02	0.42	0.21	−0.10	
Refinitiv Laggard	(l)	−0.08	−0.04	−0.07	−0.02	−0.01	−0.02	0.03	−0.16	−0.08	0.14	−0.09

4 Method

I exploit the implementation of mandatory ESG disclosure for UK companies to provide a quasi-natural experiment. Using Australian companies as the control group, both jurisdictions have the investor push for ESG disclosure, but the UK has mandated ESG disclosure on a comply or explain basis. The comparison will highlight the effectiveness of the UK mandate for ESG disclosure. Australia makes for an excellent control group because of the similarities reporting standards and overall market size. In 2019, the FSTE 350 index total market capitalisation was \$2.7 billion compared to the ASX 300 index of \$2.2 billion.

I investigate the effectiveness of the UK mandate on ESG reporting through a difference-in-differences regression framework. The regressions will follow Equation (1):

$$Y_{i,t} = \beta_0 + \beta_1 Post_t + \beta_2 Treatment_i + \boldsymbol{\beta_3 Post_t \cdot Treatment_i} + \gamma Controls_{i,t} + \epsilon_{i,t}, \quad (1)$$

where, $Post$ and $Treatment$ are dummy variables defined as:

$$Post \begin{cases} 1, & \text{if Year} \geq 2014 \\ 0, & \text{otherwise} \end{cases}$$

$$Treatment \begin{cases} 1, & \text{if UK Company} \\ 0, & \text{if Australian Company} \end{cases}$$

In Equation (1), the $Post \cdot Treatment$ variable is an interaction term that represents the UK companies after the treatment period (Henceforth referred to as the policy group).

I examine the relation between *ESG content* (*ESG sentiment*), the dependent variables Y and the policy group, using identical controls. I control for company size is through the *Large Company* dummy variable. Financial performance is controlled for through *Excess Returns*. ESG performance is controlled for through the *MSCI* (*Refinitiv*) *Leader/Laggard* dummy variables.

I examine the relation between ESG performance, *MSCI* (*Refinitiv*) *ESG Scores* as the dependent variables Y and the policy group with similar controls. I control for company size through the *Large Company* dummy variable. Financial performance is controlled for through *Excess Returns*. The nature of ESG disclosure could drive ESG performance. Considering this, I control for both *ESG Content* and *ESG Sentiment* in this regression.

5 Main Results

5.1 UK Mandate effect on ESG Content

I first examine the effect that the UK mandate for ESG reporting has on *ESG Content* by visual inspection. Figure 4 presents a comparison of country *ESG Content* by kernel density estimations for each year between 2010 and 2019. UK is represented by orange and Australia by blue. The respective orange and blue dotted lines represent the pre and post period means for UK and Australia. Visually, there is an upward trend in the overall distribution of *ESG Content*. Examining the difference in means there appears to be no effect on the policy group. This does not support my hypothesis that the UK ESG reporting mandate increases the ESG content in corporate filings. Although, there is a visible difference in variance in the pre and post period for Australia.

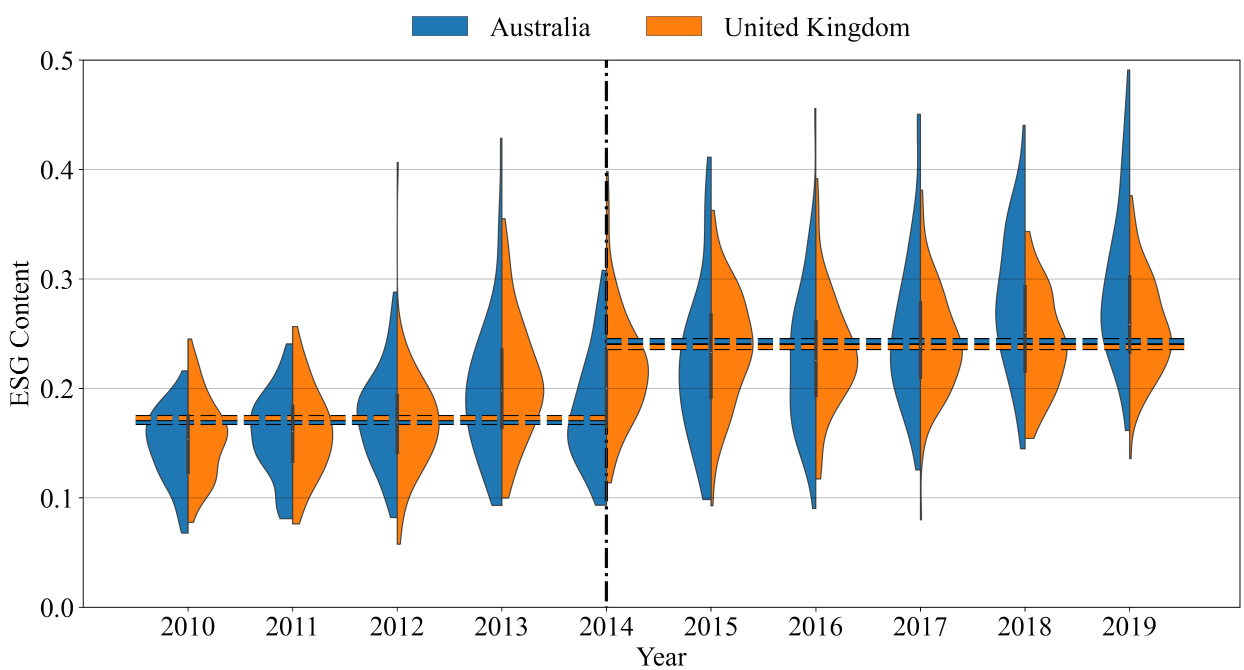


Figure 4. ESG Content.
This figure shows the kernel density estimations of *ESG Content* in each year for each company, with Australia in blue and the UK in orange. *ESG Content* is the average ratio of ESG sentences to all sentences in a filing for the year ending December 31. The pre and post periods are separated by the black dash dotted line in 2014. The orange and blue dotted lines represent the mean *ESG Content* for that jurisdiction in the pre and post years, for the sample period between 2010 and 2019.

To test this relation statistically, I employ a beta regression based on Equation (1). The variable of interest is the $Post \cdot Treatment$, it represents the difference in means for the policy group. In other terms, the effect the UK ESG reporting mandate has on the ESG content in corporate filings. Table 7 presents the results for Regressions (1), (2), and (3). The z-scores are presented in brackets.

Table 7. UK ESG reporting mandate effect on ESG content.

This table presents estimates from Beta regressions based on Equation (1). The dependent variable is *ESG Content*. The *ESG Content* of corporate disclosure is the ratio of ESG sentences to all sentences in a filling averaged over the year. The key independent variable is $Post \cdot Treatment$, it represents the effect that the UK ESG reporting mandate has on *ESG Content*. Sections 3.4 contains variable definitions. z-scores are reported in parentheses.

	ESG Content		
Mean	(1)	(2)	(3)
Intercept	−1.58***	−1.57***	−1.56***
	(−71.06)	(−65.37)	(−64.01)
Post	0.43***	0.42***	0.42***
	(15.72)	(15.41)	(15.18)
Treatment	0.02	0.02	0.01
	(0.85)	(0.92)	(0.61)
$Post \cdot Treatment$	−0.02	−0.03	−0.02
	(−0.81)	(−1.02)	(−0.62)
Large Company		−0.02	−0.01
		(−1.42)	(−1.01)
Excess Returns		0.02	0.02
		(0.96)	(0.99)
MSCI Leader		0.05***	
		(2.89)	
MSCI Laggard		−0.12***	
		(−3.08)	
Refinitiv Leader			0.00
			(0.28)
Refinitiv Laggard			−0.08*
			(−1.79)
Precision ϕ	51.58***	52.28***	51.73***
	(29.07)	(29.06)	(29.07)
Observations	1670	1670	1670
Pseudo R^2	0.27	0.28	0.27

Significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Supporting the conclusions from the visual inspection, the results from the regressions presented in Table 7 show that there is an insignificant effect with the policy group. $Post \cdot Treatment$ coefficient is insignificant, which does not support my hypothesis that the UK ESG reporting mandate increases the ESG content in corporate filings. ESG reporting mandate competes with an investor push for information. The regulatory goal for increased content in reports assumes firms would otherwise ignore the needs of their shareholders and investors. The mandates may simply add a minimum level of disclosure for a company.

The $Post \cdot Treatment$ coefficient is consistent when controlling for size, financial performance, and ESG performance in Regressions (2) and (3). Notably, there appears to be a positive relation between *ESG Content* and MSCI ESG performance as shown in Regression (2). This relation is only apparent for MSCI and not Refinitiv in Regression (3), although the *Refinitiv Laggard* control variable highlights slightly significant a consistent negative relation between low ESG performance and the *ESG Content* in reports.

The visual difference in variances in Figure 4 can be examined using a Levene test for equality of variances (Levene, 1960). Comparing the UK and Australian pre and post period variance, I can test whether the ESG reporting mandate has an effect of the overall distribution of *ESG Content*. The UK group presents a Levene test statistic of 0.11, where I fail to reject that the variance pre and post period is equal. The Australian group presents a Levene test statistic of 48.16, where I reject that the variance pre and post period is equal. In Figure 4 the variance increases in the post period. The level of UK ESG reporting is growing, but the dispersion is consistent, unlike the control group in Australia. Ultimately, introducing mandatory ESG reporting has made reporting more homogeneous and standardised year on year. Even though there is no apparent difference in means for *ESG Content* from the UK reporting mandate. The difference in variance highlights that reporting is more standardised.

I view this result in two ways. First, I could argue that mandates of this fashion are not effective because, on average, they do not add additional information to disclosure than there would have been otherwise. On the other hand, investors require the information, whether it is mandated or not. The Levene test for equality of variances shows that, in Australia, without mandates, the level of ESG content between companies varied significantly in the pre and post periods. Whilst in the UK, the level of ESG content stayed relatively homogeneous. Companies are unique and have different exposures to ESG risks. This also means they should have different levels of required disclosures to inform investors about their specific ESG risks. Investors should be able to gauge the extent of these risks that are unique to their personal situation. With homogeneous ESG reporting, disclosure is more comparable between companies. The increased comparability ultimately enhances investors' decision making.

There is notable discussion surrounding ESG reporting and its effectiveness. Primarily, it helps drive disclosure and therefore inform ESG conscious investors about the key ESG risks that face a company they in invest. This study sheds light on some of the biggest concerns surrounding ESG disclosures, such as its effectiveness in communicating useful and material information to investors. Central to effective communication is the idea of materiality. Over reporting on issues that are not material has led to the argument that ESG reporting mandates will lead to further complexity and noise in corporate reports.

The argument that ESG reporting mandates increase complexity in reports is unsupported by my results. Corporate disclosure is heavily criticised as boilerplate and uninformative. Where some argue that ESG reporting mandates will exacerbate this issue, and add further complexity to corporate filings. Regardless of the mandate, the ESG content in both Australia and the UK is growing overtime. Complexity cannot be increased by mandates if the levels of ESG related content remain consistent across jurisdictions. Investors drive information requirements, not mandates.

The added certainty from the UK ESG mandate is important in creating comparable disclosure. From a company perspective, ESG mandates add more considerations to the disclosure process. But, mandates standardise the levels of reporting and create certainty for companies. In a free market approach, firms cannot be certain that the information being disclosed is enough to satisfy investors. individual Investors require different levels of information and as a consequence investors are flooded with either overzealous or subpar disclosure. This notion is supported by results in the Levene test and visually through the increase in the variance of ESG content for Australian companies. The added certainty from the ESG reporting mandate allows for comparability. It reduces both over zealous and sub par ESG disclosure. This is consistent with the idea that ESG disclosure mandates play an important role as an intermediary between investors and companies.

Existing studies that assess ESG reporting mandates present different results. Previous attempts at assessing the UK ESG reporting mandate find that key performance indicators and narrative disclosure increase after the UK mandate was introduced (Hummel and Rötzel, 2019). My results suggest that the ESG content in does not increase after the mandate. The prior attempt employs weak textual analysis techniques using a keyword search. In comparison using BERT allows for more accurate analysis, capturing context and reducing noise. Ioannou and Serafeim (2017) examine mandates in developing countries, they find evidence to suggest that sustainability disclosure increases with disclosure mandates. My sample relates to developed markets with more mature reporting standards. The differing results suggest that the effectiveness of ESG reporting mandates may depend on the general market environment. Where the maturity of reporting could impact the effectiveness of ESG reporting mandates.

Overall, my results do not support my hypothesis that ESG reporting mandates increases the ESG content in corporate fillings. But, my results suggest that mandates have increased the comparability of unique firms. ESG disclosure mandates play an important role as an intermediary between investors and companies, ensuring comparability for investors and certainty for companies.

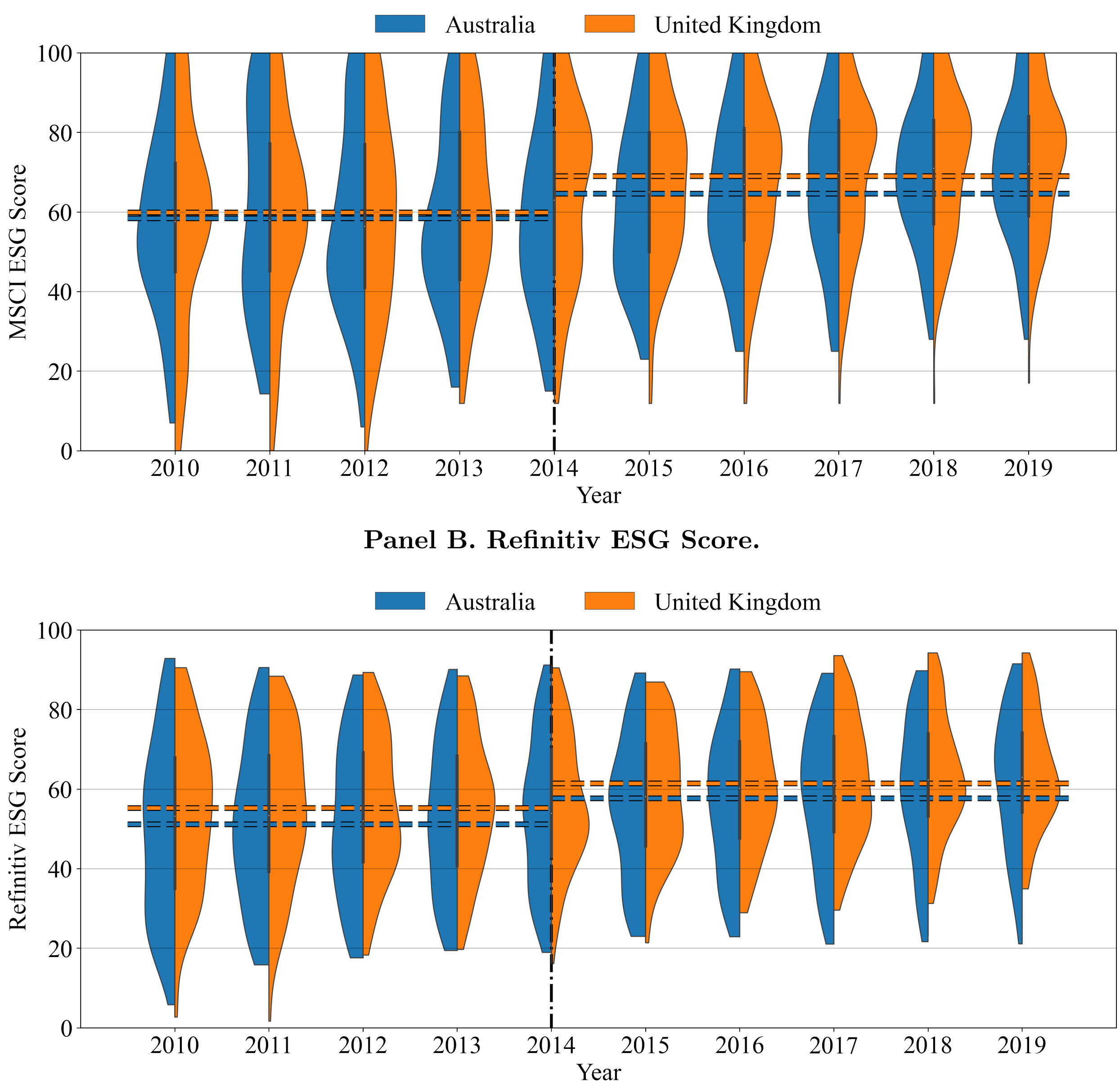
5.2 UK Mandate effect on ESG Performance

I first examine the effect that the UK mandate for ESG reporting has on ESG performance by visual inspection. Figure 5 presents a comparison of country ESG performance by kernel density estimations for each year between 2010 and 2019 for MSCI in Panel A and Refinitiv in Panel B. UK is represented by orange and Australia by blue. The respective orange and blue dotted lines represent the pre and post period means for UK and Australia. Visually, there is a slight upward trend in the distribution of *MSCI* (*Refinitiv*) *ESG scores*. Examining the difference in means there appears to be no effect on the policy group. This does not support my hypothesis that the UK ESG reporting mandate improves the ESG performance. The comparative make-up of the scoring methodology is evident through the distribution of both scores. Notably, *Refinitiv ESG Score* cuts off at bounds under the limits unlike MSCI.

To test this relation statistically, I employ a beta regression based on Equation (1). The variable of interest is the $Post \cdot Treatment$ coefficient, it represents the difference in means for the policy group. In other terms, the effect the UK ESG reporting mandate has on ESG Performance. Table 8 presents the results for Regressions (4), (5), (6), and (7). The z-scores are presented in brackets.

Supporting the conclusions from the visual inspection, the results from the regressions presented in Table 8 show that there is an insignificant effect with the policy group. $Post \cdot Treatment$ coefficient across all regressions is insignificant, which does not support my hypothesis that the UK ESG reporting mandate improves the ESG performance. There is a strong relation between the *ESG Content* in reports and ESG performance, robust to both MSCI and Refinitiv's measures of performance. The results of Regressions (1), (2), and (3) in Table 8 highlight that *ESG Content* is unchanged by the UK reporting mandate. The insignificant relation between performance and the policy group could be because of the insignificant relation between ESG content and the policy group. Examining Regressions (4) and (5) for *MSCI ESG Scores*, there is a significant relation between company size and the *ESG Content*. This supports the notion that larger companies perform better when it comes to ESG. Examining Regressions (6) and (7) for *Refinitiv ESG Scores*, there is a significant

Panel A. MSCI ESG Score.



Panel B. Refinitiv ESG Score.

Figure 5. ESG Performance.
This figure shows the distribution of both MSCI, Panel A and Refinitiv, Panel B ESG scores in each year. The figure shows the kernel density estimation of ESG scores for each company, with Australia in blue and the UK in orange. The *MSCI ESG Score* is a relative measure of a company's resilience to long-term material ESG risks. The *Refinitiv ESG Score* is a relative measure of a company's ESG performance focusing on material issues. The pre and post periods are separated by the black dash dotted line in 2014. The orange and blue dotted lines represent the mean ESG Score for that jurisdiction in the pre and post years.

Table 8. UK ESG reporting mandate effect on ESG performance.

This table presents estimates from Beta regressions based on Equation (1). The dependent variables are the *MSCI ESG Score* and *Refinitiv ESG Score*. the *MSCI ESG Score* and *Refinitiv ESG Score* are measures of ESG performance. The key independent variable is $Post \cdot Treatment$, it represents the effect that the UK ESG reporting mandate has on ESG Performance. Section 3.4 contains variable definitions. z-scores are reported in parentheses.

	MSCI ESG Score		**Refinitiv ESG Score**	
Mean	(4)	(5)	(6)	(7)
Intercept	0.73***	0.04	0.05	−0.58***
	(9.56)	(0.39)	(1.26)	(−9.40)
Post	0.18*	−0.02	0.25***	0.14***
	(1.85)	(−0.23)	(4.57)	(2.79)
Treatment	0.06	0.00	0.15***	0.10*
	(0.69)	(−0.03)	(2.68)	(2.07)
$Post \cdot Treatment$	0.06	0.10	−0.01	0.01
	(0.49)	(0.80)	(−0.09)	(0.21)
Large Company		0.54***		0.77***
		(9.06)		(25.02)
Excess Returns		−0.03		−0.15***
		(−0.36)		(−2.93)
ESG Sentiment		0.51		0.41**
		(1.46)		(2.29)
ESG Content		2.43***		1.45***
	(4.72)		(5.35)	
Precision ϕ	1.01***	1.08***	6.88***	9.68***
	(32.53)	(32.27)	(30.82)	(30.23)
Observations	1670	1670	1670	1670
Pseudo R^2	0.00	0.02	0.04	0.31

Significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

relation with all control variables. The strongest relation is between company size, followed by the *ESG Content*. This further supports the notion that larger companies perform better when it comes to ESG. Notably, there is a negative relation with *Excess Returns* and a positive relation with *ESG Sentiment* in Regression (7). Giving inference on the quality of the scoring methodology rather than the impact of the mandate.

Assessing the variance as done for the *ESG Content* is not appropriate. ESG scores are based on a comparison to peers. The variance within the ESG ratings is not an effective indicator of a distribution of performance.

My results show that an increase in ESG reporting is positively related to ESG performance. This is consistent with the view that more ESG information allows ratings agencies and investors to better assess a company's ESG performance. Further evidence highlights that the effect of the mandatory reporting on ESG performance is negligible. The insignificant relation of ESG performance could be a symptom of the insignificant relation between the ESG content and the UK ESG disclosure mandate. If the UK mandate influenced the level of ESG content in reports, it could reflect through in the ESG performance measure.

The UK mandate is based on the idea of "comply or explain." Although disclosure has become forced, there is no direct link to ESG performance, as requirements are flexible. An aim of disclosure mandates is to make companies aware of the ESG issues they face. Central to the effectiveness of mandates is the idea that self awareness may improve performance. This argument would have merit if the information disclosed after the mandate is new, but this is not the case. Prior to the mandate, investors would have made assumptions about the ESG risks that face a company. Where the addition of mandates only adds certainty of information for investors. Consistent with the idea that mandates create certainty for companies, investors also get certainty from mandates.

Overall, my results do not support my hypothesis that the ESG reporting mandate improve ESG reporting. ESG scores are a fuzzy measure of ESG performance. By design, they are a standardised measure based on a comparison to peers. Regardless, my results suggest investors gain certainty from ESG reporting mandates.

5.3 UK Mandate effect on ESG Sentiment

I first examine the effect that the UK mandate for ESG reporting has on *ESG Sentiment* by visual inspection. Figure 6 presents a comparison of country *ESG Sentiment* by kernel density estimations in each year between 2010 and 2019. UK is represented by orange and Australia by blue. The respective orange and blue dotted lines represent the pre and post means for UK and Australia. Visually, the difference in means is clearly driven by an increase in *ESG Sentiment* among Australian companies. There is no difference in the mean UK *ESG Sentiment*. It is apparent any effect from the regression will be driven by movement in Australian *ESG Sentiment*. This would show that the effect is not driven by the mandate, where I cannot support my hypothesis that ESG reporting mandates change the content tonality in a firms' ESG disclosure. The variance in the *ESG Sentiment* for UK companies appears to decrease after the post period.

Regardless, to test this relation statistically, I employ a panel regression based on Equation (1). The panel regression controls for any industry effects that are prevalent in the *ESG Sentiment*. The variable of interest is the $Post \cdot Treatment$ coefficient, it represents the

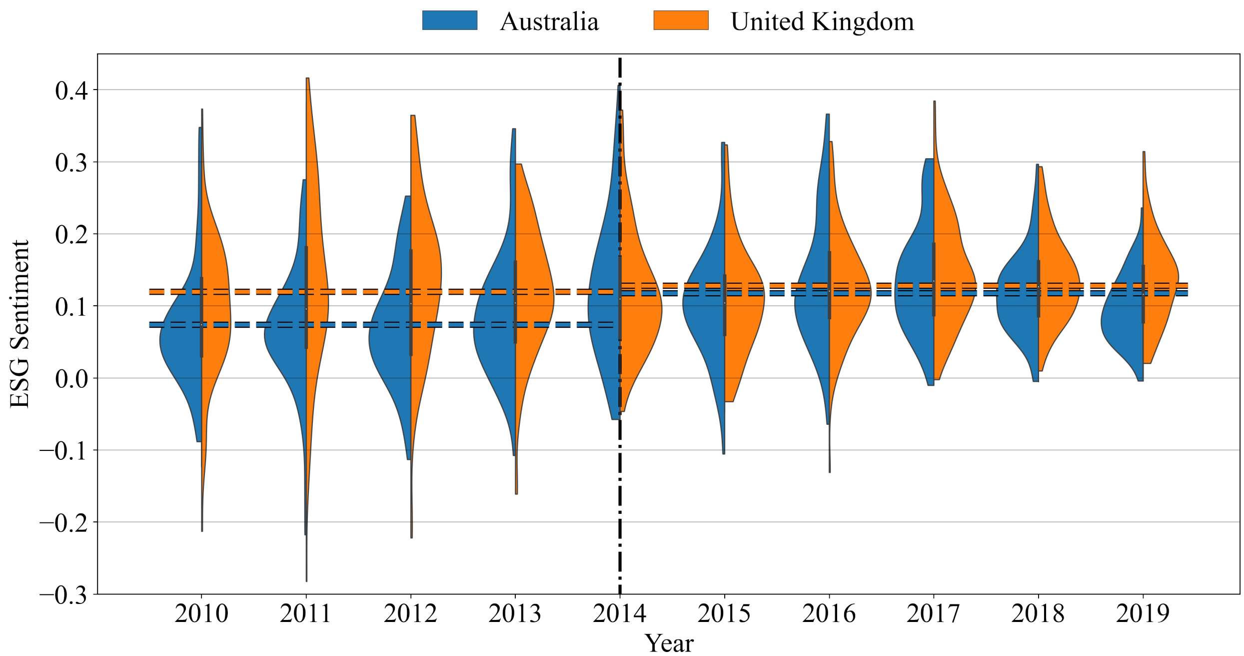


Figure 6. ESG Sentiment.
This figure shows the kernel density estimations of *ESG Sentiment* in each year for each company, with Australia in blue and the UK in orange. The *ESG Sentiment* is calculated by taking the average sentiment from company reports for the year end date December 31. The pre and post periods are separated by the black dash dotted line in 2014. The orange and blue dotted lines represent the mean *ESG Sentiment* for that jurisdiction in the pre and post years, for the sample period between 2010 to 2019.

difference in means for the policy group. In other terms, the effect the UK ESG reporting mandate has on *ESG Sentiment*. Table 9 presents the results for Regressions (8), (9), and (10). The t-statistics are presented in brackets.

The results from the regressions presented in Table 9 show that the $Post \cdot Treatment$ coefficient is significant and supports my hypothesis that ESG reporting mandates change the content tonality (sentiment) in firms ESG disclosure. The visual inspection highlights that this effect is driven by movement in Australia and not the UK. It is not certain that this effect is driven by an external factor specific to Australian companies. The $Post \cdot Treatment$ coefficient is consistent when controlling for size, financial performance, ESG performance and industry effects in Regressions (9) and 10. Notably, there are countering factors within the different ratings where being a *Refinitiv Leader* and a *MSCI Laggard* has a significant negative relation with *ESG Sentiment*. All other control variables are insignificant in Regressions (9) and (10).

A change in the overall language being used in disclosure could explain why the average

Table 9. UK ESG reporting mandate effect on ESG Sentiment.

This table presents estimates from panel regressions based on Equation (1). The dependent variable is *ESG Sentiment*. The *ESG Sentiment* of corporate disclosure is the average sentiment of ESG sentences. The key independent variable is $Post \cdot Treatment$, it represents the effect that the UK ESG reporting mandate has on *ESG Sentiment*. Section 3.4 contains variable definitions. t-statistics are reported in parentheses.

	ESG Sentiment		
	(8)	(9)	(10)
Intercept	0.07***	0.07***	0.07***
	(14.28)	(13.18)	(13.08)
Post	0.04***	0.04***	0.04***
	(6.59)	(6.51)	(6.49)
Treatment	0.04***	0.05***	0.04***
	(6.94)	(7.39)	(7.12)
$Post \cdot Treatment$	−0.03***	−0.04***	−0.03***
	(−4.14)	(−4.25)	(−4.08)
Large Company		−0.01	0.00
		(−1.15)	(−0.1)
Excess Returns		0.01	0.00
		(0.75)	(0.67)
MSCI Leader		0.00	
		(0.02)	
MSCI Laggard		−0.02**	
		(−2.18)	
Refinitiv Leader			−0.01**
			(−2.33)
Refinitiv Laggard			−0.01
			(−1.45)
Entity Effects		Industry	Industry
Observations	1670	1670	1670
Adjusted R^2	0.04		
R^2 Within		0.05	0.06

Significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

ESG Sentiment for Australia increases around the UK Mandatory ESG reporting period. Consistent with the idea that firms started became aware that the language they use in reports is being analysed for sentiment and tonality. Cao et al. (2020) highlight the Loughran and McDonald (2011) paper as a significant turning point for machine readability. Assuming some delay in response from Australian companies, an increase in overall *Document Sentiment* could be give merit to this argument. Visual examination of *Document Sentiment* in Figure 7 presents a the comparison of country *Document Sentiment* by kernel density estimations in each year between 2010 and 2019. UK is represented by orange and Australia by blue. The respective orange and blue dotted lines represent the pre and post period means for UK and Australia.

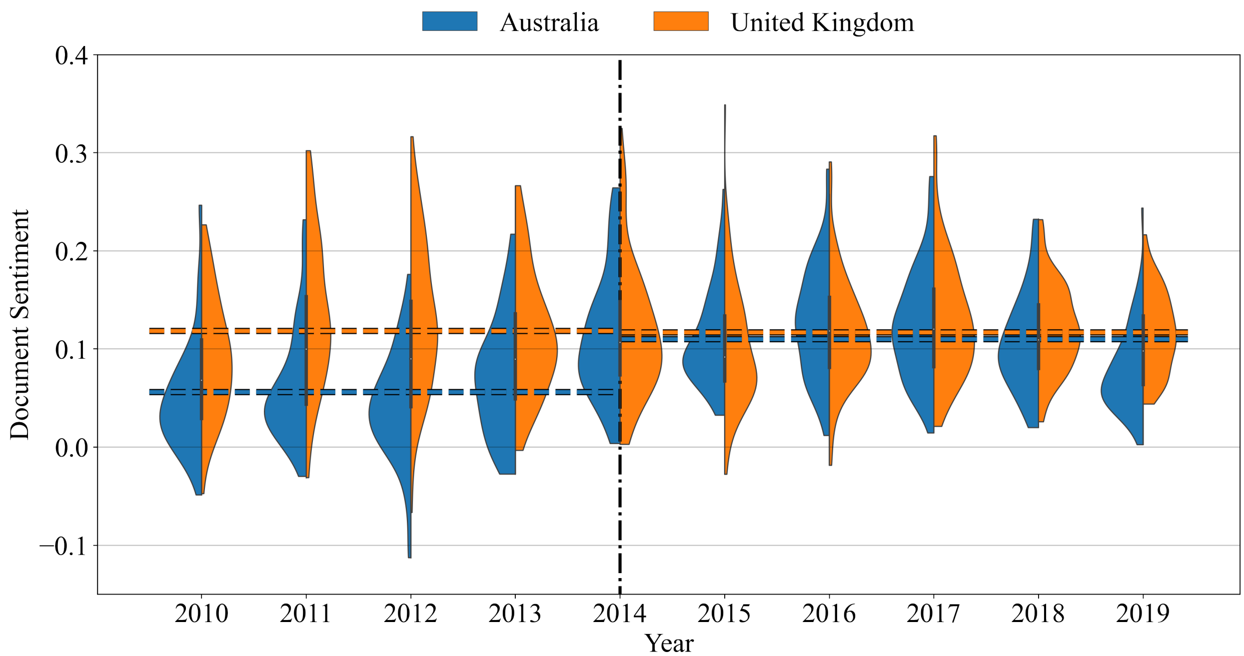


Figure 7. Document Sentiment.
This figure shows the kernel density estimations of *Document Sentiment* in each year for each company, with Australia in blue and the UK in orange. The *Document Sentiment* is calculated by taking the average sentiment from company reports for the year end date December 31. The pre and post periods are separated by the black dash dotted line in 2014. The orange and blue dotted lines represent the mean *Document Sentiment* for that jurisdiction in the pre and post years, for the sample period between 2010 to 2019.

Visually, there is a significant difference in the pre, and post means for Australia, a similar relation was observed with the *ESG Sentiment*. To test this relation statistically, I employ a panel regression based on Equation (1). The panel regression controls for any industry effects that are prevalent in the *Document Sentiment*. The variable of interest is the $Post \cdot Treatment$ coefficient, it represents the difference in means for the policy group. In other terms, the effect the UK ESG reporting mandate has on overall *Document Sentiment*. Table 10 presents the results for Regressions (11), (12), and (13). The t-statistics are presented in brackets.

The results from the regressions presented in Table 10 show that there is a significant negative relation between *Document Sentiment* and the policy group. $Post \cdot Treatment$ coefficient results in a 90% standard deviation move in the dependent variable. It is clear visually that this is specifically driven by a factor unique to Australia. The effect on the

Table 10. UK ESG reporting mandate effect on document sentiment.

This table presents estimates from panel regressions based on Equation (1). The dependent variable is *Document Sentiment*. The *Document Sentiment* of corporate disclosure is the average sentiment of sentences in a filing. The key independent variable is $Post \cdot Treatment$, it represents the effect that the UK ESG reporting mandate has on *Document Sentiment*. Section 3.4 contains variable definitions. t-statistics are reported in parentheses.

	Document Sentiment		
	(11)	(12)	(13)
Intercept	0.06***	0.06***	0.06***
	(15.33)	(15.04)	(14.57)
Post	0.05***	0.05***	0.05***
	(11.41)	(11.67)	(11.64)
Treatment	0.06***	0.06***	0.06***
	(6.94)	(12.57)	(12.43)
$Post \cdot Treatment$	−0.06***	−0.05***	−0.03***
	(−9.16)	(−9.28)	(−9.30)
Large Company		0.00	0.00
		(−0.55)	(0.62)
Excess Returns		0.01	0.02
		(1.97)**	(1.88)*
MSCI Leader		−0.01	
		(−1.51)	
MSCI Laggard		−0.01	
		(−1.19)	
Refinitiv Leader			−0.01***
			(−3.05)
Refinitiv Laggard			0.00
			(−0.56)
Entity Effects		Industry	Industry
Observations	1670	1670	1670
Adjusted R^2	0.12		
R^2 Within		0.12	0.12

Significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

ESG Sentiment is likely to be driven by the general tonality in the document.

To control for any effect that the general tone of a firm's disclosure has on the tonality of ESG sentences, I take the difference in *ESG Sentiment* and general content sentiment. This gives a measure of *Excess Sentiment*, or how the positive firms *ESG Sentiment* is greater than the general content sentiment. Where the difference will be largely positive if a firm is more positive in ESG disclosure than general financial disclosure.

Figure 8 presents a comparison of country *Excess Sentiment* by kernel density estimations in each year between 2010 and 2019. UK is represented by orange and Australia by blue. The respective orange and blue dotted lines represent the pre and post period means for UK and Australia. Visually, there is a negligible difference in the pre and post means.

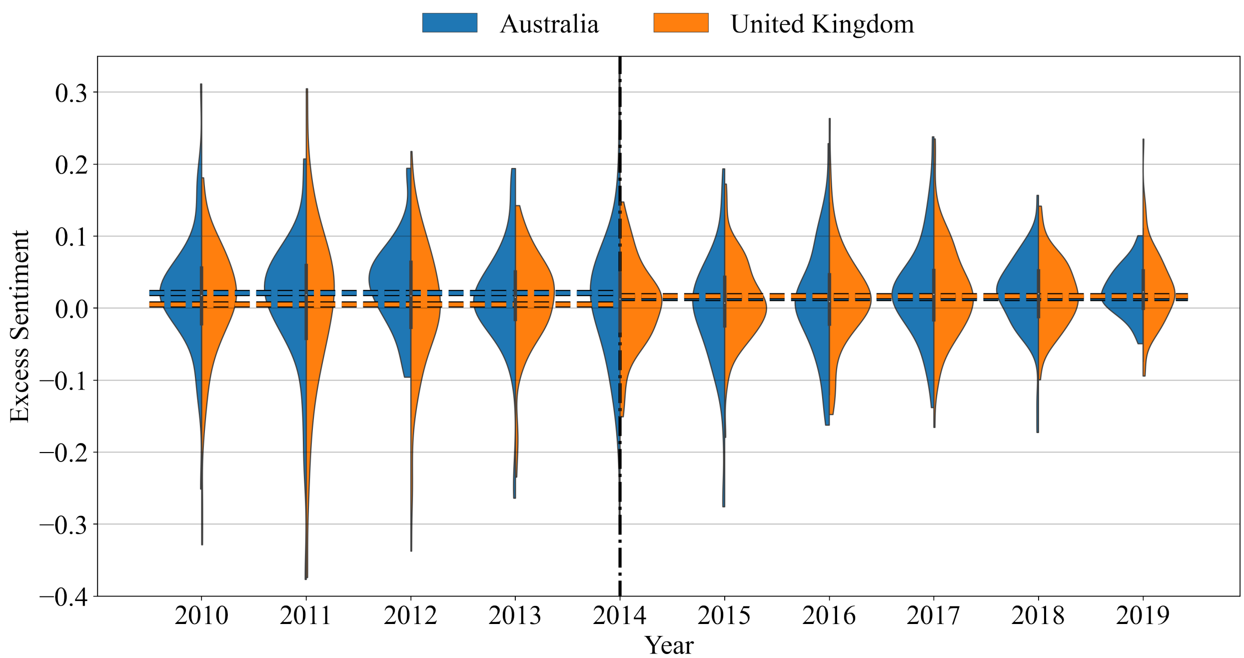


Figure 8. Excess Sentiment.
This figure shows the kernel density estimations of the difference in *Excess Sentiment* (The difference between *ESG Sentiment* and general content sentiment) in each year for each company. Australia is in blue and the UK is in orange. The sentiment scores are calculated by taking the average sentiment from company reports for the year end date December 31. The pre and post periods are separated by the black dash dotted line in 2014. The orange and blue dotted lines represent the mean difference in sentiment for that jurisdiction in the pre and post years, for the sample period between 2010 to 2019.

To test this relation statistically, I employ a panel regression based on Equation (1). The panel regression controls for any industry effects that be prevalent in the *Excess Sentiment*. The variable of interest is the $Post \cdot Treatment$ coefficient, it represents the difference in means for the policy group. In other terms, the effect the UK ESG reporting mandate has on the difference in *ESG Sentiment* and general content sentiment. Table 11 presents the results for Regressions (14), (15), and (16). The t-statistics are presented in brackets.

The results from the regressions presented in Table 11 show that there is a significant positive relation between *Excess Sentiment* and the policy group. $Post{\cdot}Treatment$ coefficient results in a 25% standard deviation increase in the dependent variable. The UK mandate for ESG disclosure has made firms more positive in their ESG disclosure relative to the tonality

Table 11. UK ESG reporting mandate effect on excess sentiment.

This table presents estimates from panel regressions based on Equation (1). The dependent variable is *Excess Sentiment*. The *Excess Sentiment* of corporate disclosure is the average difference between the *ESG Sentiment* and general content sentiment. The key independent variable is $Post \cdot Treatment$, it represents the effect that the UK ESG reporting mandate has on *Excess Sentiment*. Section 3.4.1 contains variable definitions. t-statistics are reported in parentheses.

	Excess Sentiment		
	(14)	(15)	(16)
Intercept	0.02***	0.02***	0.02***
	(4.76)	(3.66)	(4.07)
Post	−0.01	−0.01	−0.01***
	(−1.19)	(−1.38)	(−1.36)
Treatment	−0.02***	−0.01	−0.01*
	(−2.83)	(−1.49)	(−1.71)
$Post \cdot Treatment$	0.02***	0.02***	0.02***
	(2.49)	(2.32)	(2.57)
Large Company		0.00	0.00
		(−1.11)	(−0.83)
Excess Returns		0.00	0.00
		(−0.71)	(−0.70)
MSCI Leader		0.01**	
		(1.89)	
MSCI Laggard		−0.01	
		(−1.51)	
Refinitiv Leader			0.00
			(−0.09)
Refinitiv Laggard			−0.01
			(−1.35)
Entity Effects		Industry	Industry
Observations	1670	1670	1670
Adjusted R^2	0.04		
R^2 Within		0.01	0.01

Significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

in the overall document. The $Post \cdot Treatment$ coefficient is consistent when controlling for size, financial performance, ESG performance and industry effects in Regressions (15) and (16). ESG disclosure mandates force companies into disclosure, in which a company may take the opportunity to be more positive. Portraying the company image in a way that attracts ESG conscious investors.

Visually, there is a notable difference in the variance of *Excess Sentiment* overtime in Figure 8. The change in variance can be examined using a Levene test for equality of variances (Levene, 1960). Comparing UK and Australia pre and post period variance, I can test whether the ESG reporting mandate influences the overall distribution of Excess Sentiment. The UK group presents a Levene test statistic of 43.28, where I reject that the variance pre and post period is equal. The Australian group presents a Levene test statistic of 0.34, where I fail to reject that the variance pre and post period is equal. In Figure 8 it is evident that the variance decreases for UK companies in the post period. This result is consistent with increased homogeneity of ESG Content in reports. The nature of language being used in corporate filings has become more standardised and consistent because of the UK mandate for ESG reporting.

Management can use the tonality of disclosure to create biases in a firm's narrative. Cho, Roberts, and Patten (2010) highlights that worse environmental performance use more optimistic tone in environmental their disclosure. With lacking thematic content, the perception of the narrative is crucial. Firms are known to report benign impacts, misleading investors, and creating impressions of transparency (Marquis, Toffel, and Zhou, 2016). Transparency concerns are central to the ESG disclosure argument. My results show that, on average, UK companies are more positive in their ESG disclosure compared to general disclosure after the mandate was introduced. An argument could be made that mandates enhance a firm's ability to impact investor perception through disclosure. But sentiment on ESG sentences is generally more positive, and sentiment scores are affected by the use of positive language [7]. My results suggest that disclosure mandates force companies to report on that issues that are not boilerplate or filler. The nature of language being used in reports has become more impactful and reflective of genuine risks or metrics.

Overall, my results support my hypothesis that the ESG reporting mandate changes content tonality of ESG sentences. Where the mandate has made the language used by companies more impactful on investor perception. This suggests that firms are less boilerplate in their disclosure.

5.4 Robustness Tests

To examine whether these results are robust to the specific period post 2014 for the UK mandatory reporting, I run a Monte Carlo simulation. By running an identical regression based on random post periods, I can assess whether the effect is specific to the UK mandatory ESG reporting period. I estimate 1000 regressions of random post periods to get a

[7] The inclusion of words like improvements or enhancements over decreases or losses see Appendix G for examples on sentiment classification

distribution of coefficients that are time indifferent. I can then contrast the random distribution of post treatment to the true post treatment for the original regressions in Tables 7, 8, and 11 to examine if the result is specific to the time for UK mandatory reporting. Figure 9 contains a box plot to represent the distribution of the Monte Carlo simulation for Regressions (1) through (7), (14), (15), and (16).

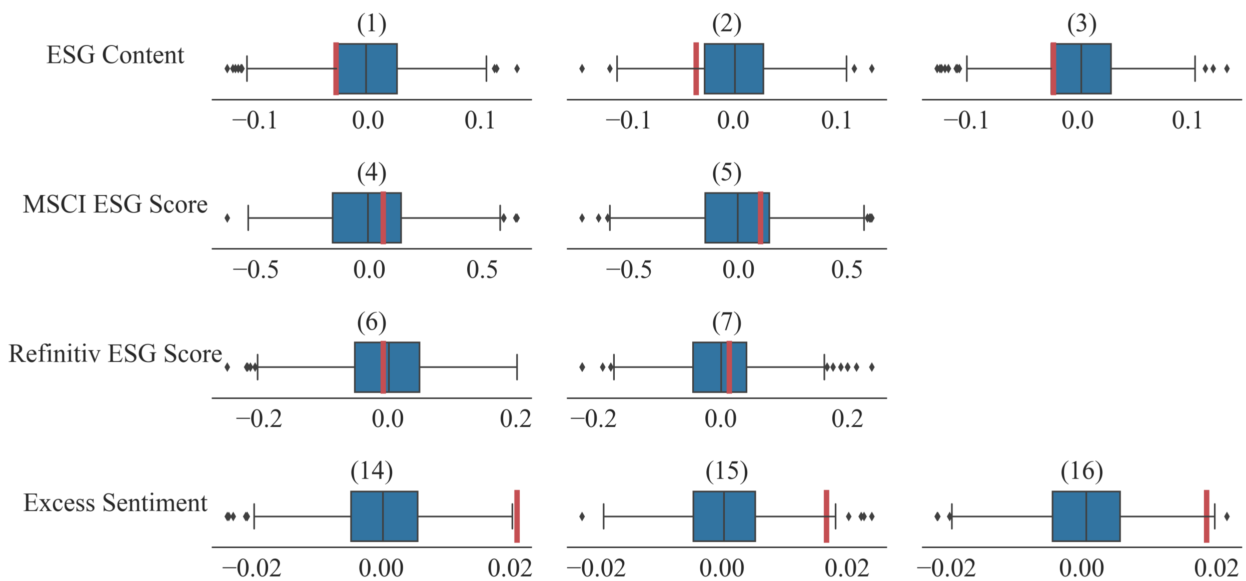


Figure 9. Simulating random post periods.
This figure plots the $Post \cdot Treatment$ box plot from Monte Carlo simulation of 1000 random post periods in each Regressions (1) through (7), (14), (15), and (16). The red line represents the real post period as presented in the true regression.

Subfigures (1), (2), and (3) represent the distribution of *ESG Content* $Post \cdot Treatment$ coefficients from a simulation of Regressions (1), (2), and (3). Visually, these coefficients fall on the lower bound of the interquartile range at Q1. These results show that the regression on ESG Content cannot be attributed to the treatment period as it falls within the bounds of the randomly assigned post period. The effect was insignificant originally, where the robustness test highlights that the effect is indifferent to the random treatment period, further supporting the insignificant result.

Subfigures (4) and (5) represent the distribution of *MSCI ESG Score* $Post \cdot Treatment$ coefficients from a simulation of Regressions (4) and (5). While Subfigures (6) and (7) represent the *Refinitiv ESG Score* $Post \cdot Treatment$ coefficients from a simulation of Regressions (6) and (7). All regressions fall within the bounds in the randomly assigned post period. The true $Post \cdot Treatment$ coefficients from Regressions (4) through (7) were all insignificant. The robustness test highlights that the results are indifferent to a random treatment period, further supporting the insignificant result.

The Subfigures (14), (15), and (16) highlight the *Post* $\cdot$ *Treatment* coefficients for a simulation of Regression (14), (15), and (16). These plots highlight that the *Excess Sentiment Post* $\cdot$ *Treatment* coefficients sit near the outer 95 percentile band of the distribution from random post periods. The true result from Regressions (14), (15), and (16) appear to be somewhat of an anomaly. This suggests that the results of the regressions on *Excess Sentiment* are specific to the UK mandatory ESG reporting period.

Overall, the robustness test supports my original results. The measures surrounding *ESG Content* and ESG performance are clearly insignificant and time indifferent. In contrast, the results from *Excess Sentiment* reflect that the UK mandate is a unique period in time that has increased the positivity of ESG sentences relative to the sentiment of general content. These results support the implications discussed in Sections 5.1, 5.2, and 5.3.

5.5 Implications

For regulators looking to implement mandatory ESG disclosures, there needs to be considerations for future proofing disclosure. My study highlights that mandates make corporate disclosure more homogeneous and standardised. The increased level of standardisation creates comparability. ESG reporting mandates have a significant focus on the systematic risks that face a company. In which an umbrella approach to ESG disclosure may force disclosure on immaterial issues. Whilst this is a shortcoming of current mandates, further defining and differentiating the systematic and idiosyncratic ESG risks that a company faces, would give investors a greater understanding of the risks that face a company.

ESG issues will change overtime, driven by investor needs. Adding further requirements to mandates could restrict future applications of ESG reporting mandates. As ESG issues change, more targeted reporting mandates will fail to be adequate. A future with clean energy may make emissions reporting redundant. Reporting requirements must change overtime. ESG issues are concerns of investors. Ultimately, the most accurate way to future proof disclosure would be through an investor driven market-oriented approach. As investor concerns change, so will investor needs for disclosure. But this comes with its own inherent problems. Mandates aid standardisation but also add certainty for companies. Different market participants have different demands for information. Where although time dynamic, the market-oriented approach creates significant uncertainty for firms. Mandates and regulatory intermediates add certainty for companies based on investor requirements.

Moreover, the investor driven approach comes with uncertainty surrounding the accuracy of information. Companies that over report immaterial disclosures—misleading investors with benign statements—create genuine concerns. But these issues are not central to disclosure mandates and could be considered a lack of adequate standardisation or assurance. Evidence suggests that assurance of ESG is effective at improving reporting definitions, scopes, and methodologies that require restatements for comparability (Ballou et al., 2018). Further, assurance enhances credibility regardless of whether the assurance provider comes from the auditing profession (Simnett, Vanstraelen, and Chua, 2009). A requirement for assurance on reporting is likely to contribute to effective and future proof ESG disclosures. ESG mandates alone are not sufficient and require additive measures like assurance to facilitate the effective transfer of information to investors.

5.6 Limitations

Due to time constraints, the UK ESG reporting mandate was examined on a generalised level. The ESG classifier I use in this study only scratches the surface of what is possible using the BERT architecture. With some refinement, it would be possible to classify sentences into environment, social, and governance themes or even further into specific ESG factors like emissions, human rights, and remuneration. Taking a more granular approach to assessing the effect of ESG reporting mandates, it would be possible to identify specific drivers in reporting, like whether reporting on social aspects has improved because of the ESG reporting mandate. This is an opportunity for future research.

To train the ESG classifier I use in this study, sentences are labelled based on my definition of an ESG sentence. I performed this task individually. Where my definition of ESG may differ from another person's view of ESG. Considering this and the overall performance of the ESG classifier, it may be a noisy indicator for ESG sentences. Although noisy, the ESG classifier is consistent. Human labelling of data is inconsistent and infeasible on a project of this size. Using a language model ensures this study is consistent and replicable.

ESG disclosure is used to enhance investor decision making surrounding the thematic ESG risks that a company may face. These risks can be broken down into idiosyncratic risk and systematic risk. Further complexity is added when firms' ESG disclosure does not focus on material risks. The ESG classifier I use in this thesis cannot identify when a sentence is at a risk that may be systematic, idiosyncratic, or immaterial. Future research could explore the effect mandates have on the materiality of ESG sentences within corporate filings.

An interesting avenue for future research would be to assess the opportunistic behaviour of firms' ESG disclosure. Companies could be opportunistic in their ESG reports releasing positive fillings prior to ESG rating assessments. It is not uncommon to see firms behaving opportunistically in reports. There are existing studies that explore the opportunistic behaviour of ESG reporting, but many rely on fuzzy indicators that only give an overview of firm behaviour (e.g., see Gonçalves, Gaio, and Costa, 2020). Leveraging a BERT model to identify any opportunistic ESG reporting is an avenue for future research.

6 Conclusion

This thesis extends the literature on corporate ESG disclosure and corporate social responsibility. Disclosure facilitates the information transfer between companies and investors. My results show the UK mandate for ESG disclosure is not effective at increasing content in corporate ESG disclosure or effective at improving ESG performance. The mandate does, however, make disclosure more standardised and comparable. Moreover, the language used in disclosure is more impactful on investor perception because of the ESG reporting mandate.

For regulators assessing ESG disclosure, my results highlight that ESG information requirements come from investors, not mandates. Different market participants have different requirements for information. Regulation creates certainty for companies and investors by acting as an intermediary. ESG reporting mandates help standardise ESG reporting and reduce boilerplate disclosure for investors. Additionally, mandates offer guidance for companies, creating certainty in disclosure requirements. Market led approaches will often lead to failure. Disclosure should inform investors and enrich the unique, independent decision-making process. Mandates are crucial in creating comparability for investors and certainty for companies.

Appendix

A Textual Data Cleaning Process

Reports from Refinitiv corporate filings data base come in single-file PDF format. The Python package `pdfminer` was used to extract the textual data from the PDF files. Only a few filings from this database were corrupted or unreadable due to compiling errors. The textual data were then tokenised into sentences for use with the BERT model. The flow of raw textual data contains several elements that are not necessarily relevant for sentence classification or sentiment extraction, for example, table extracts, headings, phone numbers, etc. For the best performance of the BERT model, context is important. That means short sentences inhibit the model. BERT is also only able to process a maximum token length of 512 without manipulation. This means there is a happy medium or optimal sentence length for use within a BERT model.

I cleaned the sentence level textual data by removing:

- Any sentence with less than 5 words
- Any sentence with less than 50 characters
- Any sentence containing over 30% numbers
- Any sentence that contained over 10 period symbols ('.')
- Any sentence with less than 65% alpha characters
- Any sentence with over 1000 characters and less than 85% alpha characters
- Any sentence with over 1500 characters
- Any sentence with over 500 words
- Any sentence that contained over 5 '/'

B Bidirectional Encoder Representations from Transformers (BERT)

The aim of this appendix is to provide a level of detail and intuition on the inner workings of BERT. For a complete explanation, refer to the original paper, or for a more digestible explanation, see Alammar's illustrative examples on BERT and transformers (Devlin et al., 2018; Alammar, 2018b,a).

BERT is an natural language processing (NLP) model that is said to have a base understanding of language. Developed by Google, it leverages transfer learning to remove the significant computational costs involved in training a large language model. BERT's key difference comes from the application of bidirectional training on transformer model architecture (Radford et al., 2019). An outline of the encoder process follows:

Each word in a sentence is converted into a vector based on an embedding algorithm. Each word is now represented by a vector of numbers (known as a word vector). The key issue with this approach is words have dual meanings and a single vector for the same word may not be appropriate. The solution to this is to add some level of context into the model. This is achieved through the attention mechanism.

Figure B.10 outlines the attention process, the first step of the process involves creating a Query, Key, and Value vector based on the word embeddings. These vectors are created by multiplying the word embedding by three matrices that are trained during the initial training process. The next step involves taking the pair-wise dot product of each query and key vector to get a matrix of scores. This can be interpreted as how similar each word vector is to each word in a sentence. The next step is to normalize the score by dividing the score by the square root of the dimension of the key vector. This is to stabilise the gradients. Softmax is then applied to each column in the matrix to normalise the score to sum up to 1 whilst also punishing words that have low similarity with each other. Appendix E has a detailed explanation of the Softmax algorithm. Following this, each value vector is multiplied by the Softmax score for that vector. This adds focus on words that have greater value whilst ignoring words with low similarity. The last step involves taking the sum of the weighted value vectors, giving the output vector for that word. A contextualised word embedding that is made up of some part of every other word in a sentence.

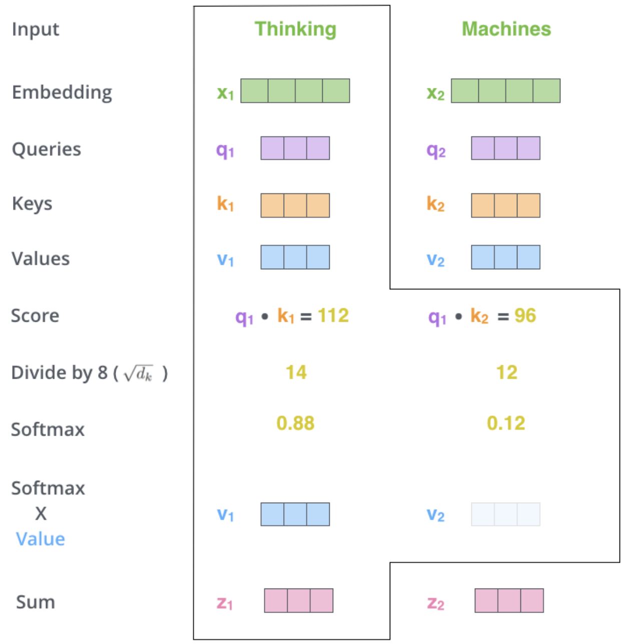


Figure B.10. Attention Process.
This figure outlines the attention process. Starting with query, key, and value vectors, the pairwise dot product is taken between a words query vector and every other words key vector. The new value is normalised and Softmax is applied. The contextualised word embedding is captured by taking the sum of the Softmax value times the value vector. This illustration is sourced from Alammar (2018b).

BERT architecture leverages multi-head attention, multi-head attention gives the model the ability to focus on unique positions of the word vector, whilst also giving the attention layer multiple representation subspaces. Having multiple heads, each with unique trained query key and value vectors, allows for different representations in each subspace of the word embeddings. Figure B.11 outlines the adaptation of the attention mechanism to a multi-head mechanism. BERT has 12 attention heads in each of its 12 layers, where each input word embedding is split up into 12 heads, with an input length of 768 for each word embedding. This means each head has an input length of 64. To reduce the 12 heads back into a vector that I can use in feed forward neural networks, each head is concatenated and multiplied by a matrix that was trained along with the model.

The complexity of the BERT model shows when multi-head attention is introduced. BERT is simply 12 layers of encoders stacked onto of each other, each taking an input from the last. The final output from this procedure is an embedded word vector that captures context and that I can use as an input in a feed forward neural network.

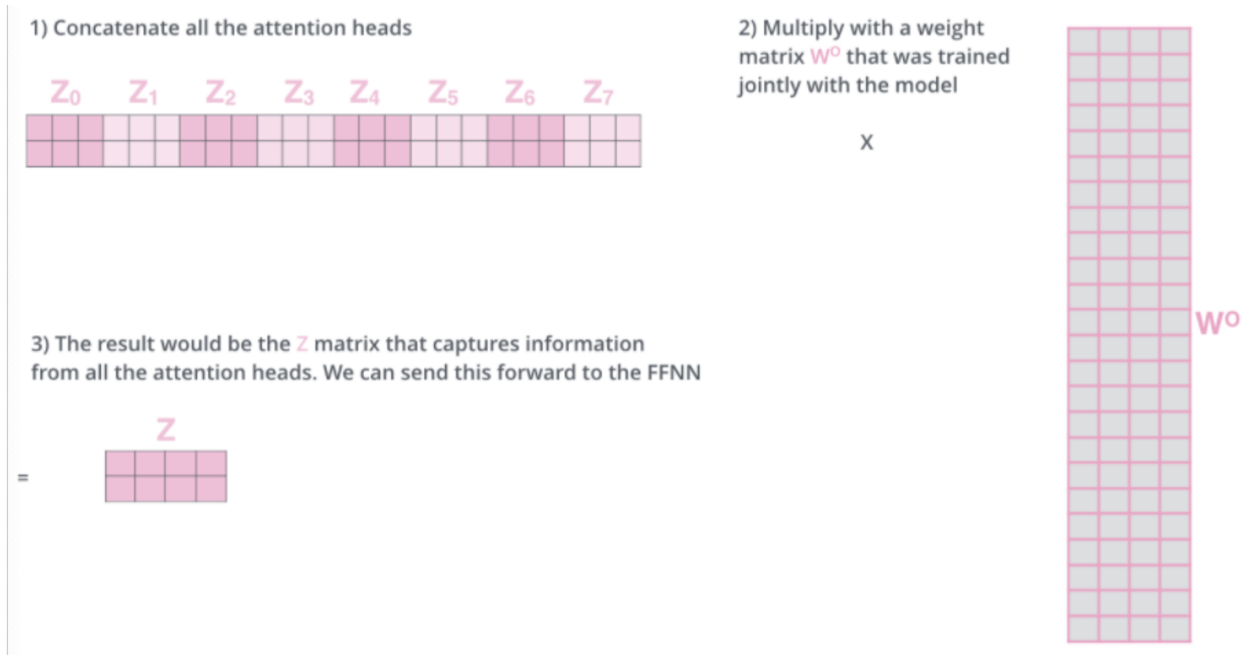


Figure B.11. Multi-head Attention.
This figure outlines the multi-head attention concatenation process. With BERT each word embedding is split across 12 heads of vectors with a length of 64. The total word embedding length is 768, where each head is concatenated and multiplied by a pretrained matrix. This captures the information from the 12 word vectors of each head and combines them into a single word embedding. This illustration is sourced from Alammar (2018b).

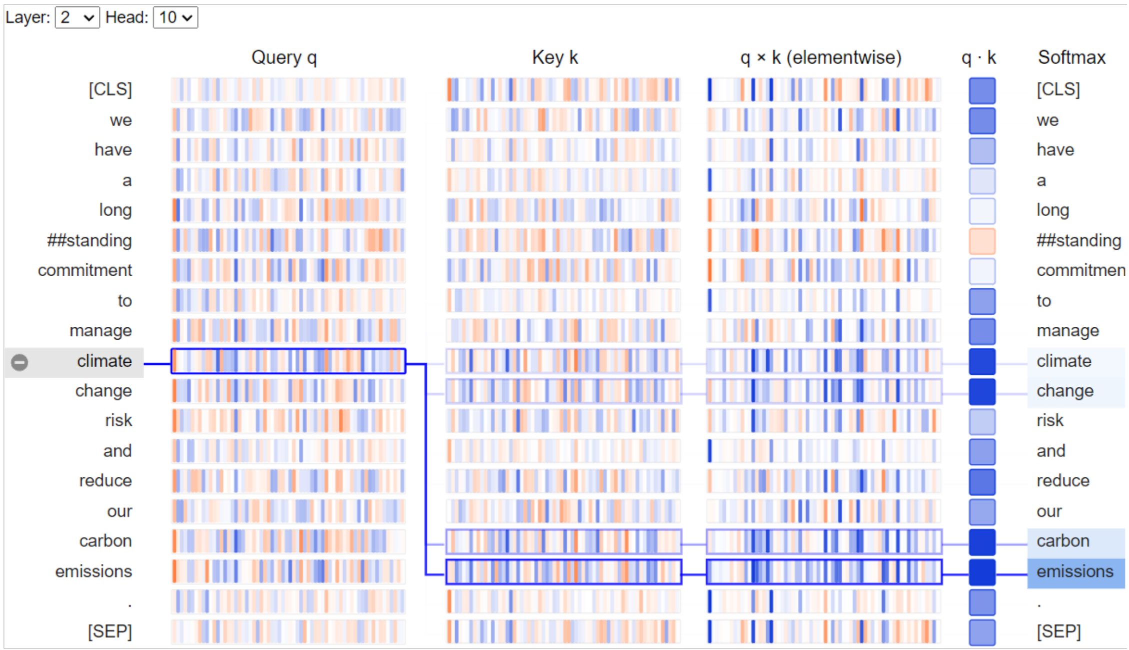
B.1 Word Embeddings

BERT requires inputs of a specific format. This format includes special tokens and positional identifiers. The [CLS] token is required to mark the beginning of a sentence and is pre-pended to the sentence embedding. Similarly, the [SEP] token is used to both mark the end of a sentence or the separation between two sentences. The application of the [SEP] token is task specific. In question/answering the [SEP] separates the question and answer and appears at the end, while in sentence classification, the [SEP] token should only appear at the end of the sentence. The BERT word embeddings are based on word piece tokens. Word piece tokens help break up compound words such as "outstanding" which gets tokenised into "out" and "##standing". '##' represents the token making up a compound word or sub word. Each word is then mapped to an embedding based on a preassigned vocabulary. BERTs vocabulary is robust. Any words that are out of vocabulary, such as misspelled words or gibberish, can comprise word piece tokens that contain single letters. The uninterpretable 'ofsif' is tokenised as: "[CLS], of, ##s, ##if, [SEP]". The design of the vocabulary allows BERT to deal with words that the model is unfamiliar with.

BERTs attention layers can be visualised to highlight where the strongest impact weights are or, in a more intuitive sense, where the model attends to each word (Vig, 2019). Visualization of the different attention layers helps to interpret model behaviour. As previously mentioned, BERT is a multi-head model with 12 layers of 12 attention heads that get concatenated to output the final word embeddings. We can take a granular approach to visualise each head in each layer. Figure B.12 representing the attention mechanism for the sentence:

> "We have a longstanding commitment to manage climate change risk and reduce our carbon emissions."

Panel A. Climate Attention Overview.



Panel B. Change Attention Overview.

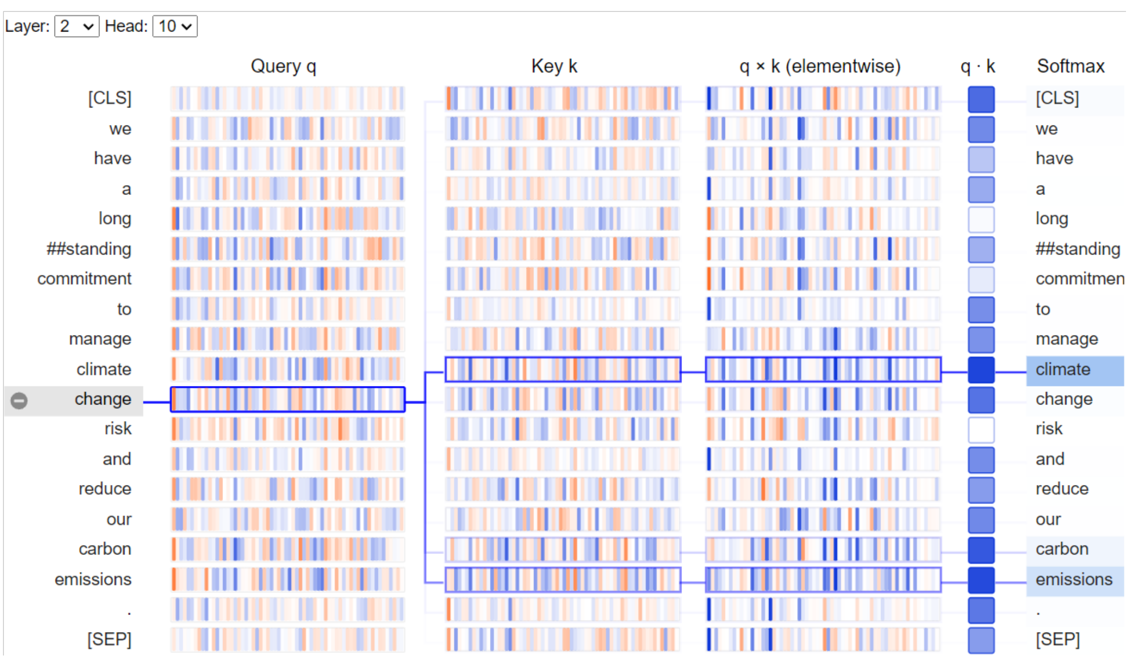


Figure B.12. Attention Visualisation.

These figures present a visual representation of the attention links in layer 2 head 10 for the sentence "We have a longstanding commitment to manage **climate change** risk and reduce our carbon emissions." Panel A shows the attention for the word "climate." The dark blue lines represent a Softmax value that approaches 1. We can see that the word for "climate" attends to the words "carbon" and "emissions". Panel B shows the attention for the word "change." We can see that the word for "change" attends to the words "climate" and "emissions". This visualisation highlights how BERT is able to capture context.

I have tokenised this sentence into BERT format with [CLS] and [SEP] tokens. In this example, we can see that for layer 2 head 10, the word "change" attends to climate and emissions, as a result, the value for the word vector for "change" is most changed by the word "climate", or in the intuitive sense, the word change attends to climate. This has grabbed the context required to understand the bigram climate change through the attention mechanism.

This mechanism happens for each layer in each head. I can also visualise the full model to observe the attention patterns, in figure B.13, the thicker the line, the stronger the attention between 2 words. Intuitively, some patterns are obvious, such as the attention to the [SEP] or [CLS] token and the attention to the previous or next word in the sequence. Attention to the special tokens adds little value to the word vectors while attention to the next word in a sequence is quite logical, as without the next word, a sentence cannot be constructed.

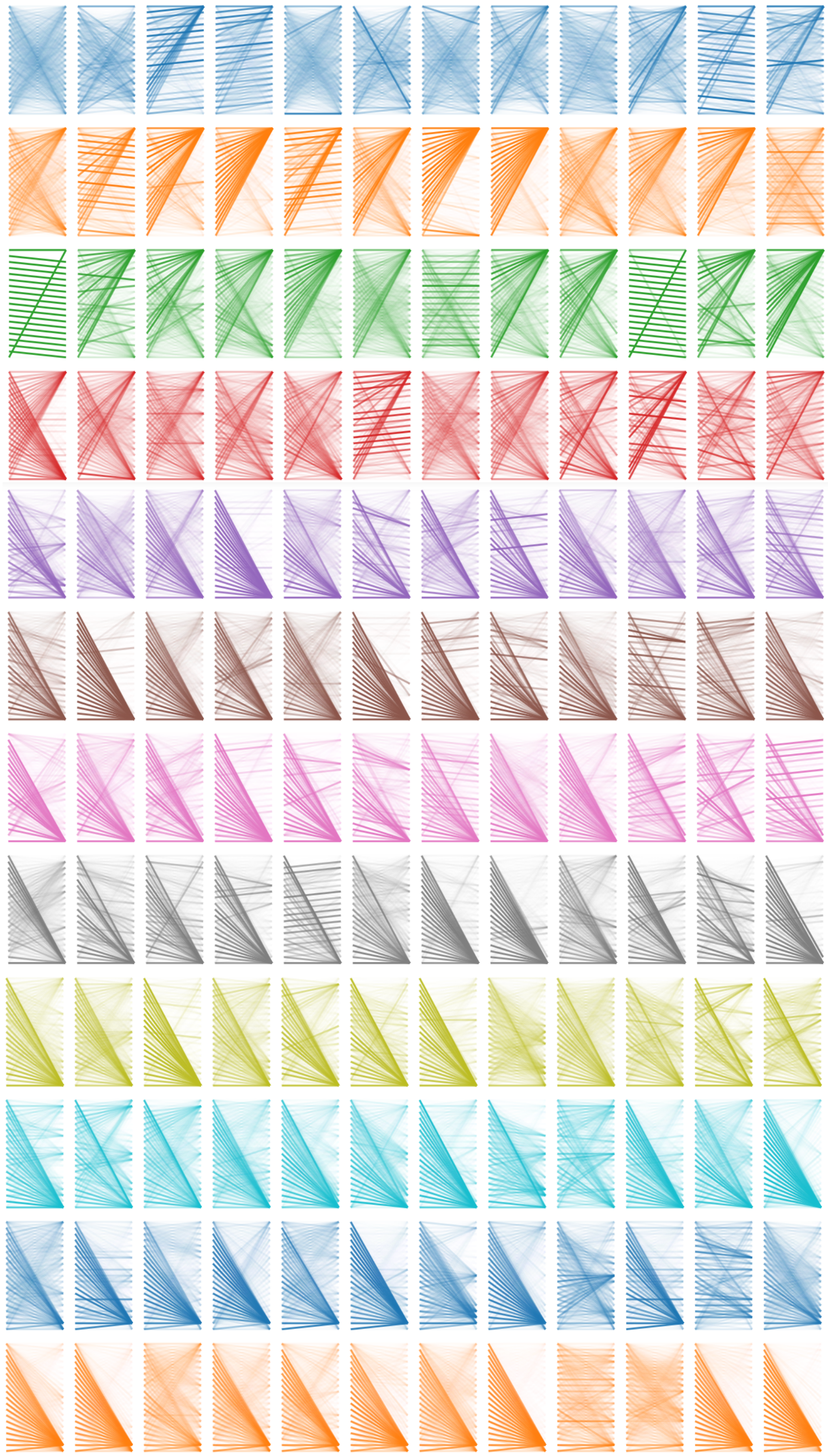


Figure B.13. Full BERT Visualisation.
This figure outlines attention process for the full BERT model (12 layers, 12 heads) on the sentence: “We have a longstanding commitment to manage **climate change** risk and reduce our carbon emissions.”

B.2 BERT Training

BERTs training is broken up into two parts, pretraining and fine tuning. BERTs pretraining helps remove the significant computational costs involved in training a language model of this size. Pretraining also decreases the cost associated with fine tuning.

B.2.1 Pretraining

BERT is pretrained to have a base understanding of language. They trained BERT on BookCorpus, a dataset containing 11,038 unpublished books from 16 different genres, along with 2,500 million words from English Wikipedia text passages. The developers of BERT conducted semi-supervised training based on 2 methods, a masked language model (MLM) and next sentence prediction (NSP).

MLM is used to give the model an understanding of the linguistic patterns within the text. The process involves randomly masking some tokens from the input embedding. Ultimately, the goal of the model is to predict these tokens. The only clue the model has to predict each token on is the context or neighbouring tokens.

BERTs application of the MLM requires it to predict 15% of the tokens in the input picked at random. 80% of the tokens in the embedding with a [MASK] token, 10% with random word tokens and 10% with the original word token. Having a combination of both random words and original words stops the model from simply reverse engineering the pattern recognition. Because the model can't identify which word tokens have been randomly assigned, it must learn the distributional contextual representation of each input token.

NSP helps the model understand context at a greater level than within the sentence. They implemented it by giving the model an input sentence followed by either the next sentence in the corpus or a random sentence in the corpus. The probability that the sentence follows the input or not is 50%. They pass the final model embeddings through a simple classification layer. The model would then output logits that can they put through a Softmax algorithm to get a proxy for the probability of whether it is the next sentence in the sequence. BERT trains both MLM and NSP simultaneously to minimise the combined loss function.

B.2.2 Fine tuning

For task specific applications, we need to fine tune BERT. We can use BERT for text classification, sentence classification, semantic similarity of sentence pairs and question/answering. For each application, we require fine tuning to leverage the base understanding of language that model gained during its pretraining phase. This paper applies BERT to classify whether a sentence is related to ESG and further uses an extension of sentence classification for sentiment analysis. Fine tuning is cost effective and fast to train compared to pretraining. The significant cost involved in training the model occurred when the model was being trained on MLM and NSP. Classification tasks require labelled data to train the model to identify the different required classes.

Fine tuning BERT for ESG sentence classification, I start by hand labelling example sentences. I label a sub sample of 3,763 sentences and maintain 500 sentences as a test set. I follow a method outlined by Kölbel et al. (2021) that improves model performance with limited training data. Because of the time constraints of this project, and the investment of time required to label data, this method provides an effective way to improve performance under these limitations.

The first stage of the model was trained on 3,263 example sentences. I trained the model for 15 epochs until there was no improvement in the validation loss. This model is then applied to a sample of 50,000 sentences. This gives me the probability (Softmax weight) that each sentence belongs to each class. In a binary classification, the weight for the opposing class is simply inverse (difference between 1 and probability). With model probabilities, the confusing examples can be examined. Confusing examples are where the model performs poorly, and classification is fuzzy. I classified confusing examples between 0.45 and 0.55 classification probability. There were 1268 confusing examples I hand labelled to improve model performance. To reduce sample bias, the newly labelled data is padded with the original dataset and confident examples. Confident examples are based on a 0.975 (0.025) classification probability. Examples are capped at 1000 to reduce sample bias.

This new training dataset contains 7030 sentences where 5030 have been hand labelled and 2000 are based on prior confident examples. The model was trained on the new dataset for 15 epochs until there was no improvement in the validation loss. The performance is improved by applying this method. Further improvements in performance are gained by using a classification probability of 0.7 to maximise the F1-score on the test dataset. Appendix C contains the hyper-parameters used in the model's training. The results from the Stage 1 iteration and 0.5 classification probability are in Appendix D.

C BERT Hyperparameters

BERT Hyperparameters Description

Variable	Description	Stage 1	Stage2
Baseline model	Model used as a baseline for training the classification model.	FinBERT	Stage 1 Model
Batch size	The batch size represents the number of training examples used in one training iteration. The larger the batch size, the more stable the gradients. They recommended for BERT models to use a batch size of 16 or 32 (Devlin et al., 2018).	32	32
Learning rate	Taken from the Adam paper, which outline an optimiser for stochastic gradient descent (Kingma and Ba, 2014). The learning rate represents how much the model will change in response to the estimated error for each update to model weights. A learning rate that is too small will cause a model that does not improve when updated, where a learning rate that's too large will cause an unstable training progress. The recommended learning rates for BERT are 5e-5, 3e-5, 2e-5 (Devlin et al., 2018).	2e-5	2e-5
Adam Epsilon	Taken from the Adam paper, Adam epsilon is a tuneable parameter that can be changed to help stabilize gradients (Kingma and Ba, 2014). Intuitively, it can be interpreted as an impact of weights, where the larger the value, the smaller the impact on the weights of the model. The ADAM paper recommends 1e-8 as a default value. This is not always the best fit, for this project, I tuned the Adam epsilon to be task specific, where the focus was on getting an interpretable validation loss schedule.	0.001	0.01
Loss function	A loss function simply calculates the distance between the output of the model and the expected output. In this project, I implemented a sample bias weight adjusted cross entropy loss (sbCEL). Appendix F contains the formula for the loss function.	sbCEL	sbCEL

(To be continued)

Variable	Description	Stage 1	Stage2
Dropout Probability	The dropout probability is used as a method of ensuring the model does not overfit. The process involves randomly dropping nodes from the neural network to prevent the model of co-adaption (Srivastava et al., 2014). In applications of BERT, I left the dropout probability at 0.1, as recommended in the original paper (Devlin et al., 2018). This means that 10% of the nodes are dropped from the model during a training schedule.	0.1	0.1
Epochs	Refers to the number of full training passess on the training dataset	15	15
Training dataset	The labelled dataset used for training the model. The data the model sees and learns from. Chosen at random.	2611	5877
Validation dataset	A sample of the training data used to as an unbiased evaluation of model fit on the training dataset. The data the model sees but doesn't learn from. Chosen at random, based on a ratio of the full training dataset.	652, 80/20	653, 90/10.
Test dataset	A sample of labelled data to provide a completely unbiased evaluation of the final model. The same test set is used to maintain comparable results	500	500

Table C.12. BERT Hyperparameters Description.
This table outlines the different the different hyperparameters used when training both Stage 1 and Stage 2 ESG classifiers. The description columns describes the variable and gives an overview of the decision making process for selecting certain values.

D BERT Results

Stage 1 training schedule and results.

Figure D.14. Stage 1 training schedule.
The Stage 1 model is trained for 15 epochs until there was no improvement in validation loss.

Table D.13. Stage 1 ESG Classifier Performance.

This table presents the results of the Stage 1 ESG classifier with a 0.5 classification probability. Precision is the ratio of correctly predicted positive examples to the total predicted positive examples. Recall is the ratio of correctly predicted positive examples to all examples in the class. The F1-score is the weighted average of the precision and recall scores. Support is the total number of sentences in each sample class.

		Precision	Recall	F1-Score	Support
General Content	0	0.96	0.92	0.94	407
ESG	1	0.70	0.83	0.76	93
Accuracy				0.90	500
Macro avg		0.83	0.87	0.85	500
Weighted avg		0.91	0.90	0.91	500

Stage 2 training schedule and results.

Figure D.15. Stage 2 training schedule.
The Stage 2 model is trained for 15 epochs until there was no improvement in validation loss.

Table D.14. Stage 2 ESG Classifier Performance.

This table presents the results of the Stage 2 ESG classifier with a 0.7 classification probability. Precision is the ratio of correctly predicted positive examples to the total predicted positive examples. Recall is the ratio of correctly predicted positive examples to all examples in the class. The F1-score is the weighted average of the precision and recall scores. Support is the total number of sentences in each sample class.

	Precision	Recall	F1-Score	Support
General Content 0	0.98	0.92	0.95	407
ESG 1	0.73	0.90	0.81	93
Accuracy			0.92	500
Macro avg	0.85	0.91	0.88	500
Weighted avg	0.93	0.92	0.92	500

E Softmax

The Softmax algorithm is a method of normalising a vector of real values such that they sum to 1. They often use it in machine learning to help scale model outputs into a range that they can be interpreted as a probability. It draws many similarities with the sigmoid function and logistic regressions. The formula is:

$$\text{Softmax}(\vec{Z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

Where: $\vec{Z}$ is an input vector, K is the number of classes and e is a standard exponential function. Example follows:

Model Output	e^z	Softmax
-1	0.368	0.047
2	7.39	0.953

F Loss Function

In this project, I implemented a sample bias weight adjusted cross entropy loss. It is calculated through the following equation:

$$weight[class] = \frac{nSamples}{nClasses * nSamples[class]}$$

$$\text{loss}(x, class) = weight[class](-x[class] + \log(\sum_j \exp(x[j])))$$

$$loss = \frac{\sum_{i=1}^{N} loss(i, class[i])}{\sum_{i=1}^{N} weight[class[i]]}$$

G BERT Examples

Table G.15. BERT and FinBERT Examples.

This tables outlines examples of sentiment scores on a small sample of sentences. The sentiment score ranges from -1 to 1, negative to positive, with 0 being neutral.

Score	ESG Examples
0.93	Environment - The launch of our latest innovation, Caroma Smart Command, an intelligent bathroom system to monitor and manage water in the built environment, further enhances Caromas reputation and commitment to reducing water usage in the built environment.
0.91	Second, in line with the revised Code, we have taken further steps during the year to facilitate improved ongoing oversight by the Board of the Groups risk management and internal control processes.
0.01	The Board has overall accountability for reviewing and approving executive remuneration as well as Non-executive Director Board and Committee fees (subject to the Board fee pool approved by shareholders).
−0.93	Failure to manage these environmental risks properly could result in litigation, regulatory action and additional remedial costs that may materially and adversely affect our financial results.
−0.91	Threats Fines may be imposed against Group companies for breaching antitrust rules,anti-corruption legislation, sanctions or human rights violations or for other inappropriate business conduct.
	General Content Examples
0.94	Our focus on productivity has improved operating performance at each of our Businesses.
0.94	Bunzl has produced another excellent set of results with growth across all business areas and strong increases in revenue, profits, earnings and dividend.
0.00	A summary of our current policies and practices regarding liquidity and funding is provided in the Appendix to Risk on page 188.
−0.93	and for a portfolio with an exposure of more than $550 billion, the losses in 2018 were $86 million1.
−0.95	Non current assets decreased by A$56.4 million (including increase in loans and receivables A$2.6 million, inventories A$9.6 million, investment properties A$28.5 million, property plant and equipment A$1.6 million and decrease in equity accounted investments by A$98.7 million).

H FinBERT Misclassification

Table H.16. FinBERT Misclassification Examples.

This tables outlines examples of sentiment scores on a small sample of sentences where there are inaccuracies. The sentiment score ranges from -1 to 1, negative to positive, with 0 being neutral. The true sentiment score should be positive.

−0.88	Brambles recorded a decline in Scope 1 and Scope 2 greenhouse gas (GHG) emissions and energy use for the Year.4
−0.80	The FY2013 employee Lost Time Injury Frequency Rate (LTIFR) was 2.1 which pleasingly was a slight decrease on last years result of 2.2.
−0.97	Tullows LTIF rate dropped to 0.28 due to reporting three lost time injuries compared to four in 2017.

I ESG Rating Methodology

Table I.17. ESG Rating Methodology.

This table contrasts the MSCI and Refinitiv ESG score assessment methodology.

	MSCI	Refinitiv
Goal	Measure a companies ESG Risk Exposure	Measure a companies ESG Performance
Focus	37 key industry material issues	A material subset of 500 company level metrics
'E'	Climate Change, Natural Capital, Pollution/Waste, Environmental Opportunities	Emissions, Innovation, Resource use
'S'	Human Capital, Product Liability, Stakeholder Opposition, Social Opportunities	Community, Human Rights, Product Responsibility, Workforce
'G'	Corporate Governance, Corporate Behaviour	CSR Strategy, Management, Shareholders
Score	Standardised comparison to peers, Risk exposure conditional on risk management	Standardised comparison to peers, Weighted on subset of material metrics

References

Alammar, J. (2018a). *The Illustrated BERT, ELMo, and co. (How NLP Cracked Transfer Learning)*. URL: https://jalammar.github.io/illustrated-bert/.

— (2018b). *The Illustrated Transformer*. URL: https://jalammar.github.io/illustrated-transformer/.

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

Araci, D. (2019). "Finbert: Financial sentiment analysis with pre-trained language models". arXiv preprint arXiv:1908.10063.

ASX (2019). *ASX Listing Rules Chapter 4 – Periodic Disclosure*. URL: https://www.asx.com.au/documents/rules/Chapter04.pdf.

ASX Corporate Governance Council (2019). *Corporate Governance Principles and Recommendations*. URL: https://www.asx.com.au/documents/regulation/cgc-principles-and-%20recommendations-fourth-edn.pdf.

Aupperle, K. E., Carroll, A. B., and Hatfield, J. D. (1985). "An empirical examination of the relationship between corporate social responsibility and profitability". Academy of Management Journal 28, 446–463.

Australian Council of Superannuation Investors (2015). *ESG Reporting Guide for Australian Companies*. URL: https://www.asx.com.au/documents/asx-compliance/acsi-fsc-esg-%20reporting-guide-final-2015.pdf.

Ballou, B., Chen, P.-C., Grenier, J. H., and Heitger, D. L. (2018). "Corporate social responsibility assurance and reporting quality: Evidence from restatements". Journal of Accounting and Public Policy 37, 167–188.

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

Berg, F., Fabisik, K., and Sautner, Z. (2021). "Rewriting history II: The (un) predictable past of ESG ratings". European Corporate Governance Institute Working Paper.

Berg, F., Koelbel, J., and Rigobon, R. (2020). "Aggregate confusion: The divergence of ESG ratings". Available at SSRN 3438533.

Bessec, M. and Fouquau, J. (2020). "Green Sentiment in Financial Markets: A Global Warning". Available at SSRN 3710489.

Bingler, J. A., Kraus, M., and Leippold, M. (2021). "Cheap Talk and Cherry-Picking: What ClimateBert has to say on Corporate Climate Risk Disclosures". Available at SSRN 3796152.

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

Bloomberg Intelligence (Feb. 2021). *ESG assets may hit $53 trillion by 2025, a third of global AUM*. URL: https://www.bloomberg.com/professional/blog/esg-assets-may-hit-53-trillion-by-2025-a-third-of-global-aum/.

Bodnaruk, A., Loughran, T., and McDonald, B. (2015). “Using 10-K text to gauge financial constraints”. Journal of Financial and Quantitative Analysis 50, 623–646.

Boffo, R. and Patalano, R. (2020). “ESG investing: Practices, progress and challenges”. OECD 1, 2021. URL: https://www.oecd.org/finance/ESG-Investing-Practices-Progress-Challenges.pdf.

Cao, S., Jiang, W., Yang, B., and Zhang, A. L. (2020). “How to talk when a machine is listening: Corporate disclosure in the age of AI”. Available at SSRN 3683802.

Chatterji, A., Levine, D., and Toffel, M. (2009). “How well do social ratings actually measure corporate social responsibility?” Journal of Economics & Management Strategy 18, 125–169.

Chatterji, A. K., Durand, R., Levine, D. I., and Touboul, S. (2016). “Do ratings of firms converge? Implications for managers, investors and strategy researchers”. Strategic Management Journal 37, 1597–1614.

Cho, C. and Patten, D. (2007). “The role of environmental disclosures as tools of legitimacy: A research note”. Accounting, Organizations and Society 32, 639–647.

Cho, C. H., Roberts, R. W., and Patten, D. M. (2010). “The language of US corporate environmental disclosure”. Accounting, Organizations and Society 35, 431–443.

Christensen, H. B., Hail, L., and Leuz, C. (2021). “Mandatory CSR and sustainability reporting: economic analysis and literature review”. Review of Accounting Studies, 1–73.

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

Cochran, P. L. and Wood, R. A. (1984). “Corporate social responsibility and financial performance”. Academy of Management Journal 27, 42–56.

Companies Act 2006 (UK).

Courtis, J. K. (2004). “Corporate report obfuscation: artefact or phenomenon?” British Accounting Review 36, 291–312.

Craja, P., Kim, A., and Lessmann, S. (2020). “Deep learning for detecting financial statement fraud”. Decision Support Systems 139, 113421.

Daines, R. M., Gow, I. D., and Larcker, D. F. (2010). “Rating the ratings: How good are commercial governance ratings?” Journal of Financial Economics 98, 439–461.

Davies, D. and Brennan, N. (2007). "Discretionary disclosure strategies in corporate narratives: Incremental information or impression management". Journal of Accounting Literature 26, 116–196.

Delmas, M. and Blass, V. D. (2010). "Measuring corporate environmental performance: the trade-offs of sustainability ratings". Business Strategy and the Environment 19, 245–260.

Delmas, M. A. and Burbano, V. C. (2011). "The drivers of greenwashing". California Management Review 54, 64–87.

Delmas, M. A., Etzion, D., and Nairn-Birch, N. (2013). "Triangulating environmental performance: What do corporate social responsibility ratings really capture?" Academy of Management Perspectives 27, 255–267.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding". arXiv preprint arXiv:1810.04805.

Dyck, A., Lins, K. V., Roth, L., and Wagner, H. F. (2019). "Do institutional investors drive corporate social responsibility? International evidence". Journal of Financial Economics 131, 693–714.

Dyer, T., Lang, M., and Stice-Lawrence, L. (2017). "The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation". Journal of Accounting and Economics 64, 221–245.

Eccles, R. and Stroehle, J. (2018). "Exploring social origins in the construction of ESG measures". Available at SSRN 3212685.

Eccles, R. G., Ioannou, I., and Serafeim, G. (2014). "The impact of corporate sustainability on organizational processes and performance". Management Science 60, 2835–2857.

Eccles, R. G. and Serafeim, G. (2013). "The Performance Frontier: Innovating for a Sustainable Strategy". Harvard Business Review 91, 17–18.

Eccles, R. G., Serafeim, G., and Krzus, M. P. (2011). "Market interest in nonfinancial information". Journal of Applied Corporate Finance 23, 113–127.

Engle, R. F., Giglio, S., Kelly, B., Lee, H., and Stroebel, J. (2020). "Hedging climate change news". Review of Financial Studies 33, 1184–1216.

ERM (2021). *Study on Sustainability Related Ratings, Data and Research.* URL: https://op.europa.eu/s/pnbs.

Fabrizio, K. R. and Kim, E.-H. (2019). "Reluctant disclosure and transparency: Evidence from environmental disclosures". Organization Science 30, 1207–1231.

Feldman, R., Govindaraj, S., Livnat, J., and Segal, B. (2008). "The incremental information content of tone change in management discussion and analysis".

Fiedler, T., Pitman, A. J., Mackenzie, K., Wood, N., Jakob, C., and Perkins-Kirkpatrick, S. E. (2021). "Business risk and the emergence of climate analytics". Nature Climate Change 11, 87–94.

Firth, J. R. (1957). *Papers in linguistics, 1934-1951*. Oxford University Press.

Flammer, C., Toffel, M. W., and Viswanathan, K. (2021). "Shareholder activism and firms' voluntary disclosure of climate change risks". Strategic Management Journal 42, 1850–1879.

Friberg, R. and Seiler, T. (2017). "Risk and ambiguity in 10-Ks: An examination of cash holding and derivatives use". Journal of Corporate Finance 45, 608–631.

Friede, G., Busch, T., and Bassen, A. (2015). "ESG and financial performance: aggregated evidence from more than 2000 empirical studies". Journal of Sustainable Finance & Investment 5, 210–233.

Friedman, M. (1970). "A theoretical framework for monetary analysis". Journal of Political Economy 78, 193–238.

Gibson, R., Krueger, P., and Schmidt, P. S. (2019). "ESG rating disagreement and stock returns". Swiss Finance Institute Research Paper.

Gonçalves, T., Gaio, C., and Costa, E. (2020). "Committed vs opportunistic corporate and social responsibility reporting". Journal of Business Research 115, 417–427.

Gray, R. (2010). "Is accounting for sustainability actually accounting for sustainability... and how would we know? An exploration of narratives of organisations and the planet". Accounting, Organizations and Society 35, 47–62.

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

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

Griffin, J. J. and Mahon, J. F. (1997). "The corporate social performance and corporate financial performance debate: Twenty-five years of incomparable research". Business & Society 36, 5–31.

Gurun, U. G. and Butler, A. W. (2012). "Don't believe the hype: Local media slant, local advertising, and firm value". Journal of Finance 67, 561–598.

Hamilton, L. C. (2016). "Public awareness of the scientific consensus on climate". Sage Open 6, 2158244016676296.

Hansen, S. and McMahon, M. (2016). "Shocking language: Understanding the macroeconomic effects of central bank communication". Journal of International Economics 99, S114–S133.

Hartzmark, S. M. and Sussman, A. B. (2019). "Do investors value sustainability? A natural experiment examining ranking and fund flows". Journal of Finance 74, 2789–2837.

Hillert, A., Jacobs, H., and Müller, S. (2014). "Media makes momentum". Review of Financial Studies 27, 3467–3501.

Hummel, K. and Rötzel, P. (2019). "Mandating the Sustainability Disclosure in Annual Reports—Evidence from the United Kingdom". Schmalenbach Business Review 71, 205–247.

Ilinitch, A. Y., Soderstrom, N. S., and Thomas, T. E. (1998). "Measuring corporate environmental performance". Journal of Accounting and Public Policy 17, 383–408.

Index Industry Association (2021). *Measurable Impact: Asset Managers on the Challenges and Opportunities of ESG Investment*. URL: http://www.indexindustry.org/wp-content/uploads/2021/07/IIA-ESG-Executive-Summary-2021-vFINAL.pdf.

Ioannou, I. and Serafeim, G. (2017). "The consequences of mandatory corporate sustainability reporting". Harvard Business School research working paper.

Jegadeesh, N. and Wu, D. (2013). "Word power: A new approach for content analysis". Journal of Financial Economics 110, 712–729.

Jouvenot, V. and Krueger, P. (2019). "Reduction in corporate greenhouse gas emissions under prescriptive disclosure requirements". Available at SSRN 3434490.

Khan, M., Serafeim, G., and Yoon, A. (2016). "Corporate sustainability: First evidence on materiality". Accounting Review 91, 1697–1724.

Kim, E.-H. and Lyon, T. P. (2011). "Strategic environmental disclosure: Evidence from the DOE's voluntary greenhouse gas registry". Journal of Environmental Economics and Management 61, 311–326.

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

Kim, I., Wan, H., Wang, B., and Yang, T. (2019). "Institutional investors and corporate environmental, social, and governance policies: Evidence from toxics release data". Management Science 65, 4901–4926.

Kingma, D. P. and Ba, J. (2014). "Adam: A method for stochastic optimization". arXiv preprint arXiv:1412.6980.

Kölbel, J. F., Leippold, M., Rillaerts, J., and Wang, Q. (2021). "Ask BERT: How regulatory disclosure of transition and physical climate risks affects the CDS term structure". Available at SSRN 3616324.

Konar, S. and Cohen, M. A. (1997). "Information as regulation: The effect of community right to know laws on toxic emissions". Journal of environmental Economics and Management 32, 109–124.

Kotsantonis, S. and Serafeim, G. (2019). “Four things no one will tell you about ESG data”. Journal of Applied Corporate Finance 31, 50–58.

Krueger, P., Sautner, Z., Tang, D. Y., and Zhong, R. (2021). “The Effects of Mandatory ESG Disclosure around the World”. Available at SSRN 3832745.

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., and Soricut, R. (2019). “Albert: A lite bert for self-supervised learning of language representations”. arXiv preprint arXiv:1909.11942.

Levene, H. (1960). “Contributions to probability and statistics”. Essays in honor of Harold Hotelling, 278–292.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). “Roberta: A robustly optimized bert pretraining approach”. arXiv preprint arXiv:1907.11692.

Loughran, T. and McDonald, B. (2011). “When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks”. Journal of Finance 66, 35–65.

— (2014). “Measuring readability in financial disclosures”. Journal of Finance 69, 1643–1671.

Marquis, C., Toffel, M. W., and Zhou, Y. (2016). “Scrutiny, norms, and selective disclosure: A global study of greenwashing”. Organization Science 27, 483–504.

Matisoff, D. C. (2013). “Different rays of sunlight: Understanding information disclosure and carbon transparency”. Energy Policy 55, 579–592.

Matsumura, E. M., Prakash, R., and Vera-Muñoz, S. C. (2020). “To Disclose or Not to Disclose Climate-Change Risk in Form 10-K: Does Materiality Lie in the Eyes of the Beholder?” Available at SSRN 2986290.

Miihkinen, A. (2012). “What drives quality of firm risk disclosure?: the impact of a national disclosure standard and reporting incentives under IFRS”. International Journal of Accounting 47, 437–468.

Mion, G. and Adaui, C. (2020). “The effect of mandatory publication of nonfinancial disclosure in Europe on sustainability reporting quality: First insights about Italian and German companies”. *Non-Financial Disclosure and Integrated Reporting: Practices and Critical Issues*. Emerald Publishing Limited.

Moniz, A. (2016). “Inferring the financial materiality of corporate social responsibility news”. Available at SSRN 2761905.

Naumer, H.-J. and Yurtoglu, B. (2020). “It is not only what you say, but how you say it: ESG, corporate news, and the impact on CDS spreads”. Global Finance Journal, 100571.

Orlitzky, M., Schmidt, F. L., and Rynes, S. L. (2003). “Corporate social and financial performance: A meta-analysis”. Organization Studies 24, 403–441.

Palmiter, A. R. (2017). "Climate change disclosure: A failed SEC mandate". Available at SSRN 2639181.

Papoutsi, A. and Sodhi, M. S. (2020). "Does disclosure in sustainability reports indicate actual sustainability performance?" Journal of Cleaner Production 260, 121049.

Patten, D. M. (1992). "Intra-industry environmental disclosures in response to the Alaskan oil spill: a note on legitimacy theory". Accounting, Organizations and Society 17, 471–475.

Plumlee, M., Brown, D., Hayes, R., and Marshall, R. S. (2015). "Voluntary environmental disclosure quality and firm value: Further evidence". Journal of Accounting and Public Policy 34, 336–361.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). "Language models are unsupervised multitask learners". OpenAI blog 1, 9.

Renneboog, L., Ter Horst, J., and Zhang, C. (2008). "Socially responsible investments: Institutional aspects, performance, and investor behavior". Journal of Banking & Finance 32, 1723–1742.

Russo, M. V. and Fouts, P. A. (1997). "A resource-based perspective on corporate environmental performance and profitability". Academy of Management Journal 40, 534–559.

Semenova, N. and Hassel, L. G. (2015). "On the validity of environmental performance metrics". Journal of Business Ethics 132, 249–258.

Serafeim, G. (2020). "Public sentiment and the price of corporate sustainability". Financial Analysts Journal 76, 26–46.

Sharfman, M. (1996). "The construct validity of the Kinder, Lydenberg & Domini social performance ratings data". Journal of Business Ethics 15, 287–296.

Simnett, R., Vanstraelen, A., and Chua, W. F. (2009). "Assurance on sustainability reports: An international comparison". Accounting Review 84, 937–967.

Sinha, N. R. (2016). "Underreaction to news in the US stock market". Quarterly Journal of Finance 6, 1650005.

Solomon, D. H. (2012). "Selective publicity and stock prices". Journal of Finance 67, 599–638.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). "Dropout: a simple way to prevent neural networks from overfitting". Journal of Machine Learning Research 15, 1929–1958.

Szwajkowski, E. and Figlewicz, R. E. (1999). "Evaluating corporate performance: A comparison of the Fortune reputation survey and the Socrates social rating database". Journal of Managerial Issues, 137–154.

Tang, D. Y., Yan, J., and Yao, C. Y. (2021). “The determinants of ESG ratings: Rater ownership matters”. Available at SSRN 3889395.

Tetlock, P. (2007). “Giving content to investor sentiment: The role of media in the stock market”. Journal of Finance 62, 1139–1168.

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). “More than words: Quantifying language to measure firms’ fundamentals”. Journal of Finance 63, 1437–1467.

The Companies (Miscellaneous Reporting) Regulations 2018 (UK).

Ullmann, A. (1985). “Data in search of a theory: A critical examination of the relationships among social performance, social disclosure, and economic performance of US firms”. Academy of Management Review 10, 540–557.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). “Attention is all you need”. *Advances in neural information processing systems*, 5998–6008.

Vig, J. (2019). “A multiscale visualization of attention in the transformer model”. arXiv preprint arXiv:1906.05714.

Waddock, S. A. and Graves, S. B. (1997). “The corporate social performance–financial performance link”. Strategic Management Journal 18, 303–319.

Widyawati, L. (2021). “Measurement concerns and agreement of environmental social governance ratings”. Accounting & Finance 61, 1589–1623.

Zhang, Z., Han, X., Liu, Z., Jiang, X., Sun, M., and Liu, Q. (2019). “ERNIE: Enhanced language representation with informative entities”. arXiv preprint arXiv:1905.07129.