Dorfleitner, Gregor; Zhang, Rongxin

Article

ESG news sentiment and stock price reactions: A comprehensive investigation via BERT

Schmalenbach Journal of Business Research (SBUR)

Provided in Cooperation with:
Schmalenbach-Gesellschaft für Betriebswirtschaft e.V.

Suggested Citation: Dorfleitner, Gregor; Zhang, Rongxin (2024) : ESG news sentiment and stock price reactions: A comprehensive investigation via BERT, Schmalenbach Journal of Business Research (SBUR), ISSN 2366-6153, Springer, Heidelberg, Vol. 76, Iss. 2, pp. 197-244, https://doi.org/10.1007/s41471-024-00185-3

This Version is available at:
https://hdl.handle.net/10419/298798

Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung (2024) 76:197–244
https://doi.org/10.1007/s41471-024-00185-3

ORIGINAL ARTICLE

ESG News Sentiment and Stock Price Reactions: A Comprehensive Investigation via BERT

Gregor Dorfleitner · Rongxin Zhang

Received: 19 July 2022 / Accepted: 12 April 2024 / Published online: 29 May 2024

Abstract In this paper, we examine in a systematic manner how investors react to the sentiment of instant ESG news. Instead of acquiring proprietary ESG news or events datasets directly from specific ESG data providers, we extract fresh ESG news directly from a plethora of raw news articles. We showcase how the latest development in NLP (i.e. the BERT model) can be applied to build a comprehensive and fresh ESG news dataset, and how company ESG news sentiment can be efficiently recognized by a machine. Overall, we find that the market reacts to ESG news based on news sentiment. On the event day, positive ESG news has an average abnormal return of 0.31% while negative ESG news leads to a mean value of −0.75%. More interestingly, we find that the impact of ESG news may depend on the company's historical ESG record. The negative impact of negative ESG news has less severe consequences for companies with an overall better ESG record, while the positive impact of positive ESG news may be more pronounced for companies with a worse ESG record.

Keywords ESG · Instant ESG News · NLP · BERT · Sentiment Analysis

JEL classification G12 · Q51 · Q56

✉ Gregor Dorfleitner · Rongxin Zhang
Department of Finance, University of Regensburg, 93040 Regensburg, Germany
E-Mail: gregor.dorfleitner@ur.de

Gregor Dorfleitner
CERMi (Centre for European Research in Microfinance), Brussels, Belgium

1 Introduction

With the increasing awareness of ethical issues such as environment protection and social care, the conception of ESG has become more and more prominent and urgent, not only in our everyday lives but also on the financial markets. As Van Duuren et al. (2016) and Amel-Zadeh and Serafeim (2018) suggest, ESG is already regarded as one of the important considerations for fund managers. In July 2020, as the booming fast fashion giant Boohoo was accused of using forced labor in a factory in Leicester, its stock price dropped more than 20% in a single day. The stock market reaction shows vividly that besides financial news, ESG news can also be an important factor and price driver on the financial markets, mainly due to their impact on reputation. Good ESG news can generally be indications of pro-ethical corporate behavior and bad ones for the opposite. Thus, the questions of how frequent good and bad news are and to what extent stock prices react to these news, help to clarify whether firms really behave ethically and to which extent the market values the apparent behavior. This article is devoted to an empirical investigation of this matter based on text analytics.

In the past decade, ESG has also become one of the hottest topics in finance literature. However, the research of ESG issues is still in its initial stage. Most ESG studies (e.g. Bennani et al. 2018; Hartzmark and Sussman 2019) rely heavily on ESG data such as different ESG ratings provided by specific ESG data providers, based on their in-house developed methodologies (Fiaschi et al. 2020). As Dorfleitner et al. (2015) suggest, there is an evident lack in the convergence of ESG measurement concepts and the different ratings neither coincide in distribution nor in risk. Therefore, empirical studies focusing on proprietary ESG performance proxies may be subjected to the problem of proxy biases. Also, the low-frequency of those ESG ratings and various rating methodologies make it almost impossible to understand how the market reacts to ESG issues in real time. Most recently, several studies (see e.g. Krüger 2015; Capelle-Blancard and Petit 2019; Taleb et al. 2020; Naumer and Yurtoglu 2022) focus more on frequent ESG information such as ESG events and ESG news. Krüger (2015) finds some evidence that investors may react to ESG events and reveal their possible pricing implications. However, due to the difficulty to process unstructured raw text data, these studies have to acquire ESG events or news data from ESG data providers.[1] The reliance on proprietary datasets may raise the concern that empirical results regarding the impact of ESG news on the financial markets could be sensitive to how data providers collect (e.g., different ESG news coverage) and process ESG news (e.g., different implementation of sentiment analysis). Therefore, despite some efforts being made, whether ESG news, or more specifically instant ESG news, influences financial markets are far from being fully understood.

In this study, we show how a comprehensive ESG news dataset is built upon a vast amount of raw ESG news and how news sentiment is extracted in a transparent way before empirical investigations are conducted. Compared to related studies

[1] To the best of our knowledge, related studies acquire ESG events or news data from ESG data providers. For instance, Krüger (2015) acquires ESG events data directly from MSCI KLD.

which often adopt ready-for-use ESG events or news data from data providers, this study builds an ESG news dataset based on raw ESG news published by more than 10,000 news sources on Thomson Reuters Eikon. We introduce the recent development in Nature Language Processing (NLP), i.e. the BERT model, to construct a comprehensive and fresh ESG news dataset from raw ESG news. Moreover, we extract sentiment signals from the unstructured textual data by applying the fined-tuned BERT sentiment classifier, which is considered more accurate than classical sentiment analysis methods such as lexicon-based sentiment analysis (Kotelnikova et al. 2021; Alaparthi and Mishra 2021).

With such a comprehensive dataset including almost all listed stocks with ESG news coverage for the past two years, we conduct for the first time a complete empirical investigation on the impact of instant ESG news on major stock markets. It sheds light on market reactions to instant ESG information. We find that the market responds to ESG news parallel to the news sentiment. The market reacts positively to positive ESG news while negatively to negative ESG news. Yet these reactions appear to be asymmetric. The market reaction to negative ESG news is stronger, compared to positive ESG news. These patterns exist not only on American stock markets, but also on European stock markets. At last, we discover an interesting point regarding the relationship between ESG news shocks and historical ESG records. When investors are confronted with ESG news, they also take the overall ESG performance of the target company into consideration. Companies with a better ESG record suffer less from market value loss due to negative ESG news, while those with a worse ESG performance enjoy more market value gain when facing positive ESG news.

These findings add to the discussion of integrating ESG factors in asset pricing (see e.g. Pedersen et al. 2021). Since ESG issues are found to be perceived seriously by investors, they should be considered and included as important factors in related research. Moreover, the empirical results question the efficiency of financial markets, as systematic arbitrage by closely monitoring ESG news could be viable. Our study also suggests that companies tend to exaggerate their ESG performance (see e.g. Kim and Lyon 2015), which is analogous to the so-called "greenwashing" phenomenon in the specific context of environmental issues. Our data shows that positive ESG news prevails on the market, which might suggest the existence of performance exaggeration regarding ESG issues. Meanwhile, the fact that the overwhelming positive ESG news is still perceived positively suggests that investors might not be able to completely detect the false claim of good ESG performance. Consequently, companies could possibly game the system by releasing more ESG information to their advantage.

The contribution of this study is twofold. First, we show how to apply the BERT model to build our own unique and massive ESG news dataset and judge news sentiment effectively and consistently. Especially, the latest breakthrough in NLP can also contribute to the advancement in financial studies focusing particularly on soft factors and provide a new and better approach in the toolbox of financial researchers to gain deeper insight into their role on the financial markets. Our study contributes to the new stream of studies leveraging the recent development in NLP in ESG related topics (Aue et al. 2022; Sokolov et al. 2021; Mehra et al. 2022; Chava

et al. 2021). Second, to the best of our knowledge, we examine the impact of ESG news in a comprehensive and complete framework for the first time. In general, we extract almost every piece of relevant instant ESG news piece for almost all listed equities, and avoid dependence on proprietary datasets and possible biases and errors associated with such tailored datasets. Therefore, the employed instant ESG news dataset is unique and comprehensive compared to other ESG events or news datasets directly sourced from ESG data vendors. The way we build the ESG news dataset enables us to come to more credible conclusions. Even though some earlier studies find that only negative ESG events or news matters (Krüger 2015; Capelle-Blancard and Petit 2019; Cui and Docherty 2020), we find evidence that investors may also value positive ones, albeit to a smaller extent. This finding has the policy implication for companies that it really matters to improve their ESG profile, but not just to avoid negative ESG news. Moreover, this study gives some clues regarding how investors deal with the relationship between newest and past ESG performance, which is rarely touched upon in ESG studies (see e.g. Serafeim and Yoon 2021). Our study suggests that a good long-term ESG profile might serve as a buffer to moderate the impact of short-term ESG news.

The remainder of the paper is organized as follows. In Sect. 2, we discuss basic background information regarding different types of ESG information, especially ESG news. In Sect. 3, we introduce briefly the recent development in NLP, i.e., the BERT model. We review the literature on market reactions to ESG performance and propose hypotheses in Sect. 4. Sect. 5 describes how we build our ESG news dataset step by step. In Sect. 6, we discuss necessary empirical methodological approaches. Sect. 7 presents the empirical results and Sect. 8 concludes.

2 ESG information processing

As the interest and demand of stakeholders in ESG issues grows, companies are subject to an increasing amount of ESG reporting guidance or requirements (KPMG 2019). According to the survey conducted by KPMG (2020), the percentage of the biggest companies which report on sustainability has increased from 53% in 2008 to 80% in 2020. Nevertheless, ESG disclosure as a source of ESG information has several obvious drawbacks for stakeholders such as low frequency, lack of credibility, timeliness, and relevance (Maniora 2017). Due to the difficulty to process ESG disclosure directly, stakeholders often rely on a third-party assessment, especially ESG ratings from ESG rating agencies (Berg et al. 2022). They usually apply a qualitative and quantitative methodology to assess corporate ESG performance by constructing ESG rating metrics based on information collected from different sources such as ESG disclosure, ESG news, and questionaries (Escrig-Olmedo et al. 2019; Del Giudice and Rigamonti 2020). However, a few studies raise the concern whether ESG ratings are good proxies of corporate ESG performance (Dorfleitner et al. 2015; Drempetic et al. 2020). Many studies show that ESG rating agencies may fail to measure (Escrig-Olmedo et al. 2019; Drempetic et al. 2020), and disagree on ESG performance (Dorfleitner et al. 2015; Berg et al. 2022; Lopez et al. 2020). Also, the fact that most ESG ratings are updated on a yearly or quarterly basis poses

a challenge for tracking corporate ESG performance in time. Even though ESG rating agencies consider various sources of ESG information including high-frequency data such as ESG news (Escrig-Olmedo et al. 2019), they are often embedded in rating scores on a periodic basis and cannot reflect the recent development of ESG performance.

Besides official ESG disclosure and ESG ratings from ESG agencies, ESG events or news can be other important sources of information for investors. In recent years, ESG events data, especially ESG incidents data, becomes more and more popular. For instance, RepRisk's incidents data is widely used by large investors (Gantchev et al. 2022). Specifically, just like traditional ESG ratings, ESG incidents indicators such as the RepRisk rating measure ESG performance quantitatively based on aggregated negative ESG news information and proprietary process. More generally, the media plays a central role in diffusing information on financial markets and contribute to the efficiency of the stock market by improving the dissemination of information (Peress 2014). On financial terminals such as Thomson Reuters and Bloomberg or main stream websites, news stories related to specific companies, including company ESG news, are updated at lightning speed. If investors care about ESG issues just like traditional financial fundamentals, they could possibly be influenced by reading these ESG news articles. However, unlike ESG ratings as numeric values, ESG news articles from different news sources are unstructured text data which is difficult to quantify. While ESG rating values can be homogeneously interpreted as the overall ESG performance, ESG news cannot be easily standardized and transformed into a common index which is easy to comprehend. Although instant ESG news may be consumed by individual or institutional investors and thus integrated into their investment decision-making process, it is unclear how and to what extent they may react to these instant non-financial information. To answer this question, a comprehensive stream of instant ESG news should be available and processed in a plausible way. Nevertheless, a ready-for-use ESG news dataset is usually not for free and should always be purchased from specific ESG data providers. Earlier related studies adopt such ESG news datasets from several popular ESG data providers such as Ravenpack and Covalence (Capelle-Blancard and Petit 2019; Cui and Docherty 2020). The key problem of this approach is that these proprietary ESG data providers may have different news coverage and textual processing methodologies, which are in most cases not transparent to researchers.

3 Advancement in Nature Language Processing: the BERT model

As the need to understand the role of soft factors extracted from unstructured text data on financial markets grows, classical textual analysis has been more commonly adopted in financial studies in recent years (see e.g. Dorfleitner et al. 2016). Despite some preliminary progress, it appears that research with classical textual analysis has reached the stage of stagnation as its benefits appear to have been fully exploited.

Progresses in NLP in the past few years, however, give new hope for further quantification of unstructured text data. Devlin et al. (2018) propose a promising language presentation model, called Bidirectional Encoder Representations from

Transformers (BERT). The BERT model is designed to pre-train deep bidirectional textual representation from unlabelled text data. Since its introduction, it has been recognized widely as the state-of-the-art language model in various language tasks. The power of the BERT model originates from several parts. First, the massive size of the BERT model is unprecedented: the base BERT model contains 110 million parameters. Second, its deliberately designed neural networks can grasp the complex relationship among words and sentences. The neutral network architecture of the BERT model is based on several encoder layers of the popular *Transformer* model proposed by Vaswani et al. (2017), of which the most important part is the so-called *self-attention* mechanism. Third, the BERT model is pre-trained with unprecedentedly massive text datasets including the BookCorpus and English Wikipedia (Devlin et al. 2018) over two different pre-training tasks.[2] With such a large training input, the BERT model can be pre-trained to the extent that meaningful word or sentence representations can arise.

The BERT model is a transfer learning framework and its usage is often separated into two stages: pre-training and fine-tuning. Various pre-trained BERT models have been pre-trained on different unlabelled text datasets with different training settings and can be accessed by researchers who seek to quantify textual information for their purposes. They can be applied directly to a wide range of down stream tasks such as text classification, named entity recognization and question answering, and has obtained the best results for many language tasks (Devlin et al. 2018). For a specific language task such as sentiment classification, researchers can continue training a pre-trained BERT model with their own labelled datasets.

After the introduction of the original BERT model (Devlin et al. 2018), some more refined and robust BERT-like models, such as RoBERTa (Liu et al. 2019) and ALBERT (Lan et al. 2019), are proposed based on the basic architecture of the BERT model and achieve better performance by slightly modifying some parts of the model design or the pre-training hyper-parameters. These models are also available to scholars and can be further fine-tuned for different language tasks.[3]

Several studies explore the application of the BERT model in ESG research. Aue et al. (2022) demonstrate how the BERT model could help predict ESG ratings by extracting signals from ESG news for US companies. Sokolov et al. (2021) also apply the BERT model to extract signals from 1000 tweets to predict ESG scores and show the potential of building an automated ESG scoring system. Mehra et al. (2022) fine-tune an ESG-BERT model which help predict environmental scores by utilizing information from 10K filings. Chava et al. (2021) leverage RoBERTa to classify ESG topics in earning calls and build an ESG dictionary.

In general, the BERT model helps advance the understanding of the impact of ESG issues on financial markets. However, it also has some limitations for ESG research. For instance, even though the BERT model offers very impressive language processing capabilities, its large size (in terms of large number of parameters) leads

[2] The discussion of the BERT model is not the key part of this study. Please refer to Devlin et al. (2018) and Vaswani et al. (2017).

[3] For example, the *Hugging Face* team maintains a list of pre-trained BERT-like models: https://huggingface.co.

to very high computing resource demand, which may restrict its application in large scale in ESG research. Moreover, like many other large machine learning models, the BERT model has also interpretability challenges. For ESG research, interpretability is of great importance for stakeholders to trust and use model outputs. Fortunately, the recent development in NLP is quite promising, which may help alleviate these problems.

4 Literature review and hypothesis development

While numerous studies report a positive relationship between ESG performance and corporate financial performance (Friede et al. 2015), there is less consensus about how investors value ESG performance on the stock markets. Although the investment community considers ESG information during investment decision-making process (Amel-Zadeh and Serafeim 2018; Van Duuren et al. 2016), the role of ESG issues on financial markets is not well understood (Bennani et al. 2018). Pedersen et al. (2021) theoretically propose an ESG-adjusted CAPM and predict that a security with a higher ESG score has a higher demand from ESG investors, which is also supported by the empirical evidence that ESG performance proxies correlate positively with institutional holdings. Hartzmark and Sussman (2019) examine the relationship between the sustainability rating rankings of the US mutual funds and fund flows and present evidence that investors do value sustainability. Regarding the market performance related to ESG investment, Mănescu (2011) find that only some ESG attributes, such as community relations, have an impact on stock returns by analyzing a long panel dataset of US firms. Bennani et al. (2018) document that the impact of ESG screening on stock performance is highly time-dependent: they find no evidence of a consistent reward for ESG integration during the 2010–2013 period but a significant excess return for the 2014–2017 period.

Despite their different perspectives and results, these earlier studies usually adopt some kind of ESG performance proxies provided by ESG data providers such as ESG rating. Very few studies address the question of whether the market reacts to high-frequency news in the field of ESG studies (see e.g. Capelle-Blancard and Petit 2019; Cui and Docherty 2020), despite the existence of a stream of studies investigating ESG events (Flammer 2013; Naughton et al. 2019; Grewal et al. 2021; Krüger 2015) and ESG incidents (Gantchev et al. 2022; Glossner 2021; Derrien et al. 2021).[4] However, there are a significant number of studies analyzing the relationship between high-frequency financial news and stock markets (Alanyali et al. 2013; Boudoukh et al. 2019). For instance, Alanyali et al. (2013) find that financial news is closely linked to trading movements. Boudoukh et al. (2019) find evidence that there is a close relationship between identified relevant firm-level financial news and stock prices. In particular, the tone of news can be of great importance to investors. Many studies apply semantic analysis to extract sentiment signals in financial news articles and investigate their possible influence. Tetlock (2007) uses a word count program to analyze texts – to investigate the interaction between financial news and the stock

[4] They do not really touch upon high-frequency ESG news in our context.

market – and observes that the extracted media sentiment predicts stock prices and trading volume. In recent years, the development of machine learning techniques has enabled researchers to investigate the role of news tonality on financial markets in deeper detail. Heston and Sinha (2017) measure news sentiment with proprietary neural network and find that daily financial news can predict stock returns for one to two days. Ke et al. (2019) introduce a supervised learning framework that can extract sentiment information from financial news articles and find that those extracted sentiment signals can predict stock returns to a large extent.

Similarly, instant ESG news as an important source of ESG information for (ESG) investors could possibly influence their investment decisions. Positive (negative) ESG news indicates the marginal improvement (deterioration) of company ESG performance and could be considered by investors in two ways. On the one hand, an improvement (deterioration) of ESG performance may lead to an improvement (deterioration) in corporate financial performance (Friede et al. 2015) and thus have an impact on the stock performance via the incorporation of this positive cash flow news into prices. On the other hand, an improvement (deterioration) of ESG performance may attract (repel) ethical investors who have the incentive to promote ESG development (Pedersen et al. 2021). Therefore, we expect that the market reaction to instant ESG news is closely related to the news sentiment.

H1: Positive (negative) instant ESG news is associated with stock over-performance (under-performance).

However, the market reaction to positive and negative ESG news could be different in terms of scale. Capelle-Blancard and Petit (2019) find that companies facing negative ESG news experience a drop of 0.10% in market value, but gain nothing on average from positive ones. Cui and Docherty (2020) also report that the market does not react to positive ESG news but overreacts to ESG controversies by analyzing ESG news processed by Ravenpack. This could be explained by investors' concern that companies have the incentive to exaggerate their ESG performance (Yu et al. 2020). With the increasing attention paid to ESG from various stakeholders, some companies find it beneficial to overstate their commitment to ESG topics (Bazillier and Vauday 2009). For instance, "greenwashing", which describes the intention of companies to label non-green products or practices as green, has been a hot topic in the past two decades (Flammer 2021). Nevertheless, a pretending of unsubstantiated ethical engagement can cause public distrust (Jahdi and Acikdilli 2009). If companies disclose ESG information more frequently or exaggerate their ESG performance, the probability that companies do good to the society decreases or the overall contribution is less valued. Therefore, investors may react less actively to overwhelming positive ESG news. Another explanation can be so-called "negativity bias", in which the market reacts significantly to negative news but remains relatively calm when good news arrives. In psychology, negativity bias refers to the phenomenon that humans give greater weight to negative events, which is manifested in different ways such as negative potency, steeper negative gradients, negativity dominance, and negative differentiation as described by Rozin and Royzman (2001). Several studies examine this negativity bias on the financial markets. Edmans et al. (2007) observe a strong negative stock market reaction to

losses of national sports teams while no evidence of a corresponding reaction to victories. Akhtar et al. (2011) investigate the market responses to consumer sentiment announcements and document the existence of negativity bias on the Australian stock market.

Likewise, it can be expected that the market reactions related to negative and positive ESG news are asymmetric. More precisely, negative ESG news may be perceived more seriously by the market and lead to stronger reactions as compared to positive ESG news. We summarize the hypothesis as follows.

H2: The market reaction related to negative ESG news is stronger than to positive ESG news.

Lastly, we discuss the possible linkage between the historical ESG record and the reaction to instant ESG news. As mentioned above, the ESG score and instant ESG news are two different types of ESG information. The former can be seen as a mid- or long-term ESG record of the company in which all past ESG information is aggregated. As opposed to that, the latter reflects short-term changes of ESG performance. Previous studies indicate that low-frequency ESG performance proxies such as ESG ratings are important to investors (see e.g. Amel-Zadeh and Serafeim 2018; Bennani et al. 2018).

To model the impact of instant ESG news in light of an existing long-term ESG rating, we propose a simple adaptive model to depict how investors adapt their perception of company ESG performance to the arrival of instant ESG news. Considering the fact that ESG agencies often update their ESG ratings based on the aggregated ESG information since the last evaluation period (e.g. Escrig-Olmedo et al. 2019), we propose a steady adaption to the arrival of ESG news. Let, $\mathrm{ESG}_{i,t-1}$ denotes the present ESG performance figure, based on past ESG information, while $esg_{i,t}$ measures the additional ESG contribution inherent in the instant news under consideration. We regard $esg_{i,t}$ as exogenous, while its expected value can depend on the company's past ESG profile to some extent. This is because past ESG ratings may have already embedded some part of future ESG activities, and positive (negative) news is more anticipated for companies with a good (bad) ESG record (Serafeim and Yoon 2021). Also, Glossner (2021) document that companies' past ESG incident rates, which may already be integrated into ESG ratings, predict more future incidents. The new ESG performance $\mathrm{ESG}_{i,t}$ then results as the sum of past ESG performance $\mathrm{ESG}_{i,t-1}$ and the ESG performance change $esg_{i,t}$ due to the news, i.e.:

$$\mathrm{ESG}_{i,t} = \mathrm{ESG}_{i,t-1} + esg_{i,t}\,. \tag{1}$$

Note that the sign of $esg_{i,t}$ is positive (negative) in case of positive (negative) ESG news, while $\mathrm{ESG}_{i,t-1}$ can without loss of generality be assumed to lie between 0 and 100, in which 100 (0) describes a perfectly sustainable (unsustainable) company. Furthermore, usually $\mathrm{ESG}_{i,t}$ is not immediately published by the ESG score provider. However, it can be seen as the theoretical new value for an investor who considers both the old ESG score and the content value of the new instant news.

As for a company with a high ESG score it is less easy to increase its ESG score compared to a company with a low ESG score, we consider the relative ESG performance change

$$\Delta \mathrm{ESG}_{i,t} = \frac{esg_{i,t}}{\mathrm{ESG}_{i,t-1}}. \tag{2}$$

Given the same value of $esg_{i,t}$, it is obvious that $\Delta \mathrm{ESG}_{i,t}$ is higher (lower) for companies with lower $\mathrm{ESG}_{i,t-1}$ when they encounter positive (negative) ESG news. Consequently, the market may behave differently to the same kind of instant news for companies with different past ESG ratings. If ESG performance enhances value, as claimed by H1, then the relative value can increase much more for a company with a low ESG score, while for a company with an already high ESG score positive and negative instant news with the same absolute value $|esg_{i,t}|$ will yield a lower value change. This view is supported by Glück et al. (2021), who argue that companies with a good ESG profile may face diminishing marginal benefits of ESG performance improvement, which is consistent with the over-investment view proposed by Goss and Roberts (2011). Combining the expectation argumentation that companies with a bad ESG record may enjoy even higher ESG performance increase from good ESG news as such news is less anticipated and more surprising to the market, we can expect stronger market reactions for these companies. However, it is less clear regarding how differently the market may react to bad ESG news for companies with different ESG records. On the one hand, the expectation argument indicates that bad ESG news is less anticipated for companies with a good ESG record and thus $|esg_{i,t}|$ may be higher. On the other hand, it should be noted that companies with a good ESG profile are still perceived as doing relatively well despite the slight downgrade of ESG performance (Glück et al. 2021) due to negative ESG news. Several studies (Lins et al. 2017; Shiu and Yang 2017; Bartov et al. 2021) show that an overall good ESG reputation can alleviate the negative impact of negative ESG events. If the latter aspect outweighs the former, we can expect that the market reacts less strongly to ESG news of companies with a good ESG record. To sum up these considerations, we state our third hypothesis as follows.

H3: The market reacts more favorably to positive ESG news of companies with a bad ESG record while less severely to negative ESG news of companies with a good ESG record.

5 Data description

5.1 The uniqueness of the employed ESG news dataset

To show the uniqueness of the ESG news dataset adopted in this study, it is essential to distinguish between ESG events and ESG news. In our context, ESG news is instant and high-frequency information which is untouched and original, while ESG events are usually "significant" events that are identified by data providers. ESG incidents data that is recently often adopted in related research specifically

Table 1 Comparison of different ESG news or events datasets

	N/E	Source	Size	# of stocks	Period
This study	News	Eikon	84,835	13,327	2019–2021
Capelle 2019	News	Covalence	33,000	100	2002–2008
Cui 2020	News	RavenPack	82,435	1500	2008–2018
Krueger 2015	Events	MSCI KLD	2116	745	2001–2007
Derrien 2021	Events	RepRisk	76,538	8054	2008–2019
Glossner 2021	Events	RepRisk	70,735	2900	2007–2017
Gantchev 2022	Events	RepRisk	37,805	N.A.	2007–2016

This table compares different ESG news or incidents datasets in related studies. The column N/E indicates whether the dataset is ESG news or events dataset. Capelle 2019 refers to Capelle-Blancard and Petit (2019); Cui 2020 refers to Cui and Docherty (2020); Krueger 2015 refers to Krüger (2015); Derrien 2021 refers to Derrien et al. (2021); Glossner 2021 refers to Glossner (2021); Gantchev 2022 refers to Gantchev et al. (2022)

refers to negative ESG events. To the best of our knowledge, very few focus directly on instant ESG news (Capelle-Blancard and Petit 2019; Cui and Docherty 2020), while the rest adopt ESG events or incidents datasets (e.g. Krüger 2015; Derrien et al. 2021; Glossner 2021; Gantchev et al. 2022). In Table 1, we compare several different ESG news or incidents datasets in recent related studies. It shows that ESG events or incidents datasets often have much lower frequency as compared to ESG news datasets. Even though we may have some understanding regarding ESG events or incidents (e.g. Krüger 2015; Derrien et al. 2021; Glossner 2021), less is known about how investors react to instant ESG news since its frequency could be far higher than ESG events or incidents. Moreover, most related studies employ proprietary datasets which directly come from data providers. This common approach has several obvious drawbacks. First, proprietary datasets may have relatively lower frequency or coverage, which may lead to biased empirical results. Second, the way ESG data providers process text data is usually opaque and empirical results based on these datasets are therefore also provider-dependent. At last but not least, these datasets are built by ESG data providers based on news sources they have and could be less representative than our ESG news dataset based on general news vendor Thomson Reuters. It is worth mentioning that positive ESG events or news is probably under-represented in the sample adopted by earlier related studies (see Krüger 2015; Capelle-Blancard and Petit 2019; Cui and Docherty 2020). For example, the ratios between the number of positive and negative ESG events or news are only about 0.37 and 2.10 in the studies of Krüger (2015) and Capelle-Blancard and Petit (2019), respectively. In contrast, positive ESG news prevails (8.86 times of negative ones) in our final ESG news sample.

5.2 Building a comprehensive ESG news dataset

In this study, we propose an alternative and general way to obtain a representative ESG news dataset which is less likely to be subject to the above problems. The original raw ESG news dataset is directly extracted from the general data provider Thomson Reuters Eikon which covers more than 10,000 news sources and serve

as one of the most important news vendors in the world. With such a wide news coverage, it is more likely that we consider the majority of instant ESG news. We first build a complete list of stocks (more than 58,000 primarily quoted stocks on Eikon) traded from all over the world and query their raw English ESG news on Eikon one by one in the period from May 2019 to March 2021. In total, we obtain a full original sample of 245,723 raw news entries tagged as ESG news by Thomson Reuters.

Before conducting empirical study, we clean the ESG news dataset in the following steps. First, we remove those ESG news records without a complete title or article text and exact duplicate news identified as those with an exact title or article text as earlier news for the same company. Accordingly, 59,519 ESG news records are dropped from the sample. Second, we further remove ESG news items for which we do not have enough data for conducting event study (i.e. those without stock or index price data). This cleaning procedure leads to a further reduction of 27,846 ESG news items.

The way we construct the ESG news sample makes sure that it is less likely to be subject to serious selection biases. Nevertheless, while we may enjoy the benefit of a wide coverage of instant ESG news, another challenge arises at the same time. There are still many fuzzy duplicate news items in the sample as more than one source may publish similar ESG news on different dates or at different times on the same date. Before further empirical investigation is conducted, we need to further identify and eliminate them.

5.3 Identifying and eliminating fuzzy duplicate ESG news

To tackle the problem of fuzzy duplicate or stale ESG news, we leverage the power of BERT-like language models. We apply the pre-trained *Sentence-BERT* model (Reimers and Gurevych 2019) to derive sentence embeddings of ESG news titles. The *Sentence-BERT* model has already been pre-trained on Natural Language Inference (NLI) datasets SNLI and MultiNLI[5] and can produce meaningful vectors for sentences. Those sentence embeddings derived from the pre-trained model are numeric representations of ESG news titles. Therefore, news titles with similar semantic meanings should be close to each other in such a high-dimensional space.

We take the following steps to figure out fuzzy duplicate or stale news entries. First of all, we sort ESG news items for the same company according to their release timestamp in an ascending order. As news titles generally represent main ideas of news articles, we identify similar ESG news items as those with a relatively high cosine similarity between sentence embeddings of ESG news titles. For each ESG news item, we calculate the cosine similarity between its title sentence embedding and those of ESG news items with an earlier timestamp and the same stock symbol. If we find any earlier ESG news item which has a value of cosine similarity higher than 0.8 with the investigated ESG news, we identify the respective ESG news as fuzzy duplicate or stale ESG news. We repeat this routine until all fuzzy duplicate ESG news items are identified. Table 2 shows some examples to demonstrate how

[5] Data: https://nlp.stanford.edu/projects/snli/.

Table 2 Examples of identifying fuzzy duplicate or stale ESG news

Date	Company	ESG news title	Cosine similarity
2019-05-13 14:41:11	Apple	CBOE Holdings Inc. – US Supreme Court Has Ruled Against Apple In App Store Antitrust Dispute	Base[1]
2019-05-13 23:12:02	Apple	iPhone owners can sue Apple over its apps, US Supreme Court decides Customers argue that company's control over the App Store is unfair	0.8517[2]
2019-11-21 13:58:04	Microsoft	Vattenfall and Microsoft pilot world's first hourly matching of renewable energy	Base
2019-11-22 17:19:33	Microsoft	Sweden : Vattenfall and Microsoft pilot world's first hourly matching (24/7) of renewable energy	0.8535
2020-09-04 08:00:00	Daimler	Daimler AG joins forces with terre des hommes and the responsible mica initiative to improve mica supply chains and eliminate child labour	Base
2020-09-09 13:04:16	Daimler	Daimler collaborates with terre des hommes and responsible mica initiative to improve mica supply chains and eliminate child labour	0.9723

[1] The earlier ESG news for the same company. [2] The cosine similarity between the title of the base (earlier) ESG news and that of the investigated ESG news. Since the similarity is over 0.8, we remove the investigated ESG news

stale ESG news items are recognized. In the end, 73,523 fuzzy duplicate news items are dropped and the final sample consists of 84,835 unique and fresh ESG news items. In this way, we alleviate the problem of duplicate observations to a large extent.

5.4 Sentiment classification with fine-tuned BERT model

Sentiment analysis identifies the overall emotion within the text in order to inspect whether the author holds a positive, neutral, or negative opinion towards the event mentioned in the news article in general. In this study, our ESG news samples from Eikon are classified into three categories: positive, neutral, or negative ESG news based on the classification results of a sentiment classifier. ESG news is classified as positive ESG news when the overall positive emotion or attitude such as praise and recognization is identified while it becomes classified as negative ESG news when it shows negative emotion or attitude such as disappointment and criticism. Otherwise, ESG news without clear indication or direction of sentiment is classified as neutral ESG news.

Sentiment analysis has long been applied in financial studies (see e.g. Kearney and Liu 2014; Li et al. 2014). However, most studies adopt classical dictionary-based sentiment analysis (Kearney and Liu 2014), which is often considered as inefficient to understand texts written by humans. As far as we know, few studies apply a BERT-like language model to do semantic analysis for finance research (e.g. Araci 2019; Bingler et al. 2021). In particular, ours is one of the few studies to introduce the recent ground-breaking development of NLP in the field of ESG studies (e.g. Bingler et al. 2021). To fine-tune an ESG news sentiment classifier, we need an extra training dataset of ESG news tagged with sentiment labels. For this, we first extract raw

Table 3 BERT models for classifying news types and news sentiment

	BERT model I	BERT model II	BERT model III
Description	Distinguish between company news and non-company news in general news sample	Distinguish between company ESG news and non-ESG news	Identify ESG news sentiment (negative, neutral, positive)
Data	20,000 non-company news items from the GDELT Event Database and 20,000 company news items from Thomson Reuters Eikon	20,000 company ESG news items and 20,000 company non-ESG news items from Thomson Reuters Eikon	50,332 ESG news items extracted from the GDELT GEG database with the help of BERT models I and II, in which 5667 are labelled as negative ESG news, 29,862 as neutral ESG news and the rest 14,803 as positive ESG news
Accuracy rate[1]	99%	99%	81%
F1 score[2]	99%, 99%	99%, 99%	66%, 84%, 79%
Max. length	128 word pieces	512 word pieces	512 word pieces
Model variant	RoBERTa	RoBERTa	RoBERTa
Key parameters	training ratio 0.8[3], training epoch 3, learning rate 2×10^{-5}, batch size 8		

[1] Accuracy rate on the evaluation dataset. [2] F1 scores for each label on the evaluation dataset. For BERT model I and BERT model II, the first F1 score is for non-company news (non-ESG news), and the second F1 score is for company (ESG) news. For BERT model III, the first F1 score is for negative ESG news, the second is for neutral ESG news and the third is for positive ESG news. [3] The original news dataset is split into a training dataset (80%) and an evaluation dataset (20%)

news records from an open-source news database called *The GDELT Project*[6]. *The GDELT Project* monitors and collects news articles from nearly every country on the planet and claims to be the largest and most comprehensive open database of human society ever created. We choose *The Global Entity Graph* (GEG), a sub database of *The GDELT Project*, as our training dataset for the sentiment classifier because of its comprehensiveness and richness.[7] Most importantly, this news database has an overall sentiment score for each news article. These news articles have already been processed by Google's nature language API and assigned with document-level sentiment scores.[8] With these sentiment scores available, we can tag news with sentiment labels.

Moreover, since our target is to classify ESG news sentiment, we explicitly focus on company ESG news in the GEG. We adopt a two-step approach to pick up company ESG news from the GEG, as we first extract company news from the whole news universe and then extract company ESG news from the identified company news. Accordingly, we train two other BERT-like classifiers (BERT models I and II) which can tell whether news is company news and whether company news is

[6] See: https://www.gdeltproject.org.

[7] On average, there are more than 100,000 news records collected by the GEG for a single day.

[8] For more information about the sentiment score, see: https://cloud.google.com/natural-language/docs/basics.

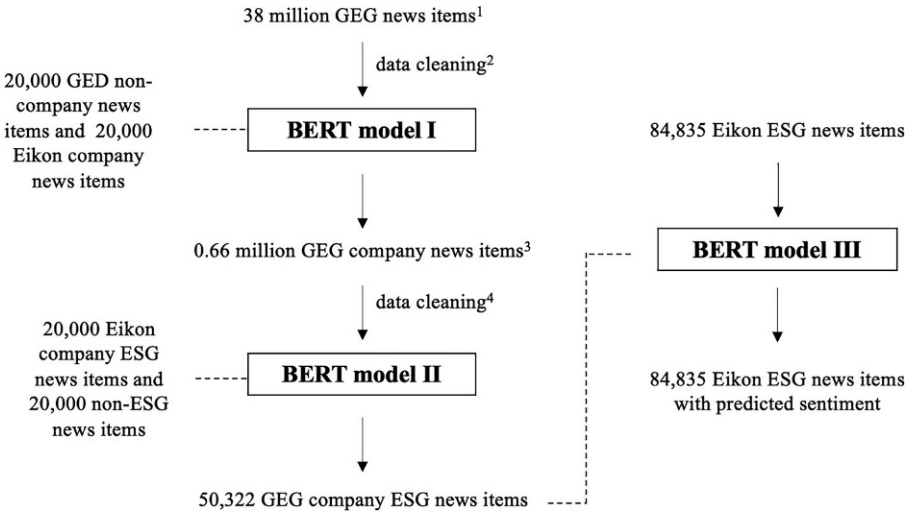

Fig. 1 How a sentiment labelled ESG news dataset is derived. The BERT models I and II are designed to pick up company ESG news from the GEG news universe. These company ESG news items with sentiment labels are used to fine-tune an ESG news sentiment classifier, i.e. BERT model III, which is supposed to help us identify Eikon ESG news sentiment. Arrows indicate the inference process and dash lines refer to corresponding training and evaluation datasets. [1] News titles are extracted from the given URLs and used as model input. [2] News items without a title or exact duplicate news items are removed. [3] Main article texts are scraped down from the corresponding servers and used as model input. [4] News items without a main article text or exact duplicate news items are removed

ESG-related. For fine-tuning the first classifier, we collect 20,000 company news items directly on Eikon and 20,000 non-company news items from another sub database of *The GDELT Project*, i.e. the GDELT Event Database (GED).[9] The GED provides news entries in which the type of event and the major event participants have been identified. We remove those news records with participant types identified as BUS and MNCs[10] and take the rest as non-company news. For fine-tuning the second classifier, we focus on company news exclusively extracted from Eikon. We collect 20,000 ESG news items and 20,000 non-ESG news items by changing the query criterium on Eikon. These two classifiers show the ability (with an accuracy of 99% on the evaluation datasets) to identify whether general news is company news and whether company news is ESG-related (see BERT models I and II in Table 3). With these two additional classifiers, we are able to extract explicitly company ESG news from the massive news sample of the GEG database. We scan over 38 million news records[11] of the GEG published in 2020 and identify 0.66 million company

[9] See: https://www.gdeltproject.org/data.html.

[10] BUS: businessmen, companies, and enterprises, not including MNCs. MNC: multi-national corporations.

[11] The GEG does not provide original text information. We extract news titles from news URLs and use them as input for the first classifier. Therefore, we drop those news entries from which we fail to extract news titles.

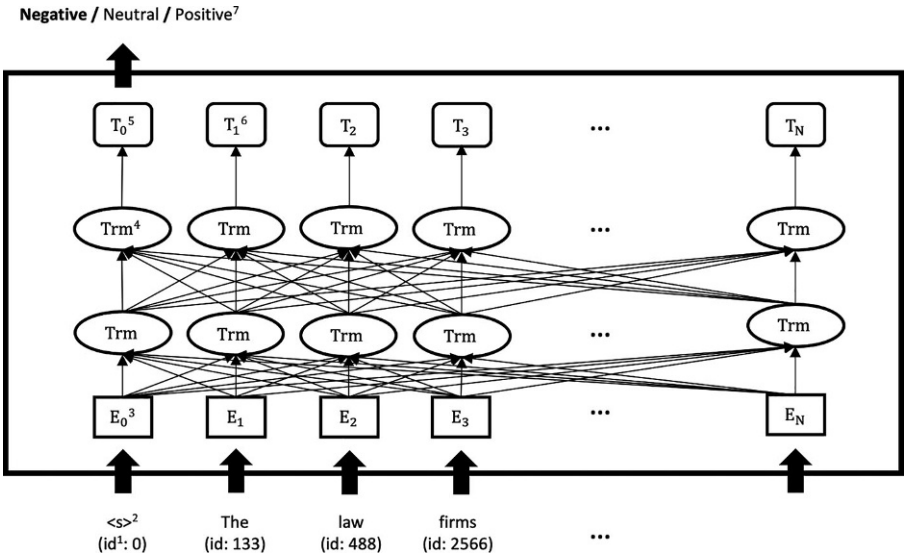

Fig. 2 BERT sentiment classifier: How ESG news is scored. This figure shows the main architecture of the BERT model and how ESG news is classified by the fine-tuned BERT sentiment classifier. The example ESG news is the last ESG news for BMW in Table 4: "The law firms of Waddell Phillips Professional Corporation and Podrebarac Barristers Professional Corporation announced today that ...". The maximum text input length is 512 word pieces. News items with longer text length are truncated. [1] id is the word piece ID in the adopted vocabulary. [2] ⟨s⟩ is a special symbol indicating the beginning of text inputs. [3] The embedding (a 768-dimensional vector) of the input token ⟨s⟩. For detailed implementation of the embedding procedure, please refer to Devlin et al. (2018). [4] The intermediate representations of the token after the *Transformer* encoder layers. For more detail about the *Transformer* encoder layers, please refer to Vaswani et al. (2017). [5] The output of the special symbol (i.e., ⟨s⟩) after all the *Transformer* encoder layers. [6] The output of a word piece after all the *Transformer* encoder layers. [7] The final sentiment label is identified by converting the previous output (corresponding to T_0 in the figure) into probabilities through a Softmax output layer and choosing the highest one

news items[12] with the first classifier, from which we identify 50,332 company ESG news items using the second classifier.

At last, we tag each company ESG news item extracted from the GEG according to its overall sentiment score. For ESG news items with an overall sentiment score not lower than 0.2, we label them as positive and those with an overall sentiment score not higher than −0.2 as negative. The rest of news items in the sample are labelled as neutral ESG news. Note that these ESG news items extracted from the GEG are only used for the purpose of fine-tuning a sentiment classifier to predict sentiment of ESG news items from Eikon.[13] We summarize all the additional news

[12] We try to scrape down original article texts for these company news items and use them as inputs for the second classifier. If it fails, we drop these news items.

[13] One proposal would be the usage of ESG news in the GEG as our research subject. However, this could be infeasible because these news items are not paired with stock tickers and thus empirical studies are not possible.

Table 4 Examples of ESG news sentiment classification

Date	Company	ESG news text	Sentiment
2019-06-05 07:35:29	BMW	As it develops its plans for the mobility of the future, the BMW Group is increasingly focusing on co-operations to help make next-level electrification technology more widely available to customers by the start of the coming decade...	Positive
2019-09-14 20:19:16	BMW	BERLIN, Sept 14 (Reuters) – BMW's engine development and purchasing expert, Markus Duesmann, is set to become the CEO of Volkswagen's Audi premium brand....	Neutral
2020-04-07 16:17:54	BMW	The law firms of Waddell Phillips Professional Corporation and Podrebarac Barristers Professional Corporation announced today that the Ontario Superior Court of Justice has certified a national class action against luxury automaker Bayerische Motoren Werke AG....	Negative
2020-03-24 15:26:23	Dow, Inc	Mar 24, 2020. Dow Inc. introduced two innovations that simplify the formulation of water-based, high temperature-resistant industrial coatings...	Positive
2020-06-17 12:35:54	Dow, Inc	Jun 17, 2020. Dow Inc. inked a joint development deal with Shell to speed up the development of technology that can electrify ethylene steam crackers....	Neutral
2020-05-21 03:06:51	Dow, Inc	Catastrophic flooding triggered by dam failures in Michigan could potentially release toxic pollution from a site contaminated by the industrial giant Dow Chemical.Dow's facility in Midland, Michigan, where the company is headquartered along the Tittabawassee River, manufactured chlorine-based products beginning in the early 1900s...	Negative

datasets and how we derive a labelled ESG news dataset as described above in Fig. 1.

Given the sentiment labelled ESG news, we can finally fine-tune a BERT-like model (BERT model III) and further apply it to infer sentiment of ESG news from Eikon, with which further empirical examination is conducted. We choose a maximum possible text length of 512 word pieces, which means that news articles more than 512 word pieces will be truncated. For more details regarding the model, please refer to Table 3 and Fig. 2. With an accuracy rate of 81% and relatively high F1 scores on the evaluation set, our fine-tuned BERT model III, for most of the time, is able to determine the overall sentiment direction of company ESG news. In fact, this model performance is quite satisfying, especially given the fact that the text input is relatively long (i.e. 512 word pieces) and there exist three sentiment labels instead of two with only negative or positive sentiment[14]. Da Silva et al. (2014) review many studies applying classical machine learning models which aim to classify tweets (relatively short texts) as positive or negative ones (only two labels) and document that most of the time the accuracy rates of these models are lower than 80%. Therefore, we are confident that this sentiment classifier can provide satisfying classification

[14] It is more difficult for models with three labels to achieve better results than those with only two labels, as they need to extract more useful information from data to distinguish among three classes.

Table 5 Descriptive statistics for categorical variables

	Obs.	%	Stocks	%		Obs.	%	Stocks	%
Continent									
America	45,688	53.86	5085	38.16	Consumer Discretionary	8371	9.87	1406	10.55
Europe	22,926	27.02	3387	25.41	Consumer Staples	4129	4.87	666	5.00
Asia	13,269	15.64	4033	30.26	Health Care	7137	8.41	1486	11.15
Oceania	2275	2.68	563	4.22	Financials	9499	11.20	1527	11.46
Africa	677	0.80	259	1.94	Information Technology	10,077	11.88	1457	10.93
					Communication Services	3521	4.15	606	4.55
Sector					Utilities	7038	8.30	406	3.05
Energy	6394	7.54	742	5.57	Real Estate	2561	3.02	746	5.62
Materials	9853	11.61	1701	12.76	None[1]	1092	1.29	442	3.32
Industrials	15,161	17.87	2142	16.07					

[1] None means there is no GICS sector classification

results and differentiate ESG news with different sentiments.[15] Moreover, to check whether the BERT model is particularly appropriate in our context, we apply other representative NLP models and compare their performance with that of the BERT model (see Sect. 9.3). Our model performance comparison shows the superiority of the BERT model over other alternative NLP models in terms of having the highest accuracy rates and F1 scores in all language tasks. To further validate the performance of the BERT model, we conduct a human audit on a subsample of 120 ESG news. See Sect. 9.4 for details. Overall, the sentiment judgement capability of the BERT model is close to human judgement, which is also supported by Fischbach et al. (2022).

Finally, we feed all unique Eikon ESG news into BERT model III and classify them into positive, neutral, and negative ESG news. See some representative ESG news entries with different sentiments in Table 4.

5.5 Basic descriptive statistics

In total, the final ESG news sample contains 84,835 ESG news items from 13,327 listed companies from all over the world. In Table 5, we show where ESG news originates from. More than half of the ESG news items final sample come from America, while around 27% items from Europe. Asia and Oceania also have a share of 16% and 3%, respectively. Moreover, we show the number of ESG news items for the top 20 countries (regions) in the full sample in Table 18. USA has the biggest share of 42% of the overall sample, followed by Canada with 11% and UK with 9%. Our sample covers almost every corner of the world and should be represen-

[15] Even when people are asked to do the classification task, it is possible that they may be inconsistent in judging news sentiment and have different opinions.

Table 6 Descriptive statistics for metric variables

	Obs.	Mean	Std.	Min	Median	Max
asset[1] (in million USD)	82,571	69,700	248,000	0.01	6020	4,320,000
esg[1]	50,722	59.08	21.51	0.92	63.15	94.47
num_news[2]	84,835	32.00	41.53	1	16	302

[1] *asset* and *esg* are firm-level data at the end of previous year of the event date. [2] *num_news* indicates how many pieces of ESG news are released for the same company during the sample period

Table 7 Sentiment distribution of ESG news

Continent	Obs.	Negative	%	Neutral	%	Positive	%
America	45,688	2147	4.70	23,171	50.72	20,370	44.59
Europe	22,926	854	3.73	11,576	50.49	10,496	45.78
Asia	13,269	828	6.24	6711	50.58	5730	43.18
Oceania	2275	355	15.60	1265	55.60	655	28.79
Africa	677	37	5.47	479	70.75	161	23.78
Total	84,835	4221	4.98	43,202	50.92	37,412	44.10

This table presents the ESG news sentiment distribution across different continents. In toal, there are 84,835 unique and fresh ESG news items for 13,327 stocks

tative to study the pricing implication of ESG news. As regards sector distribution, the top five sectors are Industrials (18%), Information Technology (12%), Materials (12%), Financials (11%), and Consumer Discretionary (10%), according to the Global Industry Classification Standard (GICS).

In Table 6, we provide more basic descriptive statistics for company level features. Note that these company basic features are from the previous year's end for each ESG news items. Overall, ESG news has an average company asset of 69.7 billion USD. Only 50,722 out of 84,835 ESG news items are paired with an Eikon ESG score. On average, ESG news has an ESG score of 59.

Moreover, we show the sentiment distribution of ESG news in Table 7. Overall, 44% ESG news items are classified as positive news by our sentiment classifier, but only 5% as negative news. The remaining half of ESG news items are classified as neutral ESG news, which means that there is no clear positive or negative sentiment revealed in texts in our context. The sentiment distributions of ESG news for America, Europe, and Asia are similar. Negative ESG news items contribute to only 4% to 6% of the corresponding continent subsamples, except for Oceania (16%).

Unlike other ESG news samples adopted in related studies (e.g. Krüger 2015; Capelle-Blancard and Petit 2019), our sample is constructed based on massive raw ESG news from comprehensive sources from all over the world. Therefore, it is much more representative and should reflect how company ESG issues are reported as a whole. As positive ESG news items clearly outweigh negative ones in our ESG news sample (no matter in the overall sample or different subsamples for different continents), it can be said that in general news media prefer to report positive ESG issues rather than negative ones. This could be partly explained by the possibility that companies may have the intention to systematically exaggerate their ESG performance due to the increasing pressure from various stakeholders. Accordingly, investors may be aware of this problem and treat positive and negative

ESG news differently. Given the sentiment classification result of the three groups of ESG news, we investigate whether there is stock performance differences among them and what are the possible determinants.

6 Empirical methodology

6.1 Event study and discussion of confounding events

In order to examine the pricing implication of ESG news, we conduct event study for each ESG news item. We define the day when ESG news is released as the event day T_0, and choose an event window which covers a period several days τ before and after the event day, i.e. from $T_0 - \tau$ to $T_0 + \tau$. We follow a standard event study procedure in which we calculate abnormal returns and cumulative abnormal returns to measure stock performance during different event windows (see Sect. 9.1 for a detailed explanation of related calculations). Moreover, we adopt the correlation robust t-statistic proposed by Kolari et al. (2018) to test the statistical significance of stock performance (see Sect. 9.2 for further explanation).

One concern of event studies is confounding events. In our context this means that synchronous non-ESG news could have an impact on the financial markets and thus could blur the real influence of ESG news. However, we regard this as very unlikely for the following reason. Even though it is not possible to completely rule out the possible influence of non-ESG news as we aim to examine high-frequency ESG news, we argue that empirical results is very unlikely to be driven by non-ESG news. As mentioned in Sect. 5, our dataset is very comprehensive and covers most ESG news in the observation period for more than 10,000 listed companies from all over the world. If the empirical results were driven by confounding events, the non-ESG news would need to be aligned with the ESG news in a systematic way. More precisely, positive (negative) ESG news would need to be systematically accompanied by positive (negative) non-ESG news from the same company—published close to the announcement date. However, while such a news disclosure behavior is unlikely but possible for any arbitrary company, there is no reason to assume that thousands of companies behave in the same manner[16]. For this reason, we hold that non-ESG news within the event window is diverse in nature and sentiment and therefore its influence on the results cancels out within our large samples. In other words, confounding events can be a problem for event studies with a relatively small number of events and a small number of different stocks. Neither is the case here.

6.2 Regressions

Apart from event study, we regress stock performance, measured by abnormal returns on the event day or cumulative abnormal returns of different event windows,

[16] To empirically examine the above argument, we additionally analyze the release pattern of instant non-ESG news which concerns the same company and is published during the ESG news event window. The results (not attached here) show that confounding events are less of a concern in our research setting.

on several independent variables to investigate whether news sentiment is a key determinant of stock performance. Moreover, we are interested in whether the past ESG ratings as assigned by ESG raters such as Thomson Reuters may have an impact on stock performance when instant ESG news is released. The regression setup is as follows:

$$R_i = \beta_0 sentiment_i + \beta_1 esg_i + \beta_2 sentiment_i \cdot esg_i + \beta_3 controls_i + e_i \quad (3)$$

in which R_i represents stock performance measured by abnormal returns ar_0 on the event day T_0 or by cumulative abnormal returns CAR_1 and CAR_2. The variable $sentiment_i$ represents the overall ESG news sentiment, i.e. positive, neutral, or negative sentiment, as predicted by our fine-tuned BERT model III. The variable esg_i is the Eikon ESG score for the company under investigation. We include interaction terms between $sentiment_i$ and esg_i to further test whether their impact on stock performance is intertwined, as predicted by H3. As regards control variables $controls_i$, we have the following setups. To control for possible size effect, we include the variable $asset_i$ in regressions. We also add num_news_i, which indicates the number of ESG news items for the same company in the sample period to control difference in media exposure. We further add $sector_i$ and $continent_i$ to control for sector and geographic differences. For detailed explanation of variable definitions, please refer to Table 19.

7 Results

7.1 Event study results from the overall sample

We show descriptive statistics of stock performance as abnormal return on the event day T_0 and cumulative abnormal returns during different sizes of event windows for each group of ESG news in Table 8. Note that we adopt robust t-statistics to test whether stock performance is significantly different from zero as described in Sect. 6. On average, the positive group shows a significant 0.31% average abnormal return, despite its disproportionally high percentage in the sample. This finding is different from that of earlier studies (Krüger 2015; Capelle-Blancard and Petit 2019), which indicate that investors do not appreciate positive ESG news or events. Moreover, the negative group has a significant −0.75% average abnormal return on the event day, indicating that investors react even more strongly to much less frequent negative ESG news. The neutral group has a relatively smaller scale of average abnormal return of 0.20% on the even day T_0. The univariate analysis on the event day provides evidence that positive ESG news is associated with outperformance while negative news may lead to underperformance, especially on the event day. Moreover, we observe that the market reactions to positive and negative ESG news may be asymmetric. This provides first evidence supporting H1 and H2. When stock performance is evaluated by CAR_1, positive ESG news leads to an average cumulative abnormal return of 1.17% while the negative group suffers from a significant loss of −1.28%. Again, neutral ESG news shows a smaller average cumulative abnormal return. When we

Table 8 (Cumulative) abnormal returns

Group	Obs.	Mean (%)	S.D. (%)	t_τ	CI (%)	Min (%)	Med (%)	Max (%)
ar_0								
Positive	37,412	0.31 ***	4.30	8.36	(0.21, 0.41)	−62.65	0.05	165.16
Neutral	43,202	0.20 ***	6.96	3.02	(0.03, 0.37)	−152.95	0.04	262.61
Negative	4221	−0.75 ***	7.03	−4.25	(−1.20, −0.30)	−119.23	−0.19	53.32
CAR_1								
Positive	37,412	1.17 ***	14.99	11.44	(0.90, 1.43)	−173.88	0.41	225.16
Neutral	43,202	0.87 ***	20.14	5.24	(0.44, 1.30)	−375.77	0.27	237.69
Negative	4221	−1.28 ***	21.89	−3.24	(−2.30, −0.26)	−341.51	−0.63	154.03
CAR_2								
Positive	37,412	1.24 ***	15.64	9.36	(0.90, 1.58)	−196.96	0.46	242.88
Neutral	43,202	0.97 ***	20.87	4.24	(0.38, 1.56)	−365.50	0.28	215.70
Negative	4221	−1.26 ***	22.66	−2.72	(−2.45, −0.07)	−358.13	−0.54	135.72
CAR_5								
Positive	37,412	0.87 ***	11.34	3.73	(0.27, 1.47)	−212.76	0.26	188.84
Neutral	43,202	0.51	16.28	1.28	(−0.52,1.54)	−361.99	0.10	231.45
Negative	4221	−1.42 **	19.80	−2.26	(−3.04,0.20)	−414.32	−0.53	155.64
CAR_{10}								
Positive	37,412	1.24 ***	15.64	2.89	(0.13, 2.35)	−196.96	0.46	242.88
Neutral	43,202	0.97	20.87	1.53	(−0.66,2.60)	−365.50	0.28	215.70
Negative	4211	−1.26	22.66	−1.51	(−3.40,0.88)	−358.13	−0.54	135.72

This table presents descriptive statistics of abnormal returns on the event day or cumulative abnormal returns during the event window. In total, there are 84,835 unique and fresh ESG news items for 13,327 stocks. The CI (confidence interval) is based on 99% confidence level, and the standard error used for the calculation of confidence interval is described in Eq. (9). Test results based on t-statistic t_d in Eq. (13) indicating whether the mean of the positive group is statistically different from that of the negative group: ar_0***, CAR_1***, CAR_2***, CAR_5***, CAR_{10}***. Note that * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

further expand the window size, i.e. change CAR_1 to CAR_2 and CAR_5, we obtain similar result patterns but do not see more obvious performance difference. For CAR_{10}, only the positive group has a significant mean cumulative abnormal return of 1.24%.

Next, we show the average abnormal returns across all ESG news for the whole event window in Fig. 3. The difference between the negative group and the other two groups is obvious. The stock performance of the negative group is most significantly negative on the event day and one day before. For the positive group, we observe notably positive abnormal return only on the event day. In contrast, the neutral group shows a milder performance throughout the whole event window. In Fig. 4, we show cumulative abnormal returns. The performance difference between the negative group and the other two groups is evident. The difference between the positive group and the neutral group only becomes more clear on the event day and thereafter. Additionally, we observe that there is a small drift for positive ESG news while a reversal for negative ESG news after the event day. This indicates that the market underreacts to positive ESG news but overreacts to negative ESG news. The rationale behind this observation could be that investors become less sensitive to

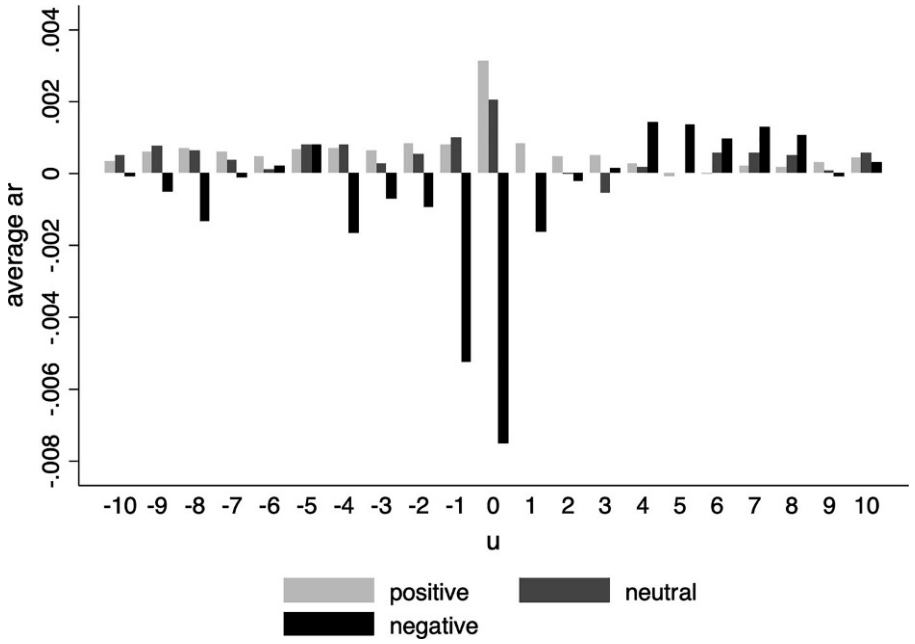

Fig. 3 Abnormal returns during $(T_0 - 10, T_0 + 10)$

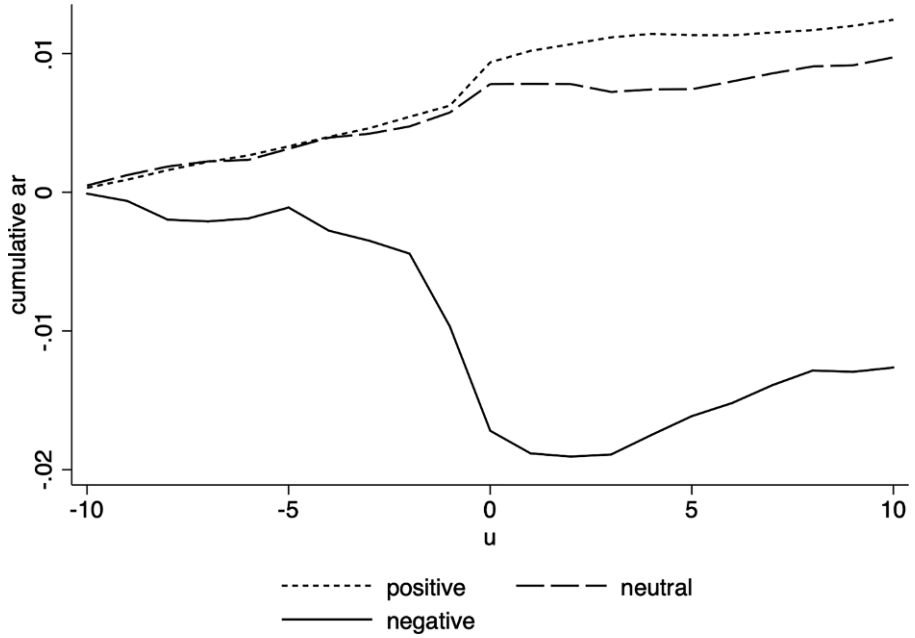

Fig. 4 Cumulative abnormal returns during $(T_0 - 10, T_0 + 10)$

Table 9 (Cumulative) abnormal returns: the America subsample

Group	Obs.	Mean (%)	S.D. (%)	t_τ	CI (%)	Min (%)	Med (%)	Max (%)
ar_0								
Positive	20,370	0.37***	4.84	6.45	(0.22, 0.52)	−62.65	0.08	165.16
Neutral	23,171	0.21*	8.05	1.92	(−0.01,0.49)	−152.95	0.04	262.61
Negative	2147	−1.01***	8.00	−3.70	(−1.71, −0.31)	−119.23	−0.33	53.32
CAR_1								
Positive	20,370	1.38***	16.84	8.81	(0.98, 1.78)	−149.62	0.46	216.51
Neutral	23,171	0.88***	22.64	3.28	(0.19, 1.57)	−199.69	0.29	219.05
Negative	2147	−2.10***	22.71	−3.60	(−3.60, −0.60)	−188.76	−1.11	154.03
CAR_2								
Positive	20,370	1.46***	17.48	7.19	(0.94, 1.98)	−162.49	0.49	242.88
Neutral	23,171	1.05***	23.34	2.72	(0.06, 2.04)	−190.68	0.33	215.70
Negative	2147	−2.07***	23.80	−2.92	(−3.90, −0.24)	−188.09	−0.98	135.72
CAR_5								
Positive	20,370	1.00***	12.71	2.80	(0.08, 1.92)	−122.81	0.30	188.84
Neutral	23,171	0.40	18.19	0.59	(−1.35,2.15)	−199.82	0.11	231.45
Negative	2147	−2.14**	20.06	−2.22	(−4.62,0.34)	−199.82	−0.85	155.64
CAR_{10}								
Positive	20,370	1.46**	17.48	2.20	(−0.25,3.17)	−162.49	0.49	242.88
Neutral	23,171	1.05	23.34	1.01	(−1.63,3.73)	−190.68	0.33	215.70
Negative	2147	−2.07	23.80	−1.49	(−5.65,1.51)	−188.09	−0.98	135.72

This table presents descriptive statistics of abnormal returns on the event day or cumulative abnormal returns during the event window for the America subsample. In the America subsample, there are 45,688 unique and fresh ESG news items for 5085 stocks. The CI (confidence interval) is based on 99% confidence level, and the standard error used for the calculation of confidence interval is described in Eq. (9). Test results based on t-statistic t_d in Eq. (13) indicating whether the mean of the positive group is statistically different from that of the negative group: ${ar_0}^{***}$, ${CAR_1}^{***}$, ${CAR_2}^{***}$, ${CAR_5}^{***}$, ${CAR_{10}}^{**}$. Note that $^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

positive ESG news due to the tendency to exaggerate the ESG performance (see e.g. Flammer 2021) and much more sensitive to negative ESG news due to the documented negativity bias (see e.g. Edmans et al. 2007; Akhtar et al. 2011). Another reason could be that negative ESG news may contain more material information than positive ESG news.

7.2 Event study results from the America subsample

The America subsample contributes 54% of the overall sample and thus is our main focus in this study. Overall, the America subsample shows a similar or even clearer picture. Table 9 presents the average one-day abnormal return on the event day and cumulative abnormal returns for whole event windows. In the America subsample, the difference among different ESG news groups is more evident. The positive group may enjoy an average abnormal return of 0.37% on the event day, while the negative group is associated with a stronger and negative abnormal return of −1.01%. When we look at the cumulative abnormal returns for different sizes

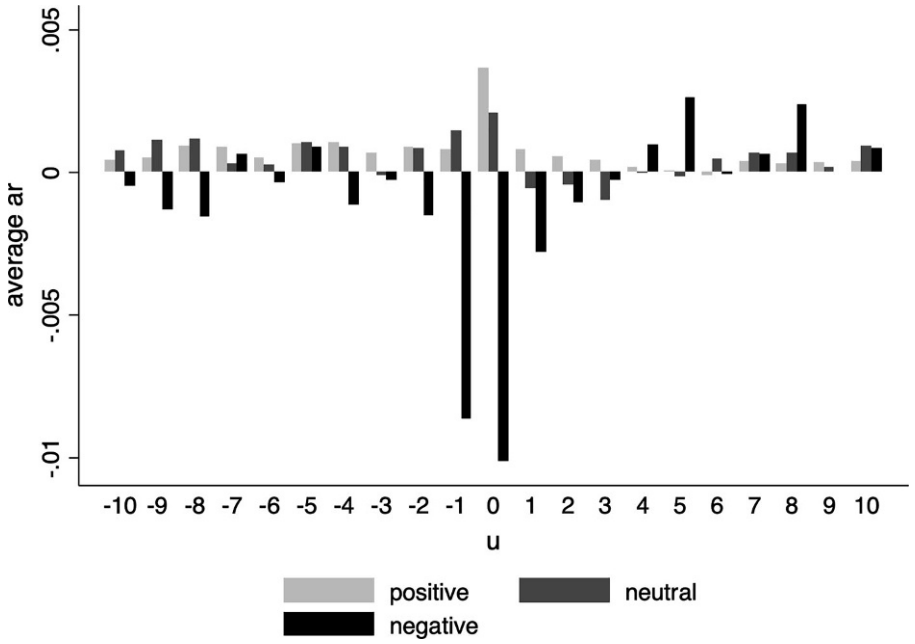

Fig. 5 Abnormal returns during $(T_0 - 10, T_0 + 10)$: the America subsample

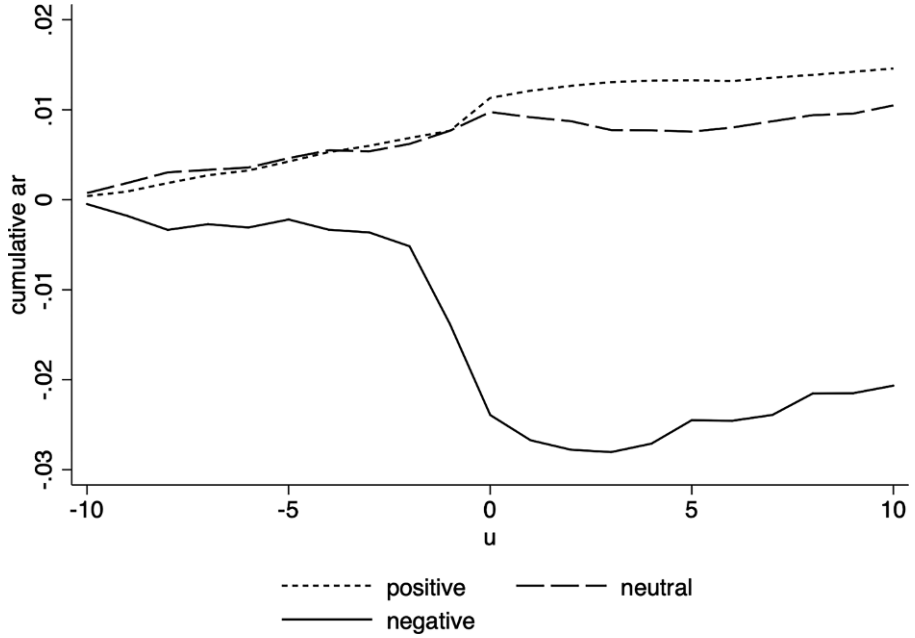

Fig. 6 Cumulative abnormal returns during $(T_0 - 10, T_0 + 10)$: the America subsample

Table 10 (Cumulative) abnormal returns: the Europe subsample

Group	Obs.	Mean (%)	S.D. (%)	t_τ	CI (%)	Min (%)	Med (%)	Max (%)
ar_0								
Positive	10,496	0.34***	3.91	6.10	(0.20, 0.48)	−60.78	0.06	142.26
Neutral	11,576	0.19**	5.79	2.23	(−0.03,0.41)	−143.91	0.05	90.93
Negative	854	−0.78%**	6.33	−2.46	(−1.60,0.04)	−61.96	−0.16	21.55
CAR_1								
Positive	10,496	1.16***	12.40	8.72	(0.82, 1.50)	−128.03	0.62	225.16
Neutral	11,576	0.45**	17.40	2.36	(−0.04,0.94)	−375.77	0.21	237.69
Negative	854	−1.82**	24.98	−1.96	(−4.21,0.57)	−341.51	−0.23	83.19
CAR_2								
Positive	10,496	1.25***	12.96	8.19	(0.86, 1.64)	−140.59	0.59	195.43
Neutral	11,576	0.44*	18.25	1.91	(−0.15,1.03)	−365.50	0.18	210.15
Negative	854	−1.54	25.87	−1.51	(−4.17,1.09)	−358.13	0.10	87.04
CAR_5								
Positive	10,496	0.83***	9.43	4.44	(0.35, 1.31)	−105.73	0.33	141.58
Neutral	11,576	0.40	14.54	1.16	(−0.49,1.29)	−361.99	0.10	176.54
Negative	854	−2.13*	24.90	−1.75	(−5.26,1.00)	−414.32	−0.14	70.40
CAR_{10}								
Positive	10,496	1.25***	12.96	3.73	(0.39, 2.11)	−140.59	0.59	195.43
Neutral	11,576	0.44	18.25	0.73	(−1.11,1.99)	−365.50	0.18	210.15
Negative	854	−1.54	25.87	−1.09	(−5.18,2.10)	−358.13	0.10	87.04

This table presents descriptive statistics of abnormal returns on the event day or cumulative abnormal returns during the event window for the Europe subsample. In the Europe subsample, there are 22,926 unique and fresh ESG news items for 3387 stocks. The CI (confidence interval) is based on 99% confidence level, and the standard error used for the calculation of confidence interval is described in Eq. (9). Test results based on t-statistic t_d in Eq. (13) indicating whether the mean of the positive group is statistically different from that of the negative group: ar_0***, CAR_1***, CAR_2***, CAR_5**, CAR_{10}**. Note that * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

of event windows (CAR_1, CAR_2 and CAR_5), we see significant difference between the positive and negative groups. Specifically, the negative group suffers an average cumulative abnormal return of −2.10% and the positive group enjoys 1.38% over a three-day event window. Again, we find evidence supporting H1 and H2, which state that stock performance of ESG news is related to the news sentiment and stock performance is asymmetric for positive and negative ESG news. Figs. 5 and 6 show daily and cumulative abnormal returns during the event window for the America subsample.

7.3 Event study results from the Europe subsample

Besides the America subsample, we take a closer look at the Europe subsample. With a share of 27%, it is the second largest subsample. We examine the Europe subsample with special care, also because English is popular or often the official language for European countries. As always, we investigate the stock performance of the three groups of ESG news in terms of abnormal return on the event day

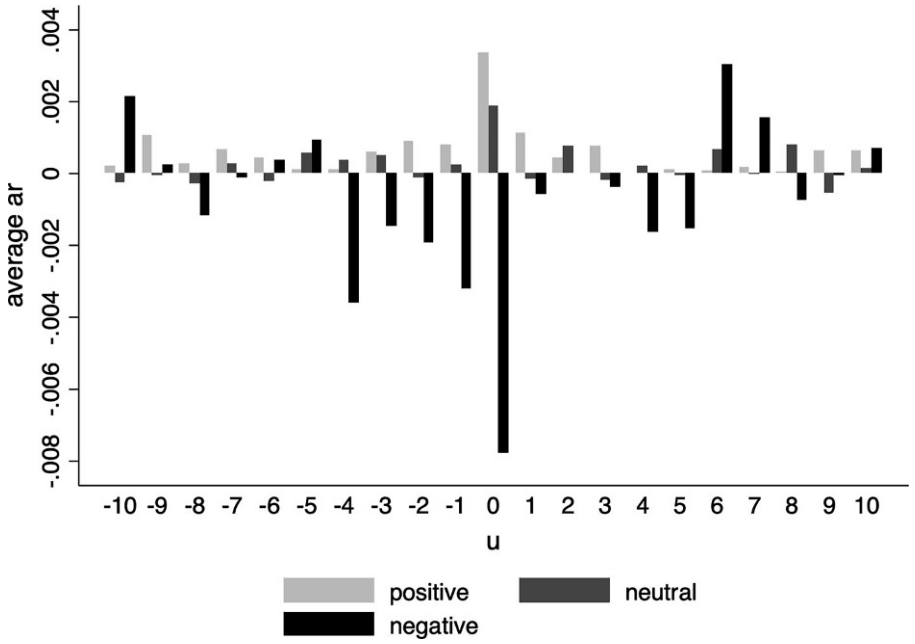

Fig. 7 Abnormal returns during $(T_0 - 10, T_0 + 10)$: the Europe subsample

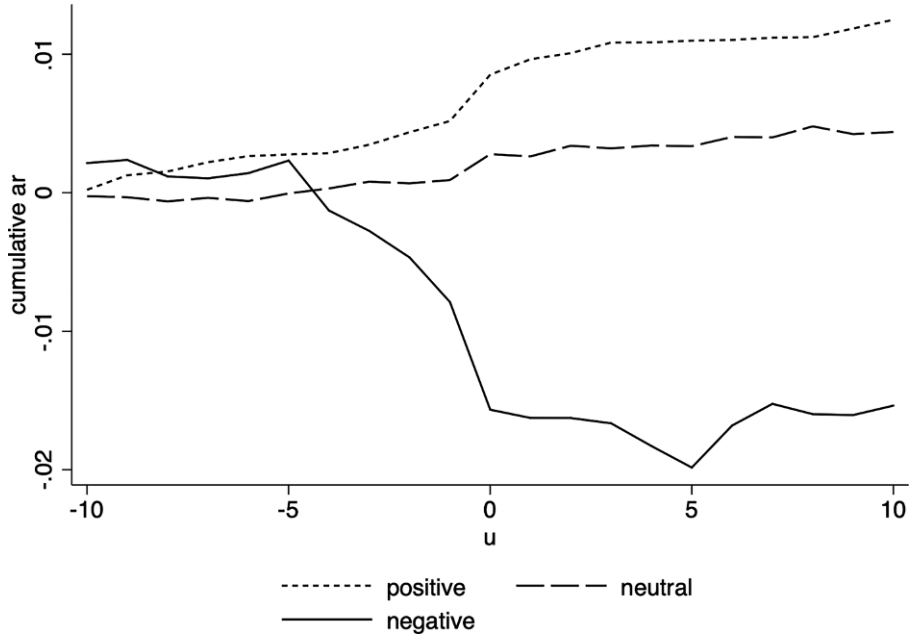

Fig. 8 Cumulative abnormal returns during $(T_0 - 10, T_0 + 10)$: the Europe subsample

and cumulative abnormal returns for the whole event window (see Table 10). The positive group enjoys a significant average abnormal return of 0.34%. In contrast, the negative group is associated with a significant negative average abnormal return of −0.78%. When stock performance is measured over a small event window (CAR_1), the positive group enjoys 1.16% average cumulative abnormal returns while the negative group suffers from a mean loss of −1.82%.

Following the same examination procedure, we present daily and cumulative abnormal returns for the whole event window in Figs. 7 and 8. The results for the Europe subsample show a very similar pattern to the America subsample. Apart from the distinct difference between the negative and the other two groups, the difference between the positive and neutral groups is more observable.

Overall, we find evidence in favor of H1 and H2 not only for the overall sample, but also for the America and Europe subsamples.

7.4 Regression results

Besides event study, we run multiple linear regressions to investigate possible determinants of stock performance related to ESG news. We regress stock performance on ESG news sentiment, ESG score, interaction terms between sentiment and ESG score, and some other control variables. Note that we choose cluster-robust standard errors at the company level in all regressions. Table 11 shows the regression results for the overall sample. We regress stock performance, i.e. ar_0, CAR_1, and CAR_2 on possible determinants and controls. In models I, III, and V, we include only the categorical sentiment variable *sentiment* and controls. In models II, IV, and VI we add interaction terms between *sentiment* and *esg* to check whether the influence of ESG news sentiment depends on the past ESG reputation of the target company.

Overall, we find evidence that whether ESG news sentiment is negative or positive has a significant effect on stock performance, which is in favor of H1. First of all, we find that the coefficient of *negative* is significantly negative across different model setups, which means that the release of negative ESG news has a noticeable and negative impact on stock performance, as compared to neutral ESG news. This indicates that negative ESG news is perceived seriously and priced by investors on stock markets. Regarding positive ESG news, we can observe significant and positive coefficients of *positive* in all models. This is evidence that positive ESG news is digested in a positive way on financial markets. Despite the fact that positive ESG news prevails, it is still positively perceived by investors. Nevertheless, when compared to *negative*, *positive* has obviously smaller coefficients across different models and thus the impact of positive ESG news may be lower than that of negative ESG news. This provides some support for H2.

When interaction terms between *sentiment* and *esg* are added, we gain more insight into market reactions to ESG news under different conditions. Interestingly, the coefficients of the interaction term *negative*esg* are significantly positive in models II, IV, and VI. One possible explanation is that the past ESG record of the company may play a role in the impact of negative ESG news on stock performance. If a company has a good ESG record, the negative impact of negative ESG news could be softened. Therefore, even though a company may suffer from bad stock

Table 11 Regressions: the overall sample

	ar_0		CAR_1		CAR_2	
	(I)	(II)	(III)	(IV)	(V)	(VI)
H1/H2						
sentiment						
negative	−0.0081***	−0.0383**	−0.0187***	−0.1484**	−0.0187***	−0.1572***
	(0.0017)	(0.0152)	(0.0055)	(0.0576)	(0.0053)	(0.0546)
positive	0.0013***	0.0159***	0.0056***	0.0364**	0.0061***	0.0346*
	(0.0004)	(0.0047)	(0.0017)	(0.0183)	(0.0018)	(0.0189)
H3						
negative*esg		0.0076**		0.0327**		0.0349***
		(0.0037)		(0.0139)		(0.0132)
positive*esg		−0.0037***		−0.0077*		−0.0071
		(0.0011)		(0.0044)		(0.0045)
Controls						
esg	0.0014*	0.0028**	−0.0012	0.0007	−0.0012	0.0004
	(0.0007)	(0.0011)	(0.0032)	(0.0044)	(0.0033)	(0.0046)
asset	−0.0003	−0.0003	−0.0011	−0.0012	−0.0013	−0.0014
	(0.0002)	(0.0002)	(0.0008)	(0.0009)	(0.0009)	(0.0009)
num_news	−0.0003	−0.0002	−0.0011	−0.0010	−0.0008	−0.0009
	(0.0006)	(0.0006)	(0.0024)	(0.0025)	(0.0026)	(0.0028)

Table 11 (Continued)

	ar_0		CAR_1		CAR_2	
	(I)	(II)	(III)	(IV)	(V)	(VI)
continent						
Africa	−0.0008	−0.0010	0.0100	0.0096	0.0108	0.0103
	(0.0025)	(0.0024)	(0.0120)	(0.0120)	(0.0136)	(0.0136)
Asia	0.0003	0.0004	0.0013	0.0014	0.0014	0.0015
	(0.0007)	(0.0007)	(0.0039)	(0.0039)	(0.0039)	(0.0040)
Europe	0.0006	0.0007	0.0022	0.0025	0.0022	0.0025
	(0.0005)	(0.0005)	(0.0020)	(0.0020)	(0.0021)	(0.0021)
Oceania	0.0049***	0.0044***	0.0070	0.0052	0.0067	0.0048
	(0.0013)	(0.0013)	(0.0056)	(0.0056)	(0.0057)	(0.0057)
sector						
communication_services	0.0016	0.0017	−0.0036	−0.0032	−0.0039	−0.0034
	(0.0016)	(0.0016)	(0.0050)	(0.0050)	(0.0053)	(0.0052)
consumer_discretionary	−0.0003	−0.0004	0.0061	0.0061	0.0073	0.0073
	(0.0008)	(0.0008)	(0.0054)	(0.0055)	(0.0054)	(0.0054)
consumer_staples	0.0001	0.0001	−0.0028	−0.0027	−0.0031	−0.0030
	(0.0008)	(0.0008)	(0.0031)	(0.0031)	(0.0033)	(0.0033)
energy	0.0012	0.0014	0.0030	0.0033	0.0032	0.0034
	(0.0011)	(0.0011)	(0.0049)	(0.0049)	(0.0051)	(0.0051)

Table 11 (Continued)

	ar_0		CAR_1		CAR_2	
	(I)	(II)	(III)	(IV)	(V)	(VI)
financials	−0.0005	−0.0005	−0.0000	−0.0002	−0.0001	−0.0003
	(0.0009)	(0.0009)	(0.0050)	(0.0050)	(0.0051)	(0.0051)
health_care	−0.0041***	−0.0041***	−0.0070*	−0.0068	−0.0083*	−0.0081*
	(0.0011)	(0.0011)	(0.0042)	(0.0044)	(0.0044)	(0.0045)
information_technology	−0.0023**	−0.0024***	−0.0148***	−0.0151***	−0.0157***	−0.0161***
	(0.0009)	(0.0009)	(0.0032)	(0.0032)	(0.0034)	(0.0034)
materials	−0.0012*	−0.0011*	0.0037	0.0038	0.0028	0.0029
	(0.0007)	(0.0007)	(0.0032)	(0.0032)	(0.0035)	(0.0035)
real_estate	0.0001	0.0003	−0.0074*	−0.0069	−0.0086*	−0.0082*
	(0.0010)	(0.0010)	(0.0043)	(0.0043)	(0.0045)	(0.0045)
utilities	−0.0012*	−0.0012*	−0.0121***	−0.0120***	−0.0142***	−0.0141***
	(0.0007)	(0.0007)	(0.0032)	(0.0032)	(0.0035)	(0.0035)
_cons	0.0005	−0.0041	0.0333*	0.0277	0.0394**	0.0353*
	(0.0041)	(0.0052)	(0.0170)	(0.0201)	(0.0179)	(0.0213)
N	50,532	50,532	50,532	50,532	50,532	50,532
F_Statistic	4.81***	4.47***	6.09***	5.81***	6.33***	5.95***
R_Squared	0.0040	0.0051	0.0046	0.0056	0.0049	0.0059
Adj_R_Squared	0.0036	0.0046	0.0042	0.0052	0.0045	0.0055

This table shows regression results for the overall sample. Standard errors are cluster-robust at the company level. All variables are defined in Table 19. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 12 Regressions: the America sample

	ar_0		CAR_1		CAR_2	
	(I)	(II)	(III)	(IV)	(V)	(VI)
H1/H2						
sentiment						
negative	−0.0092***	−0.0367**	−0.0263***	−0.1650**	−0.0264***	−0.1755***
	(0.0022)	(0.0176)	(0.0069)	(0.0683)	(0.0066)	(0.0638)
positive	0.0023***	0.0183***	0.0066**	0.0436*	0.0069**	0.0433*
	(0.0007)	(0.0062)	(0.0027)	(0.0239)	(0.0028)	(0.0245)
H3						
negative*esg		0.0072*		0.0363**		0.0390**
		(0.0043)		(0.0165)		(0.0156)
positive*esg		−0.0041***		−0.0095*		−0.0094
		(0.0015)		(0.0057)		(0.0059)
Controls						
esg	0.0016	0.0032**	−0.0028	0.0001	−0.0032	−0.0005
	(0.0010)	(0.0015)	(0.0046)	(0.0063)	(0.0047)	(0.0065)
asset	−0.0005*	−0.0005*	−0.0015	−0.0015	−0.0019*	−0.0019*
	(0.0003)	(0.0003)	(0.0010)	(0.0010)	(0.0011)	(0.0011)
num_news	0.0011*	0.0011*	0.0027	0.0024	0.0030	0.0028
	(0.0006)	(0.0006)	(0.0035)	(0.0036)	(0.0038)	(0.0039)

Table 12 (Continued)

	ar_0		CAR_1		CAR_2	
	(I)	(II)	(III)	(IV)	(V)	(VI)
sector						
communication_services	0.0036	0.0036	−0.0007	−0.0004	−0.0001	0.0003
	(0.0026)	(0.0026)	(0.0070)	(0.0070)	(0.0077)	(0.0076)
consumer_discretionary	−0.0014	−0.0014	0.0098	0.0096	0.0106	0.0105
	(0.0012)	(0.0012)	(0.0092)	(0.0094)	(0.0091)	(0.0093)
consumer_staples	0.0005	0.0006	−0.0028	−0.0025	−0.0033	−0.0031
	(0.0012)	(0.0012)	(0.0047)	(0.0047)	(0.0051)	(0.0051)
energy	0.0029	0.0030	0.0028	0.0031	0.0037	0.0039
	(0.0019)	(0.0019)	(0.0083)	(0.0084)	(0.0087)	(0.0088)
financials	0.0005	0.0005	−0.0053	−0.0054	−0.0048	−0.0049
	(0.0011)	(0.0011)	(0.0048)	(0.0048)	(0.0050)	(0.0050)
health_care	−0.0043***	−0.0041***	−0.0035	−0.0029	−0.0043	−0.0037
	(0.0016)	(0.0016)	(0.0056)	(0.0056)	(0.0058)	(0.0058)
information_technology	−0.0023*	−0.0024**	−0.0193***	−0.0196***	−0.0208***	−0.0211***
	(0.0012)	(0.0012)	(0.0046)	(0.0046)	(0.0047)	(0.0047)
materials	−0.0007	−0.0006	0.0014	0.0016	−0.0004	−0.0002
	(0.0012)	(0.0012)	(0.0055)	(0.0055)	(0.0061)	(0.0061)

Table 12 (Continued)

	ar_0		CAR_1		CAR_2	
	(I)	(II)	(III)	(IV)	(V)	(VI)
real_estate	0.0003	0.0005	−0.0128**	−0.0120**	−0.0154**	−0.0145**
	(0.0014)	(0.0014)	(0.0060)	(0.0060)	(0.0063)	(0.0063)
utilities	−0.0014	−0.0014	−0.0181***	−0.0180***	−0.0210***	−0.0209***
	(0.0012)	(0.0012)	(0.0051)	(0.0051)	(0.0055)	(0.0055)
_cons	0.0043	−0.0015	0.0486**	0.0394	0.0596***	0.0512*
	(0.0055)	(0.0071)	(0.0204)	(0.0253)	(0.0218)	(0.0270)
N	27,613	27,613	27,613	27,613	27,613	27,613
F_Statistic	3.89***	3.65***	6.72***	6.40***	6.88***	6.50***
R_Squared	0.0046	0.0057	0.0064	0.0076	0.0069	0.0082
Adj_R_Squared	0.0041	0.0050	0.0058	0.0070	0.0064	0.0076

This table shows regression results for the America sample. Standard errors are cluster-robust at the company level. All variables are defined in Table 19. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 13 Regressions: the Europe sample

	ar_0		CAR_1		CAR_2	
	(I)	(II)	(III)	(IV)	(V)	(VI)
H1/H2						
sentiment						
negative	−0.0081**	−0.0708*	−0.0175	−0.2840*	−0.0157	−0.3068*
	(0.0031)	(0.0401)	(0.0123)	(0.1651)	(0.0127)	(0.1717)
positive	−0.0003	0.0144	0.0055***	0.0149	0.0063***	0.0120
	(0.0005)	(0.0091)	(0.0020)	(0.0237)	(0.0021)	(0.0269)
H3						
negative*esg		0.0152*		0.0645*		0.0704*
		(0.0092)		(0.0378)		(0.0393)
positive*esg		−0.0035		−0.0022		−0.0013
		(0.0021)		(0.0056)		(0.0063)
Controls						
esg	0.0028**	0.0038*	0.0031	0.0013	0.0048	0.0023
	(0.0014)	(0.0020)	(0.0036)	(0.0047)	(0.0039)	(0.0054)
asset	−0.0004	−0.0005*	−0.0006	−0.0008	−0.0005	−0.0007
	(0.0003)	(0.0003)	(0.0016)	(0.0015)	(0.0016)	(0.0016)
num_news	−0.0011*	−0.0009	−0.0059	−0.0055	−0.0060	−0.0057
	(0.0006)	(0.0006)	(0.0042)	(0.0043)	(0.0046)	(0.0046)

Table 13 (Continued)

	ar_0		CAR_1		CAR_2	
	(I)	(II)	(III)	(IV)	(V)	(VI)
sector						
communication_services	−0.0031	−0.0030	−0.0123	−0.0119	−0.0145	−0.0141
	(0.0030)	(0.0029)	(0.0118)	(0.0117)	(0.0119)	(0.0117)
consumer_discretionary	0.0004	0.0007	−0.0045	−0.0033	−0.0024	−0.0010
	(0.0016)	(0.0015)	(0.0060)	(0.0055)	(0.0064)	(0.0058)
consumer_staples	0.0005	0.0006	−0.0045	−0.0045	−0.0054	−0.0054
	(0.0012)	(0.0012)	(0.0048)	(0.0048)	(0.0049)	(0.0049)
energy	0.0003	0.0004	0.0014	0.0015	−0.0001	−0.0000
	(0.0012)	(0.0012)	(0.0062)	(0.0062)	(0.0067)	(0.0067)
financials	−0.0002	0.0000	0.0021	0.0027	0.0011	0.0017
	(0.0015)	(0.0014)	(0.0077)	(0.0074)	(0.0080)	(0.0077)
health_care	−0.0046***	−0.0050***	−0.0134*	−0.0149*	−0.0146**	−0.0162**
	(0.0017)	(0.0018)	(0.0071)	(0.0078)	(0.0073)	(0.0080)
information_technology	−0.0006	−0.0008	−0.0056	−0.0061	−0.0048	−0.0053
	(0.0012)	(0.0012)	(0.0064)	(0.0063)	(0.0070)	(0.0068)
materials	−0.0008	−0.0008	0.0074*	0.0075*	0.0069	0.0069
	(0.0008)	(0.0008)	(0.0038)	(0.0038)	(0.0042)	(0.0042)

Table 13 (Continued)

	ar_0		CAR_1		CAR_2	
	(I)	(II)	(III)	(IV)	(V)	(VI)
real_estate	0.0003	0.0003	0.0025	0.0022	0.0050	0.0046
	(0.0019)	(0.0020)	(0.0089)	(0.0090)	(0.0097)	(0.0098)
utilities	−0.0015**	−0.0015*	−0.0066	−0.0064	−0.0073	−0.0070
	(0.0008)	(0.0008)	(0.0043)	(0.0043)	(0.0046)	(0.0046)
_cons	−0.0004	−0.0025	0.0093	0.0209	−0.0002	0.0142
	(0.0071)	(0.0084)	(0.0343)	(0.0309)	(0.0366)	(0.0339)
N	14,457	14,457	14,457	14,457	14,457	14,457
F_Statistic	2.95***	2.73***	2.40***	2.20***	2.45***	2.29***
R_Squared	0.0050	0.0073	0.0056	0.0082	0.0053	0.0081
Adj_R_Squared	0.0039	0.0062	0.0046	0.0070	0.0043	0.0069

This table shows regression results for the Europe sample. Standard errors are cluster-robust at the company level. All variables are defined in Table 19. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

performance when bad ESG news is released, a good historical ESG image may help relieve the problem. We also observe the significantly negative coefficient of *positive*esg* in models II and IV when the one-day performance ar_0 and three-day performance CAR_1 are taken as the dependent variables. It could be possible that when positive ESG news is released for a company with a bad ESG record, investors react more favorably since the company performs marginally better in ESG issues. Overall, our regression results suggest that stock performance related to ESG news depends not only on the news sentiment, but also on the historical ESG record. Therefore, H3 is also supported by our empirical results.

Similarly, we run the same regression routine for the America subsample. The regression results are reported in Table 12. Just like in the overall sample, we find that negative ESG news tends to have a significantly negative influence on stock performance, regardless of different model setups. We also find that the coefficients of *positive* are positive and significant, which indicates that investors react positively to positive ESG news. Again, negative ESG news appears to be taken more seriously than positive ESG news as the scale of the coefficients is larger. Moreover, we observe that the interaction term *negative*esg* is positive and significant in different models. The coefficient of *positive*esg* is also significant in models I and IV. These are indications that the historical ESG record may have an influence on investors' perception of ESG news.

We also conduct similar regressions for the Europe subsample and report results in Table 13. Even though the Europe sample presents a less clear picture, the overall patterns still hold. *negative* is significantly negative in most models except for model III and V. *positive* is significantly positive in model III and V. Moreover, *negative*esg* is significantly positive in all models at the 10% level. This provides further support to our previous findings in the overall sample and the America subsample. No matter in America or Europe, good historical reputation could be an asset when a company suffers bad ESG news coverage, while a liability when it encounters good ones.

8 Conclusion

In this study, we examine the pricing mechanism of ESG news on the major stock markets. We show how the newest development in NLP can be applied in understanding the market reactions to instant ESG news. Instead of directly adopting a proprietary ESG news dataset from ESG data providers, we construct our sample by extracting raw ESG news from Thomson Reuters Eikon and clean the news data in a consistent way. Based on a pre-trained sentence-BERT model, we are able to remove fuzzy duplicate or stale news and retain only fresh and unique ESG news to a large extent. This procedure makes sure that we have a unique and fresh news dataset while enjoying the wide coverage of ESG news from all over the world. Moreover, we fine-tune an ESG news sentiment classifier based on the BERT-like language model and achieve relatively good predictive performance. We apply it to judge the sentiment of ESG news instead of using classical lexicon-based sentiment analysis methods.

We find that the impact of ESG news is closely related to the ESG news sentiment. However, the market reactions to positive and negative ESG news are asymmetric. Positive ESG news has positive influence on the stock price while negative ESG news has stronger and negative influence on stock performance. This indicates that positive ESG news may add some value to the firm while negative ESG news do harm to the firm value to a considerable extent. Moreover, the historical ESG image of a company may influence the impact of ESG news on stock markets. More specifically, the market reaction to negative ESG news is related to the ESG record of the company. If the company had a good ESG record in the past, the negative influence of negative ESG news could be dampened and less severe. In contrast, if the company had a bad ESG record, the market reacts more favorably to marginal improvement of ESG performance.

This study has clear research implications for other financial studies. We show how the recent development in NLP could possibly facilitate and advance the research on ESG topics in different ways. The proposed text processing methodologies can also be applied in related studies, especially those investigating the role of non-financial factors. We focus specifically on the possible pricing effect of instant ESG news and provide new insight on how the market reacts to instant ESG news on the major stock markets. The empirical findings suggest the importance of ESG issues on the financial markets. Investors may incorporate daily ESG information into their investment decisions, instead of merely depending on company ESG disclosure and ESG ratings from agencies. Therefore, more attention should be given to more frequent ESG information such as instant ESG news in order to better understand the role of ESG issues.

The practical implications of this study are obvious and straightforward. Firstly, we show the importance of timely tracking of instant ESG news for investors. Investors can monitor real-time ESG news and incorporate this information in a timely manner into investment practice. Secondly, companies should not only avoid negative ESG news, but also work on improving their ESG performance since positive ESG news is also valued by investors. Moreover, companies should build up their own media monitoring system as part of their investor-relationship management, so as to build a better ESG image and avoid any misunderstandings with investors and the general public. At last, for related policy makers, our study indicates the possibility of ESG performance fraud or exaggeration in ESG news. Regulations or policies that can detect or increase the cost of such behavior should be considered and implemented. One possible countermeasure is the establishment of a third-party reviewing system in which independent external reviewers validate and evaluate ESG news on a regular basis. Moreover, the advancement in NLP could also be applied to alleviate the problem. By constructing an ESG news dataset with a label indicating the authenticity of the news, a classifier can be be trained to detect fraud or exaggeration.

We are also aware of the limitations of this study and thus provide some future research directions. First, despite the relatively good sentiment classification result, our sentiment classifier is trained on a labelled dataset pre-processed by a third party and thus its validity is restrained by the given training dataset. A better (but more costly and complicated) solution would be constructing a sentiment-labelled

dataset by designing an experiment in which participants are asked to read company ESG news and evaluate the news sentiment. By controlling the ESG news to be analyzed in this way, it may become possible to identify those positive ESG news items that elicit similarly strong reactions as the negative news (and which are masked by the many irrelevant positive ESG news items in the present study). Second, we do not differentiate various types of ESG news in this study and may not know whether investors may perceive them differently. For example, whether sustainability issues are financially material has a significant impact on the firm value (Khan et al. 2016). It would be interesting to integrate the financial materiality aspect into the pricing implication analysis of ESG news. Finally, the ESG appetite of institutional and individual investors, or investors from different countries may be different. Consequently, different groups of investors may react differently to ESG news. Therefore, to understand their behavior more comprehensively, more research effort is needed.

9 Appendix

9.1 Calculation of (cumulative) abnormal returns

For each day u in the event window, we calculate daily log-returns for ESG news item i as

$$r_{i,u} = \ln p_{i,u} - \ln p_{i,u-1}\,. \tag{4}$$

in which u is the event window days relative to the event day T_0. Next, we calculate abnormal returns for ESG news item i by estimating the market model as

$$r_{i,t} = \alpha + \beta R_{i,t} \tag{5}$$

in which $R_{i,t}$ is the daily return of the corresponding stock index[17]. We adopt an estimation period of 200 trading days which has a distance of 50 trading days to the event date. Accordingly, daily abnormal return for each ESG news event can be calculated as

$$ar_{i,u} = r_{i,u} - \widehat{\alpha} - \widehat{\beta} R_{i,u} \tag{6}$$

in which $\widehat{\alpha}$ and $\widehat{\beta}$ are estimated coefficients from the market model in Eq. (5). Moreover, cumulative abnormal returns are defined by

$$CAR_{i,\tau} = \sum_{u=T_0-\tau}^{T_0+\tau} ar_{i,u} \tag{7}$$

in which $2\tau + 1$ is the length of the ESG news event window.

[17] We choose the major stock index for each country.

9.2 Tests of significance

The test of the statistic significance of stock performance in event studies is often based on the following t-statistic:

$$t_\tau = \frac{\overline{CAR_\tau}}{\sqrt{\text{var}[\overline{CAR_\tau}]}} \tag{8}$$

in which $\overline{CAR_\tau}$ is the average of the cumulative abnormal returns across the same type of events. However, var$[\overline{CAR_\tau}]$ should be estimated with caution. Kolari and Pynnönen (2010) find that cross-sectional correlations among abnormal returns in the case of event-date clustering with the same event window may lead to biased standard tests and therefore should be considered when designing the t-statistic. In our case, we have ESG news events across many stocks and over a more than 1.5-year timeframe. Some ESG news items concern the same company and event windows may partly overlap with each other. Therefore, the corresponding cumulative abnormal returns may be subject to correlation. To address this concern, we adopt the cross-sectional and time serial correlation robust var$[\overline{CAR_\tau}]$ proposed by Kolari et al. (2018). Kolari et al. (2018) consider both cross-sectional and time serial correlation when estimating var$[\overline{CAR_\tau}]$ by grouping abnormal returns in both cross-sectional and time dimensions:

$$\text{var}[\overline{CAR_\tau}] = \frac{1}{n^2}\left(\sum_{i=1}^{n}\text{var}[CAR_{i\tau}] + \sum_{t=1}^{T}\text{var}[AR_t] - \sum_{i=1}^{n}\sum_{u=\tau_1}^{\tau_2}\text{var}[ar_{iu}]\right) \tag{9}$$

in which n is the number of events, T is the number of calendar days covered by any ESG news event for the whole sample, and AR_t is the aggregated abnormal returns on the calendar day t. The first term $\frac{1}{n^2}\sum_{i=1}^{n}\text{var}[CAR_{i\tau}]$ itself equals var$[\overline{CAR_\tau}]$ under the assumption that events are independent and can be consistently estimated by

$$\widehat{\text{var}}_a[\overline{CAR_\tau}] = \frac{1}{n^2}\sum_{i=1}^{n}(CAR_{i\tau} - \overline{CAR_\tau})^2\,. \tag{10}$$

The second term $\frac{1}{n^2}\sum_{t=1}^{T}\text{var}[AR_t]$ itself also equals var$[\overline{CAR_\tau}]$ under the assumption of serial independence and can be consistently estimated by

$$\widehat{\text{var}}_b[\overline{CAR_\tau}] = \frac{1}{n^2}\sum_{t=1}^{T}(AR_t - \overline{AR})^2 \tag{11}$$

in which $AR_t = \sum_{ar_{iu} \in D_t} ar_{iu}$ (D_t denotes the set of all ar_{iu} on the same calendar day t). The sum of the first and second term embeds both serial correlation and cross-sectional correlation terms and thus is serial and cross-section correlation robust. However, it double counts the individual variances var$[ar_{iu}]$. Therefore, we subtract the third term from the sum of the first and second term to achieve the robust var$[\overline{CAR_\tau}]$. The third term $\frac{1}{n^2} \sum_{i=1}^{n} \sum_{u=\tau_1}^{\tau_2}$ var$[ar_{iu}]$ is estimated by

$$\widehat{\text{var}_{ar}}[\overline{CAR_\tau}] = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{u=\tau_1}^{\tau_2} (ar_{iu} - \overline{ar})^2 . \tag{12}$$

Moreover, besides the significance test for the same group of ESG news, we also test whether the mean difference of stock performance between the positive group and the negative group is statistically significant. Accordingly, we adopt the following t-statistic

$$t_d = \frac{\overline{CAR_{\text{pos}}} - \overline{CAR_{\text{neg}}}}{\sqrt{\text{var}[\overline{CAR_{\text{pos}}}] + \text{var}[\overline{CAR_{\text{neg}}}]}} \tag{13}$$

in which both var$[\overline{CAR_{\text{pos}}}]$ and var$[\overline{CAR_{\text{neg}}}]$ are estimated as described in Eq. (9).

9.3 Model performance comparison

Besides the BERT model, we apply other NLP models to test whether the BERT model is the most suitable for the three language tasks, i.e., company news identification, ESG news identification and sentiment classification. We choose other three representative NLP models, i.e., Lexicon-based model, N-Gram model and Word2Vec model. We adopt accuracy and F-score as performance metrics and implement 5-Fold cross validation whenever possible. In 5-Fold cross validation, the original dataset is split into 5 equal-sized subsets. Each of these subsets is used as the evaluation dataset once, while the rest of subsets are treated as the training dataset. For comparison reasons, we set the maximum text length for the company news identification task as 128, and for the ESG news identification and sentiment classification tasks as 512. Model descriptions and other detailed model configurations are specified for each model as follows.

Lexicon-based model. For the sentiment classification task, we can alternatively choose a lexicon-based model. We identify positive and negative keywords in a news item according to a list of predefined positive and negative keywords.[18] When a positive (negative) word is identified, 1 (−1) will be added to the initial sentiment score which is set as 0 at the beginning. If the final sentiment score is no less than 5, news is classified as positive news. If the sentiment score is no more than −5, news is classified as negative news. The rest is classified as neutral news.

[18] We adopt a popular opinion lexicon maintained by Bind Liu at University of Illinois Chicago. See: https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html.

Table 14 Model performance comparison: accuracy rate

	Company news identification	ESG news identification	ESG news sentiment
Lexicon-based	–	–	57%
Unigram	96%	88%	63%
Bigram	96%	89%	62%
Word2Vec	70%	61%	58%
BERT	99%	99%	81%

Except for the lexicon-based model, the reported accuracy rates are mean accuracy rates in 5-Fold cross validation

Table 15 Model performance comparison: F1 score

	Company news identification	ESG news identification	ESG news sentiment
Lexicon-based	–	–	38%, 59%, 61%
Unigram	96%, 96%	88%, 89%	0%, 76%, 25%
Bigram	96%, 96%	88%, 89%	1%, 76%, 18%
Word2Vec	70%, 70%	68%, 45%	2%, 73%, 38%
BERT	99%, 99%	99%, 99%	66%, 84%, 79%

Except for the lexicon-based model, the reported F1 scores are mean F1 scores in 5-Fold cross validation. In each table cell, two or three F1 scores are reported according to the number of labels in a language task. F1 scores are bundled with specific labels and the sequence of F1 scores are reported as follows. For the ESG news sentiment classification task, the first F1 score is for negative ESG news, the second is for neutral ESG news, and the third is for positive ESG news. For the other two news category classification tasks, the first F1 score is for non-company (non-ESG) news, and the second is for company (ESG) news

N-Gram model. In the first step, a N-Gram (a unigram or a bigram in this study) model (see Jurafsky and Martin 2000) is applied to quantify unstructured text data. Additionally, we use term frequency-inverse document frequency (TF-IDF) features to enhance information retrieval from texts (see Ramos 2003). In the second step, we adopt a multinomial Naive-Bayes classifier to classify text inputs into different categories or sentiment.[19]

Word2Vec model. A Word2Vec model is a language model in which words are embedded as vectors of numbers (see Mikolov et al. 2013). We implement the Word2Vec model as a skip-gram model, in which the window size for searching for skip-gram samples is set as 3.[20] Given meaningful word embeddings derived from the Word2Vec model, we can further add two simple linear neural network layers, of which the first and second consist of 64 and 32 neurons respectively, before we prepare a Softmax output layer.

The model performance is summarized in Table 14 and Table 15. In all three language tasks, the BERT model always performs the best, no matter which performance metric is adopted (accuracy rate or F1 score). For the two news category identification tasks, the accuracy rates of the BERT model are 99%, which are much better than that of the Word2Vec model and slightly improvement compared with that of the N-Gram models. For the ESG news sentiment classification task, the BERT model clearly outperforms the other models with an accuracy rate of 81%.

[19] We use a Python machine learning library called Scikit-learn to implement the above steps.

[20] We use Gensim, a Python library for NLP, to train a Word2Vec model.

Table 16 Human audit on BERT sentiment classification: accuracy

	BERT	Student 1	Student 2	Student 3
BERT	100%	75%	81%	77%
Student 1	75%	100%	89%	81%
Student 2	81%	89%	100%	81%
Student 3	77%	81%	81%	100%

This table compares the sentiment classification results of three students and the BERT model. Given one classification result as the base, the accuracy rates of the rest results with respect to the base are calculated

Table 17 Human audit on BERT sentiment classification: F1 score

	BERT	Student 1	Student 2	Student 3
BERT	100%, 100%, 100%	95%, 65%, 65%	94%, 72%, 77%	93%, 65%, 73%
Student 1	95%, 65%, 65%	100%, 100%, 100%	96%, 85%, 87%	93%, 73%, 78%
Student 2	94%, 72%, 77%	96%, 85%, 87%	100%, 100%, 100%	88%, 72%, 82%
Student 3	93%, 65%, 73%	92%, 73%, 78%	89%, 72%, 82%	100%, 100%, 100%

This table compares the sentiment classification results of three students and the BERT model. Given one classification result as the base, the F1 scores of the rest results with respect to the base are calculated. The first score is for negative ESG news, the second is for neutral ESG news and the third is for positive ESG news

However, accuracy rate may be sometimes misleading as it could be possible that the model with the highest accuracy rate may not be the best model due to the imbalanced training data. To further check model performance, we choose F1 score (see Van Rijsbergen 1979), which takes the imbalance in the training data into consideration, as an additional performance metric. All F1 scores of all language tasks confirm that the BERT model is clearly superior to the other NLP models. In particular, the other four models perform badly in the ESG news sentiment classification task. The F1 scores of these four models indicate that they cannot really identify positive and negative ESG news, while the BERT model delivers much better performance.

9.4 Human audit on BERT sentiment classification

To check the validity of the BERT classification, we had three university students read a subsample of 120 news items, in which there were 40 news items per sentiment category according to the BERT model. They had to classify them into positive, neutral and negative ESG news, without knowing the sentiment classification result of the BERT model. Table 16 shows that the BERT model achieves an average accuracy rate[21] of 78%, which is close to the accuracy rate of 81% in Table 3.

Moreover, Table 17 suggests that BERT and the students have a high level of agreement in classifying negative ESG news (with an F1 score over 90%). Also note that there is clear evidence in Table 16 and Table 17 that there may be individual differences in sentiment judgement. This partly explains why the BERT model still works well, although in some cases it deviates from human judgement.

[21] When measuring the accuracy rate, we take one of the sentiment classification result of students or the BERT model as the baseline.

9.5 Additional tables

Tables 18 and 19 provide additional information on the top countries in our sample and on the definition of our variables.

Table 18 ESG news (in English) volume: Top 20 countries/regions

No.	Country/Region	ESG news	No.	Country/Region	ESG news
1.	USA	35,284	11.	Switzerland	1243
2.	Canada	9710	12.	Italy	1166
3.	UK	7319	13.	Finland	1095
4.	Japan	3372	14.	Korea	1061
5.	France	3307	15.	Spain	1057
6.	India	3090	16.	China	985
7.	Australia	2087	17.	Netherlands	888
8.	Germany	1883	18.	Norway	821
9.	Sweden	1436	19.	Thailand	815
10.	Hongkong	1325	20.	Russia	642

Table 19 Definition of variables

Variable	Description
H1	
ar_0	Abnormal return on the event date based on the estiamted market model.
CAR_1	Cumulative return during the $(T_0 - 1, T_0 + 1)$ event window.
CAR_2	Cumulative abnormal return during the $(T_0 - 2, T_0 + 2)$ event window.
CAR_5	Cumulative return during the $(T_0 - 5, T_0 + 5)$ event window.
CAR_{10}	Cumulative abnormal return during the $(T_0 - 5, T_0 + 5)$ event window.
H2	
sentiment	Categorical variable indicating ESG news sentiment judged by the BERT model, indicating whether the news sentiment is "positive", "neutral", or "negative". The reference category is "neutral".
H3	
*negative*esg*	Interaction term between *negative* and *esg*.
*positive*esg*	Interaction term between *positive* and *esg*.
Controls	
esg	ESG score for the corresponding company in the previous year of the event day, logarithmized in regressions
asset	Total asset of the company, logarithmized in regressions.
num_news	Number of ESG news items for the company during the sample period, divided by 100 in regressions.
continent	Continent where the company is located. The reference category is "America".
sector	Sector to which the company belongs. The Global Industry Classifiction Standards (GICS) is adopted. The reference category is "industrials".

Declarations

Conflict of interest Gregor Dorfleitner declares that he has no competing interests. Rongxin Zhang declares that he has no competing interests.

References

Akhtar, S., R. Faff, B. Oliver, and A. Subrahmanyam. 2011. The power of bad: The negativity bias in Australian consumer sentiment announcements on stock returns. *Journal of Banking & Finance* 35:1239–1249.

Alanyali, M., H.S. Moat, and T. Preis. 2013. Quantifying the relationship between financial news and the stock market. *Scientific Reports* 3:1–6.

Alaparthi, S., and M. Mishra. 2021. Bert: a sentiment analysis odyssey. *Journal of Marketing Analytics* 9:118–126.

Amel-Zadeh, A., and G. Serafeim. 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:1908.10063.

Aue, T., A. Jatowt, and M. Färber. 2022. *Predicting companies' esg ratings from news articles using multivariate timeseries analysis*. arXiv:2212.11765.

Bartov, E., A. Marra, and F. Momente. 2021. Corporate social responsibility and the market reaction to negative events: evidence from inadvertent and fraudulent restatement announcements. *The Accounting Review* 96:81–106.

Bazillier, R., and J. Vauday. 2009. *The greenwashing machine: Is CSR more than communication*. HAL, hal-00448861.

Bennani, L., T. Le Guenedal, F. Lepetit, L. Ly, V. Mortier, T. Roncalli, and T. Sekine. 2018. *How ESG Investing has impacted the asset pricing in the equity market*. SSRN.

Berg, F., J.F. Koelbel, and R. Rigobon. 2022. Aggregate confusion: the divergence of ESG ratings. *Review of Finance* 26:1315–1344. https://doi.org/10.1093/rof/rfac033.

Bingler, J.A., M. Kraus, and M. Leippold. 2021. *Cheap talk and cherry-picking: What ClimateBert has to say on corporate climate risk disclosures*. SSRN.

Boudoukh, J., R. Feldman, S. Kogan, and M. Richardson. 2019. Information, trading, and volatility: evidence from firm-specific news. *The Review of Financial Studies* 32:992–1033.

Capelle-Blancard, G., and A. Petit. 2019. Every little helps? ESG news and stock market reaction. *Journal of Business Ethics* 157:543–565.

Chava, S., W. Du, and B. Malakar. 2021. *Do managers walk the talk on environmental and social issues?* Georgia Tech Scheller College of Business Research Paper.

Cui, B., and P. Docherty. 2020. *Stock price overreaction to ESG controversies*. SSRN.

Da Silva, N.F., E.R. Hruschka, and E.R. Hruschka Jr. 2014. Tweet sentiment analysis with classifier ensembles. *Decision Support Systems* 66:170–179.

Del Giudice, A., and S. Rigamonti. 2020. Does audit improve the quality of ESG scores? Evidence from corporate misconduct. *Sustainability* 12:5670.

Derrien, F., P. Krueger, A. Landier, and T. Yao. 2021. *ESG news, future cash flows, and firm value*. Swiss Finance Institute Research Paper.

Devlin, J., M.W. Chang, K. Lee, and K. Toutanova. 2018. *BERT: Pre-training of deep bidirectional transformers for language understanding*. arXiv:1810.04805.

Dorfleitner, G., G. Halbritter, and M. Nguyen. 2015. Measuring the level and risk of corporate responsibility–an empirical comparison of different ESG rating approaches. *Journal of Asset Management* 16:450–466.
Dorfleitner, G., C. Priberny, S. Schuster, J. Stoiber, M. Weber, I. de Castro, and J. Kammler. 2016. Description-text related soft information in peer-to-peer lending–evidence from two leading european platforms. *Journal of Banking & Finance* 64:169–187.
Drempetic, S., C. Klein, and B. Zwergel. 2020. The influence of firm size on the ESG score: corporate sustainability ratings under review. *Journal of Business Ethics* 167:333–360. https://doi.org/10.1007/s10551-019-04164-1.
Edmans, A., D. Garcia, and Ø. Norli. 2007. Sports sentiment and stock returns. *The Journal of Finance* 62:1967–1998.
Escrig-Olmedo, E., M.Á. Fernández-Izquierdo, I. Ferrero-Ferrero, J.M. Rivera-Lirio, and M.J. Muñoz-Torres. 2019. Rating the raters: Evaluating how ESG rating agencies integrate sustainability principles. *Sustainability* 11:915.
Fiaschi, D., E. Giuliani, F. Nieri, and N. Salvati. 2020. How bad is your company? Measuring corporate wrongdoing beyond the magic of ESG metrics. *Business Horizons* 63:287–299.
Fischbach, J., M. Adam, V. Dzhagatspanyan, D. Mendez, J. Frattini, O. Kosenkov, and P. Elahidoost. 2022. *Automatic eSG assessment of companies by mining and evaluating media coverage data: NLP approach and tool*. arXiv:2212.06540.
Flammer, C. 2013. Corporate social responsibility and shareholder reaction: the environmental awareness of investors. *Academy of Management Journal* 56:758–781.
Flammer, C. 2021. Corporate green bonds. *Journal of Financial Economics* 142:499–516.
Friede, G., T. Busch, and A. Bassen. 2015. ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment* 5:210–233.
Gantchev, N., M. Giannetti, and R. Li. 2022. Does money talk? divestitures and corporate environmental and social policies. *Review of Finance* 26:1469–1508.
Glossner, S. 2021. *Repeat offenders: ESG incident recidivism and investor underreaction*. SSRN.
Glück, M., B. Hübel, and H. Scholz. 2021. *ESG rating events and stock market reactions*. SSRN.
Goss, A., and G.S. Roberts. 2011. The impact of corporate social responsibility on the cost of bank loans. *Journal of Banking & Finance* 35:1794–1810.
Grewal, J., C. Hauptmann, and G. Serafeim. 2021. Material sustainability information and stock price informativeness. *Journal of Business Ethics* 171:513–544. https://doi.org/10.1007/s10551-020-04451-2.
Hartzmark, S.M., and A.B. Sussman. 2019. Do investors value sustainability? A natural experiment examining ranking and fund flows. *The Journal of Finance* 74:2789–2837.
Heston, S.L., and N.R. Sinha. 2017. News vs. sentiment: Predicting stock returns from news stories. *Financial Analysts Journal* 73:67–83.
Jahdi, K.S., and G. Acikdilli. 2009. Marketing communications and corporate social responsibility (CSR): Marriage of convenience or shotgun wedding? *Journal of Business Ethics* 88:103–113.
Jurafsky, D., and J. Martin. 2000. *Speech & language processing. An introduction to NL processing, computational linguistics & speech recognition*. New Jersey: Prentice Hall.
Ke, Z.T., B.T. Kelly, and D. Xiu. 2019. *Predicting returns with text data*. Technical Report. National Bureau of Economic Research.
Kearney, C., and S. Liu. 2014. Textual sentiment in finance: a survey of methods and models. *International Review of Financial Analysis* 33:171–185.
Khan, M., G. Serafeim, and A. Yoon. 2016. Corporate sustainability: first evidence on materiality. *The Accounting Review* 91:1697–1724.
Kim, E.H., and T.P. Lyon. 2015. Greenwash vs. brownwash: exaggeration and undue modesty in corporate sustainability disclosure. *Organization Science* 26:705–723.
Kolari, J.W., and S. Pynnönen. 2010. Event study testing with cross-sectional correlation of abnormal returns. *The Review of Financial Studies* 23:3996–4025.
Kolari, J.W., B. Pape, and S. Pynnonen. 2018. *Event study testing with cross-sectional correlation due to partially overlapping event windows*. Mays Business School Research Paper.
Kotelnikova, A., D. Paschenko, and E. Razova. 2021. *Lexicon-based methods and BERT model for sentiment analysis of russian text corpora*. CEUR Workshop Proceedings.
KPMG. 2019. *Impact of ESG disclosures*. Technical Report. KPMG.
KPMG. 2020. *The KPMG survey of sustainability reporting 2020*. Technical Report. KPMG.
Krüger, P. 2015. Corporate goodness and shareholder wealth. *Journal of Financial Economics* 115:304–329.

Lan, Z., M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut. 2019. *ALBERT: A lite BERT for self-supervised learning of language representations*. arXiv:1909.11942.
Li, X., H. Xie, L. Chen, J. Wang, and X. Deng. 2014. News impact on stock price return via sentiment analysis. *Knowledge-Based Systems* 69:14–23.
Lins, K.V., H. Servaes, and A. Tamayo. 2017. Social capital, trust, and firm performance: the value of corporate social responsibility during the financial crisis. *The Journal of Finance* 72:1785–1824.
Liu, Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. 2019. *RoBERTa: A robustly optimized BERT pretraining approach*. arXiv:1907.11692.
Lopez, C., O. Contreras, and J. Bendix. 2020. *Disagreement among ESG rating agencies: Shall we be worried?* MPRA paper.
Mănescu, C. 2011. Stock returns in relation to environmental, social and governance performance: mispricing or compensation for risk? *Sustainable Development* 19:95–118.
Maniora, J. 2017. Is integrated reporting really the superior mechanism for the integration of ethics into the core business model? An empirical analysis. *Journal of Business Ethics* 140:755–786.
Mehra, S., R. Louka, and Y. Zhang. 2022. *Esgbert: Language model to help with classification tasks related to companies environmental, social, and governance practices*. arXiv:2203.16788.
Mikolov, T., K. Chen, G. Corrado, and J. Dean. 2013. *Efficient estimation of word representations in vector space*. arXiv:1301.3781.
Naughton, J.P., C. Wang, and I. Yeung. 2019. Investor sentiment for corporate social performance. *The Accounting Review* 94:401–420.
Naumer, H.J., and B. Yurtoglu. 2022. It is not only what you say, but how you say it: ESG, corporate news, and the impact on CDS spreads. *Global Finance Journal* 52, 100571. https://doi.org/10.1016/j.gfj.2020.100571.
Pedersen, L.H., S. Fitzgibbons, and L. Pomorski. 2021. Responsible investing: the ESG-efficient frontier. *Journal of Financial Economics* 142:572–597. https://doi.org/10.1016/j.jfineco.2020.11.001.
Peress, J. 2014. The media and the diffusion of information in financial markets: evidence from newspaper strikes. *The Journal of Finance* 69:2007–2043.
Ramos, J. 2003. Using TF-IDF to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, Vol. 242, 29–48.
Reimers, N., and I. Gurevych. 2019. *Sentence-BERT: Sentence embeddings using siamese BERT-networks*. arXiv:1908.10084.
Rozin, P., and E.B. Royzman. 2001. Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review* 5:296–320.
Serafeim, G., and A. Yoon. 2021. *Stock price reactions to esg news: The role of ESG ratings and disagreement*. Harvard Business School Accounting & Management Unit Working Paper..
Shiu, Y.M., and S.L. Yang. 2017. Does engagement in corporate social responsibility provide strategic insurance-like effects? *Strategic Management Journal* 38:455–470.
Sokolov, A., J. Mostovoy, J. Ding, and L. Seco. 2021. Building machine learning systems for automated ESG scoring. *The Journal of Impact and ESG Investing* 1:39–50.
Taleb, W., T. Le Guenedal, F. Lepetit, V. Mortier, T. Sekine, and L. Stagnol. 2020. *Corporate ESG news and the stock market*. SSRN.
Tetlock, P.C. 2007. Giving content to investor sentiment: the role of media in the stock market. *The Journal of Finance* 62:1139–1168.
Van Rijsbergen, C. 1979. *Information retrieval: theory and practice*. Proceedings of the Joint IBM/University of Newcastle upon Tyne Seminar on Data Base Systems.
Van Duuren, E., A. Plantinga, and B. Scholtens. 2016. ESG integration and the investment management process: fundamental investing reinvented. *Journal of Business Ethics* 138:525–533.
Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, and I. Polosukhin. 2017. *Attention is all you need. Advances in neural information processing systems*
Yu, E.P., B. Van Luu, and C.H. Chen. 2020. Greenwashing in environmental, social and governance disclosures. *Research in International Business and Finance* 52:101192.