International Review of Financial Analysis 89 (2023) 102745

Contents lists available at ScienceDirect

International Review of Financial Analysis

journal homepage: www.elsevier.com/locate/irfa

When ESG talks: ESG tone of 10-K reports and its significance to stock markets

Konstantin Ignatov *

Allianz Endowed Chair of Finance, WHU - Otto Beisheim School of Management, Germany

ARTICLE INFO

Keywords:
ESG
Sustainability
Market sentiment
Textual analysis

ABSTRACT

Since the ESG topic consistently gains on importance in the investment universe, companies provide investors with information regarding recent and future ESG activities through different reporting channels. The most recent research finds relevance of ESG-related corporate activities for formation of investors' opinion regarding companies' valuations and growth prospects. Based on a sample of more than seventeen thousand unique 10-K reports of US companies filed with SEC in period 2013 to 2019 and the word-power methodology proposed by Jegadeesh and Wu (2013), this study also shows evidence for significant relation of ESG textual tone of 10-K reports to stock market returns of filing companies around the report filing dates. Using the ESG linguistic dictionary recently proposed by Baier, Berninger, and Kiesel (2020), this study shows evidence for significant relation of social and governance-related topics disclosure to stock returns, while environmental narratives being ignored by the markets. When looking at individual words from the ESG lexicon, such words as "community", "health", "control" imply positive reaction of markets, while "discrimination", "embezzlement", and "crime" are related to negative returns. The robustness analysis based on the inverse document frequency word weightings and actual ESG performance scores confirms the significance of ESG information disclosure of 10-K reports for investors. Thus, this study sheds light on the mechanics of ESG information perception and its influence on capital markets.

1. Introduction

The global stock markets are driven by many factors and events, which include macroeconomic developments, monetary policies of central banks, economic and corporate cycles, and fundamentals of companies, among others. While the technical fundamentals of a company are still the major factor that determines the movements in its market valuation on stock exchanges, the verbal, or "soft", information about companies has gained in importance for investors especially in the last decade. This trend is also seen in corporate disclosure behavior, with companies publishing more and more information about their activities to meet informational needs of capital markets and general public, which leads even to an overload of information for investors (English and Schooley (2014)). However, the majority of investors advocate for even broader disclosure of qualitative data and reports, emphasizing their materiality and value for investing decisions (Ho (2020)).[1] With advances in content analysis techniques over the past twenty years as a consequence of progress in computing power and processing of information (Loughran and McDonald (2020)), the analysis of verbal corporate disclosure and the market sentiment has become commonplace in financial markets and academic research. Textual analysis methods allow investors and researchers to extract the core information out of a large corpus of textual data, provided by companies and news agencies, that is relevant for the assessment of company's current condition and perspectives and thus for changes in expectations of its future market value mirrored in the stock price (Jose and Lee (2007)).

The growing importance of sustainable behavior in the economy paved the way for increase in the corporate reporting of ESG-related information by companies in the last decade (Du and Yu (2021)). Following the disclosure guidelines of market regulators such as Securities and Exchange Commission (SEC) and authorities as well as informational demand of market participants, the majority of companies around the world dedicate special attention to reporting of policies and activities in environmental, social, and governance areas, with investors

* Corresponding author at: Allianz Endowed Chair of Finance, WHU - Otto Beisheim School of Man- agement, Burgplatz 2, 56179 Vallendar, Germany.
E-mail address: konstantin.ignatov@whu.edu.

[1] In comparison to institutional investors, the retail investors do not account for ESG information disclosed by companies while making adjustments to their portfolio structure as shown by the latest research (Moss, Naughton, and Wang (2020)).

https://doi.org/10.1016/j.irfa.2023.102745
Received 5 October 2022; Received in revised form 23 April 2023; Accepted 27 June 2023
Available online 7 July 2023
1057-5219/

thoroughly analyzing sustainable plans and opportunities of companies and adjusting their investment strategies accordingly (Brogi and Lagasio (2019)). The recent research finds significant relation between reporting of ESG-related information and various performance characteristics of companies (Amel-Zadeh and Serafeim (2018)), including stock price informativeness (Grewal, Hauptmann, and Serafeim (2021)), capital cost (Dhaliwal, Li, Tsang, and Yang (2011)), profitability (Brogi and Lagasio (2019)), ESG performance (Lopez-de-Silanes, McCahery, and Pudschedl (2019)), and market valuation (Aureli, Gigli, Medei, and Supino (2020)). While the significance of the tone of textual information for investor sentiment and thus stock prices has been the focus of content analysis in both practice and academia for over a decade (Tetlock (2007)), the reaction of capital market participants to ESG news and information disclosure constitutes a relatively new area of research interest. For example, Landau, Rochell, Klein, and Zwergel (2020) find positive relation of inclusion of ESG information in annual reports of European companies to changes in their market values, while Serafeim and Yoon (2022) observe stock price reactions to certain types of ESG news and explain reasons for different perception of news by investors. Da Silva (2020) also emphasizes the relevance of inclusion of ESG information in annual reports for investors, confirming prior findings with respect to informativeness of ESG statements about prospective earnings of a company being dependent on the extent of ESG disclosure.

To derive key messages out of large volumes of textual information from corporate reports, analysts and scientists predominantly resort to methods of textual analysis from linguistics and machine learning fields (Tremblay, Parra, and Castellanos (2015)). The majority of approaches used in analyzing textual information can be clustered into two categories - counting methods following specific rules and methods based on statistical inference (Li (2010)). The dictionary-based approaches form classifying rules that distinguish semantic categories in a text corpus according to the words contained in thematic word lists, while the machine learning procedures use statistical rules to map the text content to specific categories based on correlation observations. While the dictionary-based analysis requires specific word lists that have to be manually prepared in advance, this type of content exploration has gained considerable credibility in the research community (Meier, Esmatyar, and Frost (2018)), especially in the financial context after introduction of the first finance-tailored linguistic dictionary by Loughran and McDonald (2011). Li (2010) also emphasizes the importance of customized word lists dependent on the type of documents analyzed and semantic topics of interest for the rules-based procedures.

While the textual analysis of ESG information is an object of interest for both academics and practitioners that seek to assess ESG efforts of a company and their consequences for its future business development (Barbeito-Caamano and Chalmeta (2020)), the lack of a standardized, comprehensive ESG dictionary has been a hurdle for application of dictionary-based approaches on large text datasets in the ESG context. The recent proposition of a broad structured ESG dictionary by Baier, Berninger, and Kiesel (2020), based on annual reports of US companies, creates an opportunity to analyze ESG tone of information provided by companies in their annual reports on a large scale using established methods of textual analysis. Besides that, the application of a market-based weighting scheme within the dictionary framework allows to evaluate investors' attitude to ESG topics and their reaction to disclosure of sustainability-related information, which is the primary focus of this paper. A combination of the word weighting design suggested by Jegadeesh & Wu, 2013 with the ESG dictionary also allows to decompose the market reaction to the ESG tone of reports by determining the most significant ESG words for investors in terms of their positive or negative impact on stock prices. The estimation procedure based on market movements ensures that the ESG sentiment is not predetermined by subjective opinion of researchers that usually assign the tone characteristics to the words from the dictionary and neglect the differences in their semantic strength (Jegadeesh and Wu (2013)). With the annual report being one of the major sources of ESG information for investors (Da Silva (2020)), the framework allows to empirically analyze the role of the ESG information disclosure of companies for investors and hence their investing decisions on a large scale.

Thus, this paper implements textual analysis techniques to examine the reaction of financial markets to the ESG sentiment of companies based on their 10-K reports and to determine the most important ESG words for investors in terms of the magnitude of changes in abnormal returns following the release of 10-K reports. The empirical results based on a sample of more than seventeen thousand unique 10-K reports of 3142 listed US firms released in years 2013 to 2019 provide evidence for a positive significant reaction of financial markets to the ESG tone of 10-K[2] reports, with investors emphasizing the social and the governance disclosure while ignoring the environmental narrative of 10-K reports. Besides the analysis of the overall ESG sentiment including separate sustainability pillars, the framework implemented in this paper estimates the impact of each individual ESG word on changes in abnormal stock returns and thus the influence on investors' expectations. The market-based regression method following Jegadeesh and Wu (2013) finds "community", "health", and "people" among the words with the most positive impact on stock returns, while "embezzlement", "crimes", and "discriminate" among terms that induce the most negative reaction of investors. To the best of our knowledge, this study is the first to analyze the impact of the ESG tone and individual words of 10-K reports on financial markets using a market reactions-based word weighting approach within the ESG context.

This paper makes contribution to the literature in the fields of textual analysis, capital market sentiment, and investors' awareness and perception with focus on environmental, social, and governance pillars of sustainability. The study sheds light on the role of ESG tone conveyed in 10-K reports of companies in changes of investors' expectations regarding their market valuation. Besides applications for the asset and investment management purposes, the framework implemented in this study could serve as a management tool for companies to optimize their communication with investors in terms of ESG information reporting, thus improving the delivery of intended meaning of messages to recipients through qualitative information disclosure. In addition, the determination of ESG words with the most negative effect on stock prices could help companies to determine the critical aspects of sustainability for investors and adjust their business strategy accordingly. Last but not least, the identification of the most relevant ESG topics could enhance the work of regulatory agencies, such as the SEC, in terms of the development of new ESG disclosure regulations for companies, as in case of upcoming climate disclosure rules which cover both the material risks and direct and indirect emissions from companies' activities.[3]

The remaining sections of the paper are organized as follows. Section II provides with an overview of related research studies on investors' reaction to ESG information along with description of the methodological framework implemented in this paper. Section III describes the dataset used in this study and empirical results including the robustness analysis. Section IV interprets and discusses the obtained results, while Section V draws conclusion and considers future research opportunities arising from this paper.

[2] We use 10-K reports instead of annual reports in our study because of more detailed information provided with 10-K reports compared to annual reports of companies.

[3] "Upcoming SEC climate disclosure rules bring urgency to ESG data strategy planning", Reuters, 30.01.2023, available at https://www.reuters.com/legal/legalindustry/upcoming-sec-climate-disclosure-rules-bring- urgency-esg-data-strategy-planning-2023-01-30

2. Related literature and methodology

2.1. Literature overview

The empirical studies provide evidence for a significant relationship between 10-K reports and the absolute value of excess return, such as in the study of Griffin (2003) that analyzed the relationship in the event window covering the filing date of a 10-K report, which showed more elevated reaction of market prices to 10-K reports than to 10-Q reports filed with SEC. In another study, Li and Sun (2022) show a significant relationship between the search for 10-K reports in the EDGAR database through investors' IPs and the return predictability of stock returns. In addition, Dyer, Lang, and Stice-Lawrence (2017) look at the evolution of 10-K reports over the period from 1996 to 2013 and find substantial increase of disclosure information in 10-K reports, which could serve as an additional factor for investors to adhere to the 10-K reports as a profound source of relevant corporate information in recent years. Following the exploration of the general tone of qualitative information disclosed by companies and its role for financial markets, the current research increasingly devotes itself to the study of ESG information relevance for stakeholders and its impact on expectations of the market value of companies. For example, Capelle-Blancard and Petit (2019) evaluate the market reaction on regular ESG news covering one hundred multinational corporations by focusing on the 2002–2010 time frame and find a negative impact of bad ESG news on companies' market valuations, with investors showing no reaction to the positive announcements. By differentiation between types of the news' sources, the study finds a pronounced preference of investors towards media outlets and negligence of explicit ESG comments provided by companies. In another study, Du and Yu (2021) focus on corporate social responsibility reports (CSR) of >260 companies included in Fortune 500 in period 2002 to 2014 and find a positive impact of the tone and the readability of reports on changes in the stock prices of listed companies. On the other side of the Atlantic, Landau et al. (2020) examine integrated reporting of ESG topics of the STOXX 50 index members from 2010 to 2016 and observe a positive effect of disclosure of high-quality reports on companies' valuations 6 months after the end of the financial year. Adding to the European evidence, Brinckmann and Ender (2019) look at the CSR online news related to the Austrian listed companies-members of the ATX index and find that the news covering environmental and social pillars of sustainability induce positive abnormal returns within a five-day window around their publication. Looking at the Asian markets, Yen, Shiu, and Wang (2019) find positive reaction of markets to good CSR news of socially responsible companies provided to the public in period 2009 to 2013 and available in the main English-based information sources. Furthermore, Hummel, Mittelbach-Hormanseder, Rammerstorfer, and Weinmayer (2020) by analyzing more than one thousand adverse ESG events and the level of sustainability-related information disclosure of 655 companies available in the REPRISK database from 2011 to 2014 find a mitigating effect of the ESG disclosure on investors' reaction in the aftermath of negative ESG accidents. Analyzing the discourse of ESG topics in conference calls on a sample of more than one hundred and forty thousand documents of US companies from 2006 to 2020, Henry, Jiang, and Rozario (2021) find a positive link between the degree of discourse and the market value of companies, with explicit positive effect of social and governance pillars within the discourse. Focusing on the type of ESG news in terms of their financial relevance for the company, Serafeim and Yoon (2022) find on a sample with more than three thousand companies in period 2010 to 2018 that the news' impact on investors and thus their reaction depend on the ESG news' relevance for fundamental metrics of companies, with social topics earning highest attention of investors in terms of the abnormal returns. Krüger (2015) observes more than two thousand CSR-related events between 2001 and 2007 of companies contained in the KLD database and also finds significant response of investors in the aftermath of them, especially to cases concerning the societal and environmental areas of sustainability. Looking at the recent economic shock induced by the COVID-19 pandemic, Cheema-Fox, LaPerla, Wang, and Serafeim (2021) discover that the positive news sentiment around social- and governance-focused actions of more than three thousand global companies evaluated by Truvalue Labs mitigate negative stock returns in crisis time and thus soften the adverse reaction of investors. With regards to the stock price informativeness and the volatility risk of stocks, Grewal et al. (2021) and Sabbaghi (2022) find positive influence of the disclosed ESG information on firm-specific volatility of returns on a sample of more than one thousand US companies in the period 2007 to 2015 and on the asymmetric stock price volatility of companies[4] between 2015 and 2019, accordingly.

Although the relevance of qualitative ESG information for investors is emphasized in many studies, there is still no consensus in the literature on the magnitude of the effect on stock prices, the significance of both positive and negative types of news for investors, the role of individual ESG pillars for markets, and the dependence of the impact on news' relevance to companies' financials (Da Silva (2020)). Furthermore, Capelle-Blancard and Petit (2019) argue that the focus of many studies lies only on selected industries and their findings do not necessarily represent the general linkage between the ESG information and the stock markets' reaction, while Serafeim and Yoon (2022) emphasize small sample sizes of conducted studies in periods when the relevance of ESG for the economy and investors was not at the forefront of decision-making. The recent developments in natural language processing (NLP) and computing power allow to reevaluate mentioned relations on a large sample using ESG-specific dictionary and a market-based term weighting scheme on structured reports that contain ESG information. Since capital markets pay particular attention to annual reports[5] of companies (Hummel et al. (2020)), the ESG dictionary of Baier et al., 2020, developed on a sample of 10-K reports, and the word-power technique of content analysis suggested by Jegadeesh & Wu, 2013 are best suited to analyze the importance of ESG information for investors. We rely on 10-K reports in our study since high quality and reliability of data contained in 10-K reports, which companies are obliged to file with SEC in contrast to the voluntary reporting, are valued by investors and have shown to have significant impact on market prices (Jegadeesh and Wu (2013)). Besides that, the integration of ESG information in 10-K reports, defined as integrated reporting, serves the investors as a reliable source of sustainability-relevant information and in combination with financial information helps the investors to assess the impact of ESG on the core business model of a company (Baier et al. (2020), Maniora (2017)). Thus, we choose the 10-K reports instead of more frequently published 10-Q reports or stand-alone ESG reports because of their quality, reliability, standardized reporting of information, as well as integration of financial information relevant to investors' decisions.

2.2. Market reactions-based word weighting approach

Besides a thematic word list, the content analysis based on a

[4] Members of the MSCI USA ESG Leaders Index and the MSCI USA Small Cap ESG Leaders Index.

[5] A 10-K report contains generally more detailed information than the annual report, although many companiesjust copy the information from their 10-K reports into annual reports (https://www.sec.gov/files/reada10k.pdf). Apart from a more detailed information, 10-K reports also have a standardized reporting format which enables a comparison of companies from different industries in our study by means of EDGAR data base scraping. With respect to the differences between the 10-K report and the annual report in terms of ESG content, the literature has examples which show that the disclosure patterns of ESG disclosure information are similar in both types of reports (Gamble, Hsu, Kite, and Radtke (1995)). Given a more detailed disclosure of information in a 10-K report compared to an annual report, we rely on 10-K reports sample in our study to get a more accurate and comprehensive picture of ESG disclosure of exchange-listed companies.

dictionary approach also implies the weighting of the terms in the overall sentiment score that reflects the tone of a document. While the standard methodology is based on a simple counting of words from the dictionary contained in a text and thus assumes an equal semantic weight of each word (Meier et al. (2018)), the research in linguistics came up with more profound weighting schemes, such as the inverse document frequency (*idf*), that take into account the frequency of the word's usage in both the individual document and the whole sample, thus lessening the weight of frequently used words with less semantic meaning in the score. However, such term weighting method ignores the differences in impact of particular words on investors and thus neglects the semantic strength of words.[6] To address these issues, Jegadeesh and Wu (2013) suggested a weighting methodology for dictionary-based approaches of content analysis that assigns term weights based on the market reaction evoked by disclosure of information. The objective reaction of investors allows to determine both the degree and the direction of the impact of each word without prior assumptions, thus considering market perception of information in the quantitative tone score.

Since the sentiment score has to reflect these factors, Jegadeesh and Wu (2013) define three main features that serve as its fundament, namely a positive relationship between the frequency of dictionary words used in the text and the value of the score, a positive relationship between the weight of each dictionary word and the score value, and a negative relationship between the score value and the number of all terms in the text. Following these characteristics, the ESG sentiment score based on the word-power approach of Jegadeesh and Wu (2013) is defined as follows:

$$ESGSent_i^{wpw} = \sum_{j=1}^{J} \left(w_j F_{i,j}\right) \frac{1}{a_i} \tag{1}$$

where $ESGSent^{wpw}$ is the ESG sentiment score of 10-K report i based on the word-power weights, w_j is the weight of ESG word j from the dictionary, $F_{i,j}$ is the frequency of use of ESG word j in 10-K report i, and a_i is the overall number of words[7] in 10-K report i. To ensure that the sentiment score is related to the tone passed on to investors, the ESG word weights should be quantified based on the abnormal stock price changes around the disclosure time of information to the markets. Jegadeesh and Wu (2013) suggest following functional form to incorporate this relationship:

$$\Delta r_i = a + b\left(\sum_{j=1}^{J} \left(w_j F_{i,j}\right) \frac{1}{a_i}\right) + \epsilon_i = a + \left(\sum_{j=1}^{J} \left(bw_j F_{i,j}\right) \frac{1}{a_i}\right) + \epsilon_i \tag{2}$$

where r_i is defined as the abnormal return around the filing date of 10-K report i, b is the true value of the regression coefficient of each ESG word weight w_j, and ϵ_i is the error term. The abnormal return around the disclosure date of 10-K report i is defined as follows:

$$\Delta r_i = \prod_{t=0}^{3} ret_{i,t} - \prod_{t=0}^{3} ret_{vwi,t} \tag{3}$$

where $\prod_{t=0}^{3} ret_{i,t}$ the four-day is cumulated stock return of the company that released the 10-K report i at $t = 0$, and $\prod_{t=0}^{3} ret_{vwi,t}$ is the four-day cumulated return of the market index[8] starting at the 10-K report's release date $t = 0$.

In comparison to parameters $F_{i,j}$ and a_i contained in eq. (2) that can be computed based on the documents' statistics, the ESG term weights have to be estimated using a regression approach to ensure the link between the stock market reaction and the weight of each particular ESG word. In this way, the methodology allows to trace the reaction of investors to specific ESG words through changes in stock prices induced by them. To estimate the ESG word weights, we follow the regression approach of Jegadeesh and Wu (2013):

$$\Delta r_i = a + \left(\sum_{j=1}^{J} \left(B_j F_{i,j}\right) \frac{1}{a_i}\right) + \epsilon_i \tag{4}$$

where B_j is the coefficient of the regression and an unbiased estimator of bw_j.[9] To get the estimates of each ESG word weight w_j, the results of the estimation (4) are standardized as follows:

$$\widehat{w}_j = \frac{\widehat{B}_j - \overline{B}}{Standard\ Deviation\ \left(\widehat{B}_j\right)} \tag{5}$$

where $\widehat{w}_j$ defines the calibrated measure for each ESG word's true weight, $\widehat{B}_j$ is the coefficient of the estimated regression slope in (4), and $\overline{B}$ is the average of $\widehat{B}_j$'s among all ESG words. The derived ESG word weights from eq. (5) are implemented to determine the ESG sentiment score for each 10-K report as defined in eq. (1). To analyze the relationship between the sentiment score and the abnormal changes in stock prices, Jegadeesh and Wu (2013) use the specified regression:

$$\Delta r_i = a + b\left(\sum_{j=1}^{J} \left(\widehat{w}_j F_{i,j}\right) \frac{1}{a_i}\right) + \epsilon_i \tag{6}$$

where $\sum_{j=1}^{J} \left(\widehat{w}_j F_{i,j}\right) \frac{1}{a_i}$ is the estimated sentiment score of 10-K report i. Since the word weights are also measured based on abnormal returns, it is necessary to split the sample into the training, or estimation, set that is intended to estimate the word weights using the regressions, and the test set to explore the relationship between the ESG sentiment of documents and the market reaction.[10] The split makes sure that the abnormal returns are not compared to the sentiment scores that include the word weights which are determined using the very same abnormal returns. Thus, to quantify and analyze the ESG sentiment of textual information, this weighting methodology should be used in combination with an ESG-focused dictionary.

2.3. ESG dictionary

To quantify the tone of a specific topic within a text by means of the dictionary approach requires a compilation of a thematic word list that comprises the words which are clearly associated with this particular topic. While the researchers focused on the ESG sentiment applied different approaches from linguistics, including the counting of ESG words, until recently there was no comprehensive ESG dictionary with a broad coverage of the major sustainability topics (Baier et al. (2020)). Besides affiliation of words with ESG subjects, the creation of an ESG word list should also be based on a single type of information source to account for stylistic differences in lexicon used in diverse text genres, such as official documents like 10-K reports with formal language and

[6] Besides that, such term weighting schemes are based on subjective assumptions of researchers with regards torelationships between the content of a document and the induced reaction - for example, a predefined classification of words into positive and negative categories.

[7] Including non-ESG words.

[8] The market index is represented by the value weighted CRSP stock market index, which includes all stocks traded on the NYSE, AMEX, NASDAQ, and ARCA stock exchanges.

[9] The parameters b and wj are not estimated independently at this point to prevent the relative scaling of word weights in an arbitrarily way (see Jegadeesh and Wu (2013), p. 715 for further details).

[10] The common standard in the data science literature defines the ratio of the training set to the test set between 70% to 30% and 80% to 20% of observations. In this paper, we adhere to the 70% to 30% ratio for our analysis based on the number of observations, with the training set based on older observations of the time series and the test set on the latest ones.

online articles with more casual use of vocabulary. Since the annual report is considered as one of the most credible sources of ESG information with respect to incidents, new policies and activities of companies (Hummel et al. (2020)), Baier et al. (2020) developed a broad ESG dictionary based on a sample with 100 10-K reports of companies-members[11] of the S&P 100 index including their proxy statements. Since those companies belong to nine different sectors,[12] the dictionary represents a broad lexicon of ESG terms used in the corporate world among industries to communicate sustainability-related information to different stakeholder groups. To ensure the possibility to analyze particular topics withing a broad ESG context, the researchers designed the ESG word list into several categories and subcategories, thus allowing to evaluate the relative importance of particular topics within the main three ESG pillars using the content analysis. The manual judgement and classification of words ensures that the ESG word list gets only words that are predominantly used in context of sustainability, with the researchers deliberately discarding the word chains from their analysis to avoid the double counting of terms which could miscalibrate the word weights.

The ESG dictionary of Baier et al. (2020) includes 482 words structured into 10 categories and 34[13] subcategories, which are distributed under the three ESG pillars. The Governance topic comprises "corporate governance", "business ethics", and "sustainability management and reporting" categories, while the Environment pillar is represented by "climate change", "ecosystem service", and "environmental management" categories. With respect to the Social pillar, "public health", "human rights", "labor standards", and "society" categories incorporate all words that describe social-related subjects in 10-K reports. Since some of ESG words cannot be assigned to a specific subcategory due to their broad interpretation and use across several subcategories, the words are also distributed on the higher levels in the dictionary structure than subcategories.[14] Despite an imbalance in the number of words from the three major ESG topics used in 10-K reports selected for the sample,[15] the researchers managed to create a comprehensive ESG dictionary that covers all sustainability aspects in detail. Thus, the ESG dictionary of Baier et al. (2020) is best suited to analyze the reporting of ESG information of companies in 10-K reports, and in combination with the market-based word weighting scheme, to explore the reaction of markets to particular ESG topics and to determine their relevance for investors. However, the dictionary with 482 items also contains inflected forms of words together with their base forms, which results in an undue duplication of words with the same semantic meaning. Since the estimation of term weights within the word-power approach is performed by means of the regression analysis, the inclusion of inflected word forms can distort the estimation results due to multicollinearity in the regression model. To avoid such misestimation of the term weights, we apply the Porter stemming algorithm to the word list of Baier et al. (2020) which eliminates "the commoner morphological and inflexional endings from words"[16] and transforms them into their stem forms. Thus, the final version of the dictionary used in this study contains 299 unique ESG words, while the stemming procedure does not impact the number of categories and subcategories originally defined in the dictionary.

3. Data sample and empirical results

3.1. Data sample construction and descriptive statistics

Our sample consists of 10-K reports of US companies[17] in period between January 2013 to December 2019 obtained through the SEC's EDGAR[18] database by means of a web-scraping algorithm developed in Python. Our choice of the time frame is based on the increase in relevance of ESG topics in economic and social matters in the last decade[19] and the availability of fundamental and market data of companies analyzed, while the concentration on the US market is related to several factors, including high efficiency of capital market, standardized filing of reports with SEC via a centralized database, and the availability of 10-K reports of all companies in an electronic format. In addition to textual data from EDGAR, the sample comprises stock and market return data from the CRSP as well as fundamentals data from the Compustat databases.[20] Since the match of 10-K reports from the EDGAR database and market/fundamentals data from CRSP/Compustat databases is possible only through the Central Index Key (CIK) number, we begin the derivation of the dataset with identification of US companies that are available in the Compustat database with historic[21] CIK numbers. In the next step, the historic CIK numbers of 3401 companies are used to select 10-K reports available in the EDGAR database, thus reducing the sample size to 3162 companies. The resulting 18,729 10-K reports are processed and merged with data from CRSP and Compustat databases, which reduces the number of observations to 17,009. Table 1 summarizes the procedure of the data sample construction. Since the word weighting scheme of Jegadeesh and Wu (2013) requires the estimation of weights with regressions, we split our data sample into the training and the test sets according to data science benchmarks. Thus, the training set, which is used for estimation of ESG word weights, contains 12,296 unique 10-K reports from January 2013 to December 2017 period, whereas the test set includes 4713 files of 2446 companies from January 2018 to December 2019 period.

Since the EDGAR database contains 10-K reports in the HTML-format, all reports are cleaned from HTML language attributes along with tables, images, exhibits, and the standard cover text using customized algorithms and data processing libraries in Python. Furthermore, the resulting text corpora are tokenized and passed through the English dictionary[22] to screen out all non-word items, while common stop words as well as abbreviations and proper names are also deleted from texts. All remaining words are modified with the Porter stemming algorithm to ensure their comparability with stemmed forms of words from the ESG dictionary. The words which are contained in both text corpus and ESG dictionary build the core of the term-document matrix for each 10-K document in the sample that is used for estimation of ESG sentiment scores. In addition to calculated abnormal returns, the dataset also contains several control variables used in the regression analysis. Table 2 summarizes descriptive statistics of the complete data

[11] The sample includes the 25 largest firms in terms of their market value.

[12] Industrials, Energy, Consumer Discretionary, Financials, Consumer Staples, IT software, Health Care, IT hard- ware, Telecommunication Services (Classification according to the Global Industry Classification Standard).

[13] While Baier et al. (2020) define 40 subcategories in their dictionary, 6 of them do not contain any ESG words. Nevertheless, the authors include these subcategories into the structure of the ESG dictionary since they are addressed by words with a broader meaning that belong to the higher-level categories.

[14] Topics (Environment, Social, Governance) and the ten categories, see Appendix B for a detailed description.

[15] The authors mention that the vast majority (over 80%) of ESG words contained in their sample belong to the Government topic, followed by Social and Environment pillars.

[16] See https://tartarus.org/martin/PorterStemmer for further information.

[17] Companies with a country domicile in the United States of America.

[18] Electronic Data Gathering, Analysis, and Retrieval System; see https://www.sec.gov/edgar/about for further information.

[19] Du and Yu (2021) analyze a sample of CSR reports of Fortune 500 companies in period 2002 to 2015 and observe the growth in number of issued reports leveling off in years 2013 and 2014. See also Serafeim and Yoon (2022).

[20] Both databases are accessed through the Wharton Research Data Services, see https://wrds-www.wharton. upenn.edu for further information.

[21] Since the CIK numbers can be reassigned to companies throughout their stock trading history, e.g., in cases of M&A activity or a relisting after bankruptcy, it is crucial to identify the historic CIK numbers of companies to ensure a proper merge of 10-K reports with corresponding company data.

[22] We follow Jegadeesh and Wu (2013) and apply the 2of12inf general dictionary of English language, accessible athttp://wordlist.aspell.net/12dicts.

Table 1

Derivation of the data sample. This table presents the compilation process of the 10-K data sample with applied adjustment steps and resulting observation numbers. The complete sample consists of 10-K reports of US companies in period January 2013 to December 2019 along with corresponding market and fundamentals data of companies. All 10-K reports are cleaned and processed prior to the merge with other data, including removal of HTML language elements, tables, images, exhibits, and the standard cover text of a 10-K report. All identified words in each 10-K report, after removal of non-word items, common stop words, abbreviations, and proper names, are modified with the Porter stemming algorithm to ensure the comparability with stemmed words from the ESG dictionary. The final sample is constructed by matching EDGAR, CRSP, and Compustat databases using the historic CIK, CUSIP numbers, and tickers as unique company identifiers.

Data sample/adjustment	Number of companies/reports	Removed companies/reports
US companies with historic CIK numbers available in Compustat	3401	
of which 10-K reports available in EDGAR	3162	239
EDGAR 10-K 2013–2019 US sample	18729	
without duplicates	18709	20
merge with Ticker/CIK library from Compustat	18075	634
merge with price returns database from CRSP	17009	1066
10-K/Abnormal return training sample	12296	
10-K/Abnormal return test sample	4713	
Complete 10-K/Abnormal return US 2013–2019 sample	17009	

Table 2

Summary statistics of the complete dataset (2013–2019). This table presents the descriptive statistics for the complete dataset with 10-K reports of US companies in period between January 2013 to December 2019 including the filing period abnormal stock returns of companies together with the control variables used in regressions. Abnormal return is defined as company's cumulated stock return over a four-day 10-K filing period window ($t-1$, $t+2$) in excess of the market return represented by the CRSP value-weighted market index. Earnings announcement return is defined as company's cumulated stock return over a three-day window around the earnings announcement ($t-1$, $t+1$) in excess of the market return represented by the CRSP value-weighted market index. Book-to-market is defined as the ratio of the book value of equity to the market value of equity at the end of the quarter before 10-K report filing. Size is defined as the natural logarithm of the market capitalization of company at the end of the quarter before 10-K report filing. Accruals are defined according to Sloan (1996) as one-quarter difference in current assets without cash subtracting change in current liabilities without change in debt included in current liabilities and change in income taxes payable minus depreciation and amortization expense. Volatility is defined as a 250-day rolling standard deviation of company's firm-specific (excess) stock return. Shares turnover is defined as the natural logarithm of the number of shares traded over 250 days before the 10-K report filing date divided by the number of shares outstanding on the filing date. ESG, Environmental, Social, and Governance performance are Arabesque S-Ray@ performance scores of a company in the overall ESG category as well as singular pillars available on the 10-K report's filing date. The extreme values of book-to-market, accruals, and volatility variables do not affect the coefficients and their significance in econometric models.

Variable	Observations	Mean	SD	Min.	25%	50%	75%	Max.
Total words	17009	30991	15356	120	23272	29112	36188	785354
ESG words	17009	1648	724	2	1201	1504	1935	11944
Environmental words	17009	116	171	0	40	63	109	2654
Social words	17009	379	240	0	234	312	442	3209
Governance words	17009	1273	531	1	952	1184	1488	8887
Abnormal return	17009	0.008	0.114	−0.659	−0.021	0.001	0.027	10.529
Earnings announcement return	16999	0.005	0.087	−0.701	−0.036	0.002	0.043	1.056
Book-to-Market	16511	0.459	2.645	−318.469	0.208	0.402	0.671	35.881
Size	16537	7.562	1.683	0.462	6.420	7.463	8.59	13.686
Accruals	16543	0.000	1.475	−2.139	−0.023	−0.007	0.003	189429.
Volatility	16377	0.427	4.457	0.079	0.202	0.281	0.413	517.630
Shares turnover	16376	−4.97	0.730	−9.814	−5.388	−4.959	−4.529	−0.638
ESG performance	8097	50.76	8.343	22.91	45.04	51.02	56.83	77.21
Environmental performance	8097	46.93	13.461	22.03	34.35	43.45	57.53	87.89
Social performance	8097	51.08	8.712	22.78	44.41	50.22	57.27	80.93
Governance performance	8097	51.81	12.870	11.33	43.41	52.69	61.18	86.69

sample, whereas Table 3 presents the description of the test set used in regressions to analyze the relationship between ESG sentiment and abnormal stock returns.

The number of unique US firms per year in the dataset remains stable around 2500 companies, with firms issuing larger reports in terms of the total number of words year after year. The average 10-K report contains around thirty-one thousand words, with three-quarters of 10-K reports not exceeding the level of thirty-seven thousand words. The mean number of ESG words used in a 10-K report is 1648, with 75% of reports containing less than two thousand ESG words. Of the three sustainability pillars, the Governance category topics get the greatest coverage in 10-K reports in terms of the average number of words (1273), followed by Social (379) and Environmental (116) pillars. The average filing period abnormal return in the complete set is 0.8%, whereas the average earnings announcement return is 0.5%. With regards to the test set that includes 4713 10-K reports filed in the period between January 2018 to December 2019, the average number of total words per 10-K report amounts to 32348, with 75% of reports providing around thirty-eight thousand words or less. Similar to the complete data set, the Governance category shows the highest number of average words used in a report (1353), with Social (403) and Environmental (122) pillars trailing behind. The word-power ESG sentiment score ranges between 0.09 and 0.6 with a mean value of 0.274, while the mean abnormal return of the test period is 0.46%. Fig. 1 demonstrates the frequency distribution of abnormal returns in both the complete and the test datasets along with the distribution of ESG sentiment scores of 10-K reports in the test set.

The distribution of abnormal returns in the test set is similar to the distribution observed in the complete sample, with both graphs showing expected positive kurtosis with fat tails.[23] The ESG sentiment score has a positively skewed frequency distribution, with the minimum value of 0.092 and the maximum value of 0.601, while the difference between 25% and 75% quantiles amounts to 0.057. The training data set, which

[23] We do not exclude the extreme values of abnormal returns from our sample to avoid the selection bias of market reactions that could be induced by ESG tone of 10-K reports.

Table 3

Summary statistics of the test dataset (2018–2019). This table reports the descriptive statistics for the test dataset with 10-K reports of US companies in period between January 2018 to December 2019, including the filing period abnormal stock returns of companies along with the control variables used in regressions. The test set data are used in regressions to examine the market's reaction to the ESG sentiment of 10-K reports. Abnormal return is defined as company's cumulated stock return over a four-day 10-K filing period window ($t-1$, $t+2$) in excess of the market return represented by the CRSP value-weighted market index. ESG sentiment score is the estimated ESG tone of 10-K reports based on the word weights calibrated using data from the training set and Eqs. (4) and (5). Earnings announcement return is defined as company's cumulated stock return over a three-day window around the earnings announcement ($t-1$, $t+1$) in excess of the market return represented by the CRSP value-weighted market index. Book-to-market is defined as the ratio of the book value of equity to the market value of equity at the end of the quarter before 10-K report filing. Size is defined as the natural logarithm of the market capitalization of company at the end of the quarter before 10-K report filing. Accruals is defined according to Sloan (1996) as one-quarter difference in current assets without cash subtracting change in current liabilities without change in debt included in current liabilities and change in income taxes payable minus depreciation and amortization expense. Volatility is defined as a 250-day rolling standard deviation of company's firm-specific (excess) stock return. Shares turnover is defined as the natural logarithm of the number of shares traded over 250 days before the 10-K report filing date divided by the number of shares outstanding on the filing date. The extreme values of book-to-market, accruals, and volatility variables do not affect the coefficients and their significance in econometric models.

Variable	Observations	Mean	SD	Min.	25%	50%	75%	Max.
Total words	4713	32348	12866	1528	24831	30568	37.931	219.502
ESG words	4713	1756	737	23	1295	1602	2058	11.944
Environmental words	4713	122	174	0	43	68	112	2243
Social words	4713	403	251	0	249	330	470	2164
Governance words	4713	1353	534	23	1023	1265	1579	8887
Abnormal return	4713	0.005	0.082	−0.654	−0.025	−0.000	0.028	1.419
ESG Sentiment$_{wpw}$	4712	0.274	0.050	0.092	0.240	0.265	0.297	0.601
ESG Disclosure$_{idf}$	4712	14.240	3.163	3.347	11.990	13.645	16.029	27.064
Earnings announcement return	4710	0.005	0.092	−0.591	−0.039	0.002	0.046	1.056
Book-to-Market	4329	0.406	4.883	−318.469	0.195	0.409	0.698	9.189
Size	4339	7.759	1.702	0.789	6.649	7.648	8.800	13.686
Accruals	4343	−0.010	0.063	−0.671	−0.023	−0.007	0.003	0.799
Volatility	4644	0.504	7.677	0.087	0.212	0.286	0.420	517.630
Shares turnover	4644	−4.93	0.711	−9.814	−5.327	−4.925	−4.513	−1.331

includes 10-K reports filed in the period January 2013 to December 2017 and is used for calibration of word power weights, shows descriptive characteristics similar to the complete dataset. The results of the estimation of ESG word weights using the calibration set allow to identify the words with the highest positive and negative impact on markets, while the frequency of words' usage reveals the communication patterns of companies with their stakeholders.

3.2. ESG lexicon analysis

To get a general picture of the ESG lexicon usage of companies, we apply the "word cloud" analysis tool – a technique of graphical visualization of text data widely used in the data mining applications and Big Data research, which depicts the words according to the frequency of their usage in text (Jin (2017)). The relative appearance of single words is reflected through their font size in the cloud, which eases the identification of the most relevant topics in reports and the comparison of different semantic categories, such as positive and negative textual tones (Barbeito-Caamano and Chalmeta (2020)). The word cloud visualization approach finds application in various disciplines including linguistics, economics, politics, and medicine (Atenstaedt (2012)). We apply the world cloud algorithm in Python on the test set that includes 4712[24] 10-K reports to explore the most frequently used ESG words by US companies, whereas the word clouds with positive and negative ESG words are obtained based on the word-power weights calibrated on the training sample consisting of 12296 10-K reports. Fig. 2 shows the word cloud with a complete lexicon used in 10-K reports as well as the clouds with the most impactful negative and positive ESG words.

The analysis of ESG lexicon used by companies indicates the prevalence of Governance topics in 10-K reports, with "control", "compensation", "award", "compliance", "disclosure", and "audit" being among the words most frequently mentioned in texts. The focus on Governance pillar matches the results of other studies[25] that look at annual reports of companies, while the word cloud analysis allows to identify most relevant words used to convey governance-related information to stakeholders. The word cloud containing ESG word stems with negative weights identifies several word groups that negatively affect the stock prices, among them "discrimin", "gay", and "lesbian", emphasizing possible investors' dissatisfaction in cases of discrimination of the LGBT community, or word groups "crime", "payout", "embezzle", pointing out illegal cases of misuse of funds. However, socially necessary actions encouraged by the public and depicted by word stems "veteran", "charit", "philantrop" also induce negative market reaction, which can be associated with additional spending of company funds that is not supported by investors. Positive stock returns, in turn, are associated with ESG word stems such as "oversight", "harass", "inspect", "sustain", highlighting most likely the support of investors for oversight activities regarding harassment cases, as well as measures aimed at supporting employees and their professional development represented by word stems "labour", "scholarship", "learn", "talent", and "success". Along with the direction of the impact, the word-power weights provide opportunity to identify the most impactful ESG words in both positive and negative categories. Fig. 3 presents words with largest weights based on the regression estimates in each category.

The analysis of word weights confirms initial observations obtained using the word cloud technique, with topics covering endowments, embezzlements, discrimination, and crimes being negatively perceived by investors. At the opposite end of the tonal spectrum, words such as "community", "health", "people", and "control", among others, cause the strongest reaction of markets, with words that belong to the Governance category amounting to the half of the top 20 most impactful positive and negative words. With regards to the Environmental pillar, emissions in the atmosphere and deforestation are associated with negative stock returns, whereas positive issues concerning water, coal, and wilderness are particularly welcomed by markets. Social terms make up around a third of the most effective words in both tonal categories, with words that describe community actions and people's health

[24] We exclude one document with a missing ESG sentiment score to align the dataset used in the word cloud analysis with the sample used in the regression analysis part.

[25] See, for example, Baier et al. (2020).

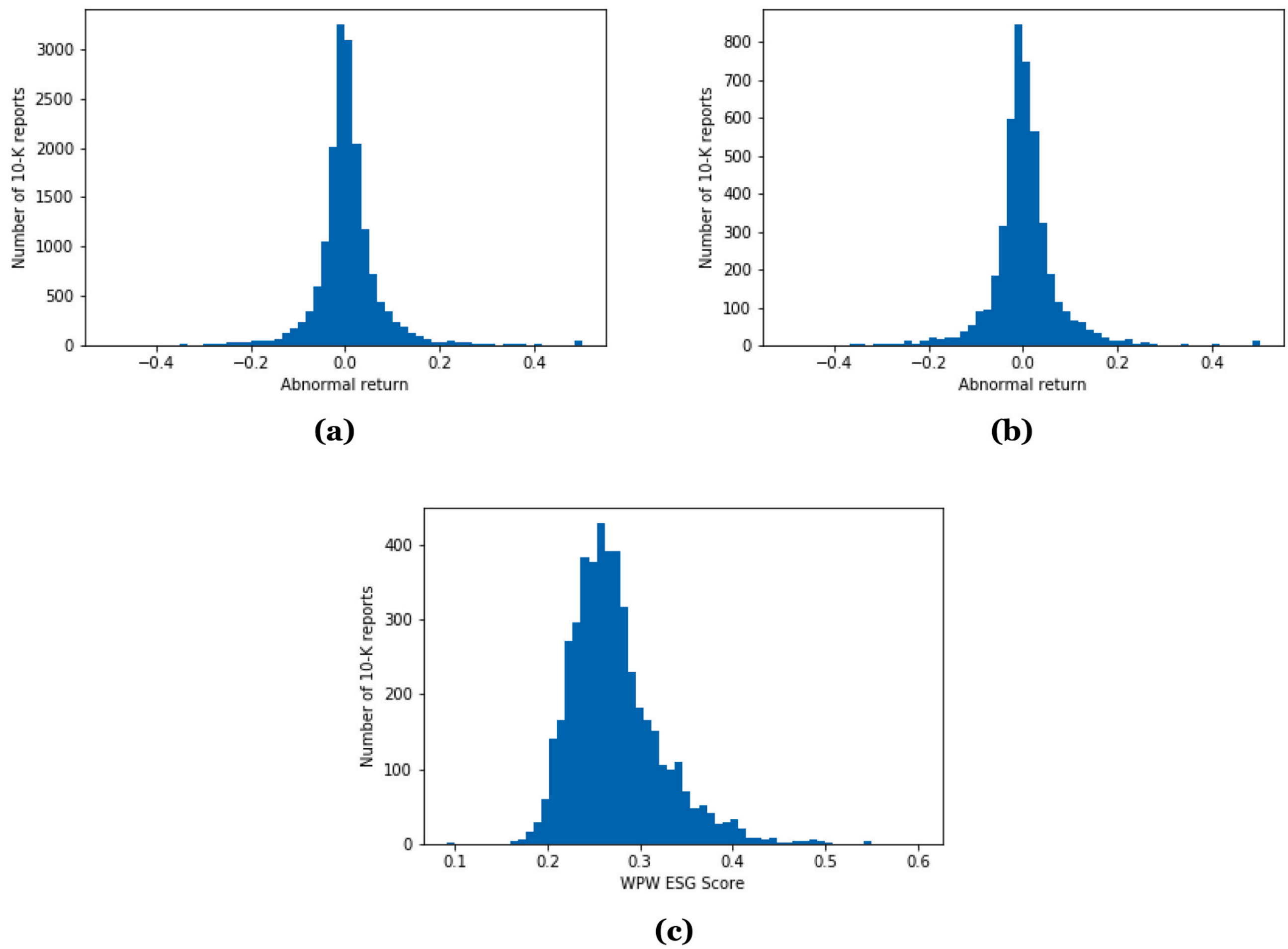

Fig. 1. Distribution of abnormal returns and ESG sentiment score. This figure presents the frequency distribution of abnormal returns and ESG sentiment score estimated based on the word-power weights. Graph (a) depicts the frequency distribution of abnormal returns in the complete dataset which covers the period between January 2013 and December 2019. Graph (b) depicts the frequency distribution of abnormal returns in the test dataset which covers the period between January 2018 and December 2019. Graph (c) presents the frequency distribution of the ESG sentiment score of 10-K reports filed in the period between January 2018 and December 2019 that are contained in the test dataset. Abnormal return is defined as company's cumulated stock return over a four-day 10-K filing period window $(t-1, t+2)$ in excess of the market return represented by the CRSP value-weighted market index. ESG sentiment score is the estimated ESG tone of annular reports based on the word weights calibrated using data from the training set and Eqs. (4) and (5).

showing a positive tone, and words covering discrimination issues and endowment activities a negative tone, respectively. The analysis on the individual words level using a large sample enables to identify terms and expressions that attract the most attention from investors. However, to examine the overall impact of the ESG tone on markets, it is necessary to look at the interaction of different ESG topics which result in a particular picture of investors' understanding of and their attitude to company's ESG stand and agenda. The interplay of different tones is captured by the weighted ESG sentiment score, which is suited for regression analysis to examine the general effect of ESG tone on financial markets.

3.3. Regression analysis

Word-power weights. The ESG tone of 10-K reports is represented with a quantitative sentiment score based on word weights estimated using the word-power approach as defined in Eq. (1). While the word weights are estimated based on the training set with 10-K reports filed between January 2013 and December 2017, the ESG sentiment scores are calculated for reports contained in the test set of the sample, i.e., filed between January 2018 and December 2019, based on word weights calibrated on the training set of the complete sample. Such design ensures that word weights are not estimated using the same abnormal returns that are used in regressions on sentiment scores calculated based on these word weights. To examine the relationship between the ESG sentiment score and associated stock market movements, we fit an Ordinary Least Squares (OLS) regression by regressing the abnormal stock returns on ESG sentiment score with various control variables. Based on Eq. (6), we define the regression as follows:

$$AbnReturn_i = a + b \times ESGSent_{wpwi} + \epsilon_i \tag{7}$$

where $AbnReturn_i$ is a four-day abnormal return of company i around filing of the 10-K report as defined in Eq. (3), and $ESGSent_{wpwi}$ is the ESG sentiment score of 10-K report of company i based on word-power weights. We include time and industry fixed effects in our regression analysis to account for differences in ESG reporting of companies that belong to industries with contrasting sustainability imprints[26] and levels of litigation risk,[27] and use standard errors clustered on the firm level dimension to control for potential correlation between observations that

[26] Current research finds substantial differences in reporting patterns of companies from "good" and "bad" industries(in terms of their sustainability performance; see, e.g., Nazari, Hrazdil, and Mahmoudian (2017)).

[27] Prompted by possible consequences in case of a breach of obligations and the resultant substantial harm to the environment and the public that must be reimbursed by the company (see Da Silva (2020)).

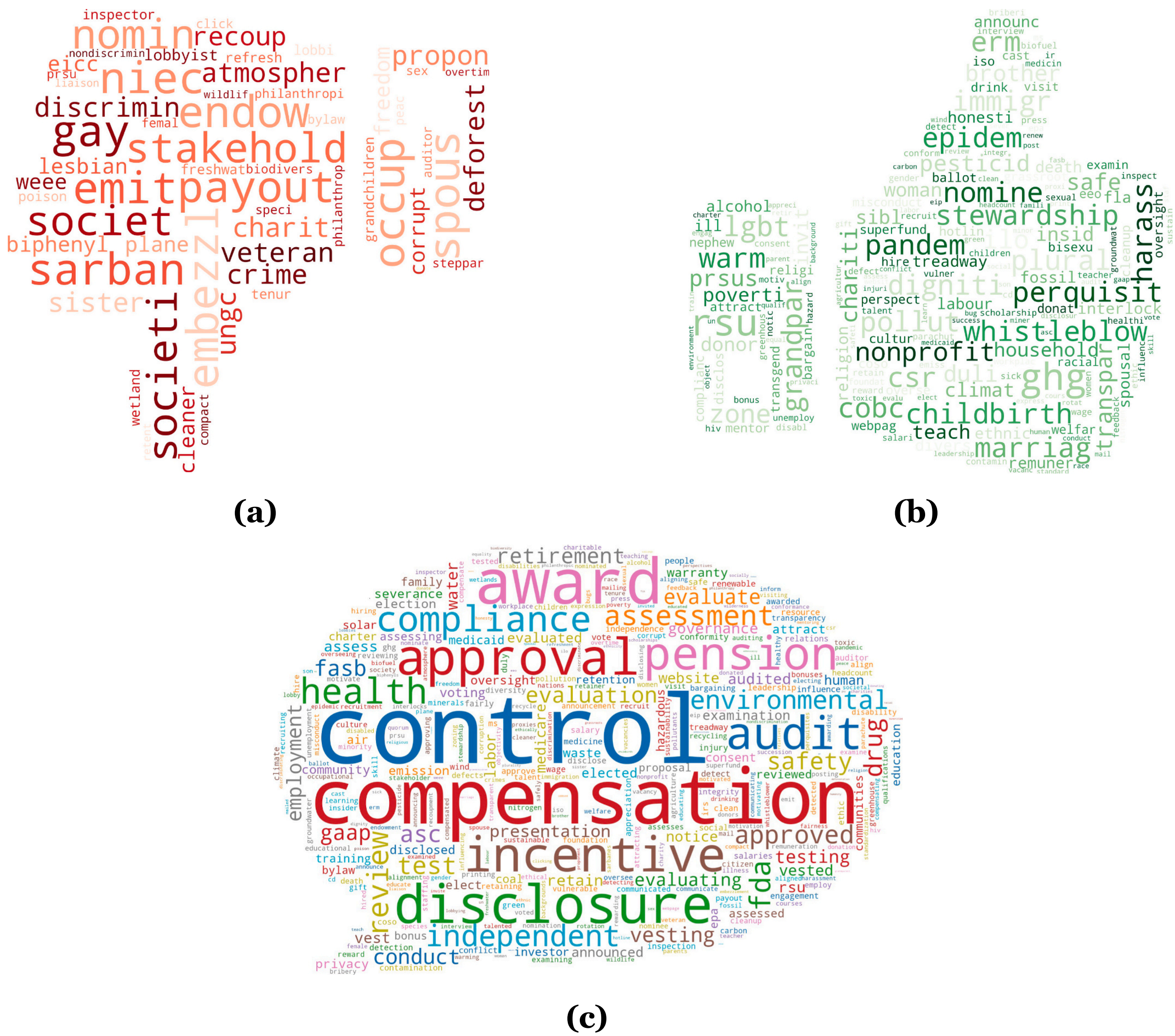

Fig. 2. ESG word clouds. This figure shows the word clouds with ESG lexicon used in 10-K reports. Graph (a) presents the most impactful negative ESG word stems in terms of induced market reaction according to the word weights estimation (words with negative regression coefficients only). Graph (b) presents the most impactful positive ESG word stems in terms of induced market reaction according to the word weights estimation (words with positive regression coefficients only). Graph (c) presents the word cloud with a complete ESG lexicon used in 10-K reports contained in the test sample, where more frequently used words have larger font size. The test sample includes 4712 10-K reports of US companies filed in the period January 2018 to December 2019. The training sample, which is used for estimation of word weights based on Eqs. (4) and (5), includes 12296 10-K reports of US companies filed in the period between January 2013 to December 2017. The word stems presented in graphs (a) and (b) are transformed using the Porter stemming algorithm on the regression estimates in each category.

are distributed over the time frame but belong to the same company (Petersen (2009)). To control for other effects[28] on stock returns known from the literature, we include additional variables in our regression, extending the regression (7) as follows:

$$AbnReturn_i = a + b \times ESGSent_{wpwi} + c \times Size_i + d \times BM_i + e \times Accruals_i. + f \times Vola_i + g \times ShareTurn_i + h \times EarnDayRet_i + \epsilon_i \tag{8}$$

where $Size_i$ is defined as the natural logarithm of the market capitalization of company at the end of the quarter before 10-K report filing, BM_i is the ratio of the book value of equity to the market value of equity

[28] Control for other textual tones, such as pure positive tone based on the dictionary of Loughran and McDonald (2011), does not significantly impact the results of our tests.

Top 20 negative ESG words

endowment	nieces	occupational	atmosphere	crimes
payout	embezzlement	stakeholder	veteran	UNGC
sarbanes	gay	society	proponent	discriminate
emit	spouse	nominate	deforestation	charitable

(a)

Top 20 positive ESG words

citizen	fair	health	control	compensate
community	drug	people	governance	approve
relatives	nation	test	coal	wilderness
water	pension	quorum	inform	vest

(b)

Fig. 3. The most impactful positive and negative ESG words. Figure (a) presents top 20 ESG words with most negative impact on stock returns according to the regression analysis, where "endowment" and "payout" have the largest negative weights, and "discriminate" and "charitable" the smallest negative weights. Figure (b) presents top 20 ESG words with most positive impact on stock returns according to the regression analysis, where "citizen" and "community" have the largest positive weights, and "wilderness" and "vest" the smallest positive weights. All word weights are obtained based on Eqs. (4) and (5) using the training set with 10-K reports of US companies filed in the period between January 2013 to December 2017. The words are presented in their initial forms as defined in the dictionary of Baier et al. (2020), while the estimation of weights was conducted using the stemmed versions of words to eliminate the impact of word forms contained in the dictionary.

Table 4
Regression analysis. This table presents the regression estimates of the filing period abnormal stock returns on ESG sentiment and disclosure scores with various control variables on the sample of 10-K reports filed in the period between January 2018 and December 2019. $ESGSent_{wpw}$ is the ESG sentiment score of 10-K report based on word-power weights as defined in Eq. (1). $ESGDisc_{idf}$, $EnvDisc_{idf}$, $SocDisc_{idf}$, and $GovDisc_{idf}$ are respectively ESG, Environmental, Social, and Governance disclosure scores of 10-K reports based on *idf* word weights computed as defined in Eqs. (9) and (10). *Size* is defined as the natural logarithm of the market capitalization of company at the end of the quarter before 10-K report filing. *BM* is defined as the ratio of the book value of equity to the market value of equity at the end of the quarter before 10-K report filing. *Accruals* are defined as one-quarter difference in current assets without cash subtracting change in current liabilities without change in debt included in current liabilities and change in income taxes payable minus depreciation and amortization expense. *Vola* is a 250-day rolling standard deviation of company's firm-specific (excess) stock return. *ShareTurn* is the natural logarithm of the number of shares traded over 250 days before the 10-K report filing date divided by the number of shares outstanding on the filing date. *EarnDayRet* is company's cumulated stock return over a three-day window around the earnings announcement $(t-1, t+1)$ in excess of the market return represented by the CRSP value-weighted market index. All OLS regression models are estimated with time and industry fixed effects as well as standard errors clustered on the firm level. The regression models with "†" sign denote the regressions run only on the Mining division companies according to the SIC codes. The number of observations changes throughout the models due to limited availability of control variables for each regression model. All independent variables are standardized to a mean of 0 and a standard deviation of 1. The coefficients' estimates are not affected by the presence of outliers in the control variables. "***" denotes the 1% significance level, "**" the 5%, and "*" the 10% level, respectively. The values in parentheses report the standard errors of estimated coefficients.

	Models								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)†	(9)†
$ESGSent_{wpw}$	0.0338**	0.0238**						−0.1140	−0.0776
	(0.016)	(0.012)						(0.110)	(0.089)
$ESGDisc_{idf}$			0.0399***	0.0322***					
			(0.011)	(0.009)					
$EnvDisc_{idf}$					−0.0139	0.0023	−0.0015		
					(0.009)	(0.008)	(0.008)		
$SocDisc_{idf}$					0.0414***	0.0376***			
					(0.012)	(0.008)			
$GovDisc_{idf}$					0.0302*		0.0324**		
					(0.018)		(0.013)		
Size		−0.0435***		−0.0459***		−0.0436***	−0.0440***		−0.0314
		(0.009)		(0.009)		(0.009)	(0.009)		(0.046)
BM		0.0185		0.0153		0.0214	0.0197		−3.1673
		(0.036)		(0.037)		(0.034)	(0.037)		(4.060)
Accruals		−0.0754*		−0.0783**		−0.0797**	−0.0747*		−0.0160
		(0.040)		(0.040)		(0.040)	(0.040)		(0.199)
Vola		0.0243*		0.0275**		0.0232*	0.0234		0.4430
		(0.015)		(0.014)		(0.014)	(0.015)		(0.381)
ShareTurn		0.0233		0.0176		0.0134	0.0185		0.0613
		(0.015)		(0.015)		(0.015)	(0.016)		(0.061)
EarnDayRet		0.6625***		0.6613***		0.6585***	0.6617***		1.0338***
		(0.045)		(0.045)		(0.045)	(0.045)		(0.203)
Observations	4712	4259	4712	4259	4712	4259	4259	185	170

at the end of the quarter before 10-K report filing, $Accruals_i$ are defined[29] according to Sloan (1996), $Vola_i$ is a 250-day rolling standard deviation of company's firm-specific stock return, $ShareTurn_i$ is defined as the natural logarithm of the number of shares traded over 250 days before the 10-K report filing date divided by the number of shares outstanding on the filing date, and $EarnDayRet_i$ is defined as company's cumulated stock return over a three-day window around the earnings announcement $(t-1, t+1)$ in excess of the market return represented by the CRSP value weighted market index following Jegadeesh and Wu (2013) and Eugster and Wagner (2021). Models 1 and 2 in Table 4 report the estimated coefficients of regressions. In both regression models, the ESG sentiment coefficient (0.0338 and 0.0238, respectively) is positive and significant at the 5% level, implying a positive effect of ESG tone of 10-K reports on abnormal stock returns of companies, with stronger ESG tone conveyed to investors being related to higher returns during the filing event window. With regards to the control variables, size and earnings announcement return show estimates significant at the 1% level, with

[29] One-quarter difference in current assets without cash minus change in current liabilities without change in debt included in current liabilities and change in income taxes payable minus depreciation and amortization expense.

$Size_i$ having a negative effect on abnormal returns and coefficient of −0.0435, and $EarnDayRet_i$ a positive effect with the coefficient of 0.6625, respectively. Along with size, accruals also negatively impact the stock returns, with the coefficient of −0.0754 being significant at the 10% level. Volatility, in turn, adds positively to abnormal returns at the 10% significance level with 0.0243 coefficient. The regression results underpin our expectations about the role of qualitative ESG information, which is conveyed through 10-K reports, in investors' opinion making on companies' fair market value. The results of estimated coefficients for control variables largely match the observations obtained in other studies from the literature.

Litigation risk industries. Companies that belong to environmentally harmful industries are exposed to additional ESG risks, since every incident associated with a hazard for nature or society can potentially result in costly lawsuits, reimbursements, and public embarrassment for the company. Due to a close attention paid by regulators and the public, companies have incentives to notably emphasize ESG topics in their reporting and to try to shape public opinion on the performance of business operations in terms of their ESG impact, thus lessening the risk of potential accusations. Since the 10-K report is one of the major sources of ESG information for a broad range of stakeholders, companies with a high litigation risk can have higher ESG disclosure levels in comparison to their peers (Da Silva (2020)). In particular, the current research finds elevated disclosure levels in annual reports of companies that belong to the Mining industry,[30] which is exposed to a high litigation risk in the environmental area due to the nature of business operations. To analyze the reaction of financial markets to the ESG tone of 10-K reports issued by such companies,[31] we run the regressions (7) and (8) on a reduced sample that includes only companies from the Mining sector. Models 8 and 9 in Table 4 report the results of regression estimates. In contrast to the full sample with all industries, the coefficient of $ESGSent_{wpwi}$ is negative in both regression models, with coefficient estimates of −0.1140 in model 8 and − 0.0776 in model 9. This result implies that extensive ESG reporting of mining companies does not lead to more favorable reaction of investors on stock markets, but rather the other way round. However, both coefficients are not statistically significant, while the sample size decreased from 4712 to 185 observations in the first case and from 4259 to 170 in the second, respectively. From the control variables, only the earnings announcement return shows strong significance with 1.0338 coefficient estimate. While a small sample size of companies from the Mining sector does not yield statistically significant results with regards to the ESG tone, this outcome still doesn't support the trend seen in other industries, which may serve as motivation for further studies on companies with high litigation risk based on larger datasets.

Inverse document frequency (*idf*) weights. While the market-focused weighting methodology suggested by Jegadeesh and Wu (2013) rests on estimation procedure that uses a part of sample observations to determine the term weights, the standard approach applied in the finance literature on disclosure analysis relies on relative frequency of words used in all documents included in the sample. The method assumes that the more often a certain word is used across all documents, the less informational load it carries, thus lowering the value of words that belong to the standard lexicon of all companies. Whereas the regression estimators of word-power weights can be sensitive to the ratio of observations included in the training sample and their distribution over the timeline, the *idf* -based weighting methodology enables to determine the term weights based on the same observations that are used to examine the relationship between the extent of ESG disclosure[32] and the reaction of markets. Thus, to compare the results obtained with the regression-based estimation of word weights, we compute the ESG document scores with *idf* -based weights for 10-K reports contained in the test sample that is used to analyze the relationship between the ESG tone and abnormal stock returns, while the *idf* weights are also estimated using the same 10-K reports from the test set[33] and thus without the exclusion bias. We compute the *idf* -based word weights according to Loughran and McDonald (2011) as follows[34]:

$$w_{i,j}^{idf} = \begin{cases} 1 + log\left(tf_{i,j}\right)^{*}log\left(\dfrac{N}{df_j}\right) \; if \; tf_{i,j} > 0, \\ 0 \; otherwise \end{cases} \tag{9}$$

where w^{idf} is the weight of ESG word j in the ESG score of 10-K report i, $tf_{i,j}$ is the use frequency of ESG word j in 10-K report i, N is the number of 10-K reports contained in the sample, and df_j is the number of sample's 10-K reports in which ESG word j is used at least once. Since the computed word weights already include their relative proportion in the ESG disclosure score of each 10-K report, we can compute the score for the 10-K report by simply adding up relative word weights of words that are contained in that particular report. Thus, the ESG disclosure score is defined as follows:

$$ESGDisc_i^{idf} = \frac{1}{(1 + loga_i)} \sum_{j=1}^{J} w_{i,j}^{idf} \tag{10}$$

where $ESGDisc^{idf}$ is the ESG disclosure score of 10-K report i, and a_i is the number of words in 10-K report i. To examine the relationship between the ESG disclosure extent and the reaction of investors, we substitute the ESC tone scores estimated based on the word-power approach in regressions (7) and (8) with ESG scores computed based on *idf* weights as defined in eq. (10). Models 3 and 4 in Table 4 report estimated coefficients of regressions with abnormal stock returns. In line with models 1 and 2, both regressions yield significant positive coefficients for the ESG score, with the single variable model estimate of 0.0399 and 0.0322 with control variables, respectively. This result implies that the markets react also positively to the extent and quality of ESG disclosure, awarding companies that devote more attention to ESG-related topics in their 10-K reports with higher capitalization values. With regards to other control variables used in model 4, both the coefficients' values as well as their significance show similar results to model 2. Thus, the analysis based on *idf* word weights corroborates our results obtained in the previous part.

While the estimation of term weights based on regressions ideally requires a large number of words in the dictionary to yield a balanced calibration process, the *idf* -based approach enables to compute the word weights and thus the scores even with a limited number of words of interest. Since the Governance category has a significant overweight in the ESG dictionary as well as in 10-K reports of companies relative to the other two categories, we do not compute the tone scores for separate

[30] According to the SIC, the Mining division includes companies operating in metal and coal mining, oil and gasextraction, mining and quarrying of nonmetallic minerals areas.

[31] Companies belonging to the Mining sector are identified according to their SIC codes (1000–1499).

[32] However, the *idf* -based weights do not convey the tonal spectrum of words in terms of their positive, neutral, or negative tone, so that in this case we can analyze rather the disclosure extent of ESG topics than the ESG tone itself. Therefore, we use the word-power approach of Jegadeesh and Wu (2013) for our main analysis.

[33] We also compute the *idf* -based weights using the complete 10-K sample from January 2013 to December 2019 with no significant impact on the results of our regression tests. For comparison reasons, we focus on the weights obtained using the test set to examine the results based on the same data set.

[34] See Jegadeesh and Wu (2013) p. 714 for a detailed derivation of formulas.

ESG categories using the word-power approach. However, the mechanics of *idf* weights, which are not dependent on regression estimations, allow us to analyze each particular ESG pillar in detail. For that purpose, we compute the disclosure scores for Environmental, Social, and Governance categories separately based on corresponding sections of the ESG dictionary. To examine the relationship between the abnormal returns in the filing period and individual ESG categories, we modify the regression (7) as follows:

$$AbnReturn_i = a + b \times EnvDisc_i^{idf} + c \times SocDisc_i^{idf} + d \times GovDisc_i^{idf} + \epsilon_i \tag{11}$$

where $EnvDisc_i^{idf}$, $SocDisc_i^{idf}$, and $GovDisc_i^{idf}$ are the Environmental, Social, and Governance disclosure scores of 10-K report i based on *idf* word weights. Model 5 in Table 4 reports the results of the fitted regression. While the coefficient of the environmental score is statistically insignificant, the social and governance scores show the opposite result, with the social disclosure coefficient of 0.0414 significant at the 1% level, and 0.0302 significant at the 10% level of the governance category, respectively. However, the two categories show a high degree of correlation between them, thus raising the multicollinearity problem for the fitted regression model. To overcome this weakness, we run additional regressions that separate the social and governance variables, and in addition include further controls for the firm-specific effects. For that purpose, we modify and extend the regression (11) as follows:

$$\begin{aligned} AbnReturn_i = a &+ b \times EnvDisc_i^{idf} + c \times SocDisc_i^{idf} + d \times Size_i + e \times BM_i \\ &+ f \times Accruals_i + g \times Vola_i + h \times ShareTurn_i \\ &+ i \times EarnDayRet_i + \epsilon_i \end{aligned} \tag{12}$$

$$\begin{aligned} AbnReturn_i = a &+ b \times EnvDisc_i^{idf} + c \times GovDisc_i^{idf} + d \times Size_i + e \times BM_i \\ &+ f \times Accruals_i + g \times Vola_i + h \times ShareTurn_i \\ &+ i \times EarnDayRet_i + \epsilon_i \end{aligned} \tag{13}$$

Models 6 and 7 in Table 4 report the estimation results. While the environmental score remained insignificant, both social and governance scores show similar results to the model 5, with the coefficient of the governance disclosure increasing its significance to the 5% level. With regards to the control variables, both regression models yield coefficients and their significance levels akin to the regression models 2 and 4 in Table 4. Thus, the regression analysis shows significant positive relationship between the Social and Governance disclosure scores of 10-K reports and respective abnormal stock returns during the filing period.

3.4. Robustness analysis

To implement the robustness analysis of our results, we make two adjustments to the setting of our tests, which affect the sample size and control variables used in the regression estimations. First, we extend the time frame of the data set back to January 2013, thus considering both the training and the test data sets for the regressions, but focusing only on the disclosure scores with *idf* -based term weights.[35] Second, we include additional control for the actual ESG performance of companies in our regressions to account for differences in sustainability performance levels of different companies besides their affiliation to a particular industrial division. Furthermore, the current research finds differences between the "perceived" and "actual" ESG performance of companies, which can significantly differ due to opportunistic disclosure strategies of company's management and misinterpretation of disclosed information by analysts (Schulz (2017)).[36] For those reasons, we use Arabesque S-Ray@ ESG performance scores[37] that analyze and evaluate "all financially material environmental, social, and governance issues"[38] which affect companies on a daily basis, and thus report the actual sustainability performance of companies. The ESG performance evaluation includes both the overall and individual pillars assessment of companies, which allows us to include the performance controls for the general and supplementary regression models presented in the main part. Thus, we modify the regression eqs. (8), (12), and (13) as follows:

Table 5
Robustness analysis. This table reports the regression estimates of the filing period abnormal stock returns on ESG disclosure scores with various control variables, including the Arabesque S-Ray@ ESG performance scores, on the sample of 10-K reports filed in the period between January 2013 and December 2019. *ESGDisc*$_{idf}$, *EnvDisc*$_{idf}$, *SocDisc*$_{idf}$, and *GovDisc*$_{idf}$ are respectively the ESG, Environmental, Social, and Governance disclosure scores of 10-K reports based on *idf* word weights computed as defined in Eqs. (9) and (10). *ESGPerf*, *EnvPerf*, *SocPerf*, and *GovPerf* are respectively the ESG, Environmental, Social, and Governance Arabesque S-Ray@ performance scores of companies available on the 10-K report filing day. All OLS regression models are estimated with time and industry fixed effects as well as standard errors clustered on the firm level. The number of observations is limited due to restricted availability of control variables for each regression model. All independent variables are standardized to a mean of 0 and a standard deviation of 1. The coefficients' estimates are not affected by the presence of outliers in the control variables. "***" denotes the 1% significance level, "**" the 5%, and "*" the 10% level, respectively. The values in parentheses report the standard errors of estimated coefficients.

	Models		
	(1)	(2)	(3)
ESGDisc$_{idf}$	0.0156**		
	(0.007)		
ESGPerf	−0.0109**		
	(0.005)		
EnvDisc$_{idf}$		−0.0014	−0.0015
		(0.006)	(0.005)
SocDisc$_{idf}$		0.0190***	
		(0.006)	
GovDisc$_{idf}$			0.0012**
			(0.001)
EnvPerf		−0.0010	−0.0071*
		(0.005)	(0.004)
SocPerf		−0.0097	
		(0.006)	
GovPerf			−0.0045
			(0.004)
Size	−0.0294***	−0.0272***	−0.0267***
	(0.008)	(0.009)	(0.009)
BM	−2.4205**	−2.2182*	−2.3386**
	(1.129)	(1.142)	(1.147)
Accruals	−5.9942	−6.1754	−6.0225
	(3.796)	(3.776)	(3.742)
Vola	2.6428	2.6429	2.6557
	(2.107)	(2.102)	(2.102)
ShareTurn	0.0173	0.0169	0.0176
	(0.013)	(0.013)	(0.013)
EarnDayRet	0.5969***	0.5962***	0.5969***
	(0.039)	(0.039)	(0.039)
Observations	7743	7743	7743

[35] Since the original training set is also used for the regression analysis in this part and thus cannot be used for estimation of word-power weights needed for ESG sentiment scores.

[36] Serafeim and Yoon (2021) also find that the market response to ESG information depends on ESG ratings of companies.

[37] Arabesque S-Ray@ ESG performance scores are provided by Arabesque S-Ray GmbH, UK Branch under exclusivelicense agreement.

[38] Besides 250 sustainability metrics collected on the annual basis, Arabesque scans daily over 30 thousand news sources published in over 170 countries for sustainability related controversies and NGO campaign activities looking back one year on a rolling basis for each sustainability topic, which is used in a multi-faceted procedure to create a comprehensive ESG score available on a daily basis. See https://www.arabesque.com/s-ray/our-scores/ for further information.

$$AbnReturn_i = a + b \times ESGDisc^{idf} + c \times ESGPerf_i + d \times Size_i + e \times BM_i + f \times Accruals_i + g \times Vola_i + h \times ShareTurn_i + i \times EarnDayRet_i + \epsilon_i \tag{14}$$

$$AbnReturn_i = a + b \times EnvDisc_i^{idf} + c \times SocDisc_i^{idf} + d \times EnvPerf_i + e \times SocPerf_i + f \times Size_i + g \times BM_i + h \times Accruals_i + i \times Vola_i + g \times ShareTurn_i + k \times EarnDayRet_i + \epsilon_i \tag{15}$$

$$AbnReturn_i = a + b \times EnvDisc_i^{idf} + c \times GovDisc_i^{idf\,idf} + d \times EnvPerf_i + e \times GovPerf_i + f \times Size_i + g \times BM_i + h \times Accruals_i + i \times Vola_i + g \times ShareTurn_i + k \times EarnDayRet_i + \epsilon_i \tag{16}$$

where $ESGPerf_i$, $EnvPerf_i$, $SocPerf_i$, and $GovPerf_i$ are respectively the ESG, Environmental, Social, and Governance performance scores of company i on the 10-K report filing day. Table 5 reports the estimation results of regressions. Due to a limited availability of ESG performance scores for companies contained in our sample, the dataset used in regressions with performance scores includes 7743 observations. The results presented in model 1 confirm a significant positive relationship between the ESG disclosure score and abnormal stock returns, with coefficient of 0.0156 significant at the 5% level. The ESG performance score, in turn, has a negative coefficient estimate with 5% significance level. With regards to other control variables, size, book-to-market, and earnings announcement return show significant estimation coefficients, whereas accruals, volatility, and shares turnover demonstrate insignificant results. Similar picture can be seen in models 2 and 3 which analyze the impact of individual ESG pillars' disclosure on abnormal returns. Comparable to previous results in the main part, the social disclosure score has a positive coefficient significant at the 1% level in model 2, whereas the governance score also yields a positive estimated coefficient at the 5% level in model 3. The estimated coefficients for control variables do not significantly deviate from the results obtained in model 1. Overall, the robustness analysis supports our findings obtained in the main part, with the disclosure of Social- and Governance-related topics attracting particular attention of the market.

4. Discussion and limitations

The analysis of ESG lexicon by means of the word clouds shows a strong prevalence of Governance topics in 10-K reports, with particular attention of companies' management to narratives related to i.a. compliance, audit control, compensation and awards, and disclosure regulations. The most impactful ESG words for investors, in turn, also highlight subjects related to gender and social discrimination, financial crimes, communities and people's health. The regression models analysis based on the word weighting approach as defined in models (1) and (2) in Table 4 show a significant relationship between the abnormal market reaction and the ESG disclosure metric, which is also shown in the models (3) and (4) with *idf* -based term weights and thus regardless of the term weighting approach. Thus, the results imply that the general ESG information disclosure of public companies via 10-K reports serves as an object of attention for the investors and their investment decisions shown by the abnormal returns during the 10-K report disclosure date window. The regression analysis in models (5), (6), and (7) in Table 4, which focuses on the singular pillars of sustainability area, emphasizes the role of Social and Governance pillars in the overall positive effect of ESG disclosure on market valuation, underpinning the results of lexicon evaluation. Such outcome can be related to the focus of our sample on US companies and thus on the specifics of social and political agenda in the USA, which also affects the demands and expectations of investors. While the Environmental pillar does not stand in the spotlight of the agenda in contrast to European countries, the social and governance matters take centre stage in the United States, with social activism movements like MeToo and BlackLivesMatter, or Occupy Wall Street and increased governance controls after the global financial crisis shaping both political and economic landscapes. Our results also show significant positive reaction of markets to both the general ESG tone and the ESG disclosure extent and quality of 10-K reports, even after controlling for the firm-specific effects, including the control for earnings announcement abnormal returns. However, the sample with companies that belong to the Mining division shows the opposite effect, with negative reaction of investors to a more extensive ESG disclosure of companies. While the estimated coefficients presented in regressions (8) and (9) in Table 4 fail to reach statistically significant levels, this outcome should serve as a motivation for further studies on industries with bad ESG imprint and high litigation risk based on larger sample sizes. The robustness analysis, which includes additional controls for the actual ESG performance of companies, supports the results obtained in the main part, while also corroborating outcomes of other studies.[39] Both the general ESG disclosure in model (1) in Table 5 and Social and Governance disclosure scores in models (2) and (3) based on *idf* term weights show significant results, which could be examined further in potential future studies on a different sample with other time frames. Thus, our empirical results provide evidence for a significant effect of ESG sentiment on changes of US companies' market valuations around the 10-K report filing day, with particular role of Social and Governance narratives.

However, this study is also subject to several limitations. We constrain our sample to US companies, thus ignoring other regions and markets of the world. The inclusion of European companies, where the emphasis on environmental topics is much higher than in the USA, could provide a different picture regarding the role of the Environmental pillar in the overall effect of the ESG tone. Besides that, we approximate the tone of individual pillars with their information disclosure scores due to an unbalanced composition of the ESG dictionary and emphasis of 10-K reports' narratives on governance topics. Furthermore, we limit our assessment of the ESG tone to 10-K reports of companies while ignoring other sources of information such as 8-K reports,[40] stand-alone sustainability reports[41] prepared by companies and analysts' comments, news from media outlets and comments in social networks. Also, the sample with 10-K reports does not represent all industries in the same proportion, which complicates the application of general results and relationships to particular industries. In addition, the estimated word-power weights used in ESG tone scores are subject to sensitivity of calibration results depending on size and composition of the training sample. The use of a larger sample in future studies could reinforce the evidence obtained with sample and test settings applied in this paper.

5. Conclusion

We investigate the role of ESG sentiment in market capitalization changes of companies using a novel ESG dictionary and a market reactions-based word weighting scheme within a dictionary-based approach of content analysis. By focusing on more than seventeen thousand 10-K reports of US companies filed with SEC, we find a positive reaction of markets to the ESG tone of 10-K reports during a four-day

[39] See, e.g., Du & Yu, 2021, Serafeim & Yoon, 2021, Da Silva (2020), Lopez-de-Silanes et al. (2019).

[40] 8-K is "the current report companies must file with the SEC to announce major events that shareholders should know about", see https://www.sec.gov/fast-answers/answersform8khtm.html for further information.

[41] However, the use of stand-alone CSR reports only can lead to a sample selection bias, since companies are not obliged to issue such reports and publish them voluntarily, which makes this option attractive to companies with a very good ESG record (Du and Yu (2021)).

filing event window. In contrast to similar studies on this topic, we rely on a word weighting scheme that determines both the magnitude and the tone of each particular ESG word conveyed to market participants through 10-K reports. Thus, we also determine the most relevant ESG words in each particular sustainability category, while the analysis of ESG pillar disclosure scores emphasizes a particular significance of social- and governance-related topics for investors. Also, we observe a different reaction of markets to extensive ESG reporting of companies that belong to the environmentally unfriendly Mining sector with a high litigation risk, which requires further examination in future studies. Our robustness analysis, which includes controls for active sustainability performance, also shows the relevance of ESG disclosure in 10-K reports to investors beyond the sustainability rankings provided by rating institutions. The use of an alternative word weighting scheme for the ESG disclosure quantification also supports our initial results obtained with a market reactions-based approach, with the latter identifying the words inducing the most positive and negative reactions of the markets. While the ESG reporting rapidly gains on strategic importance and becomes indispensable for companies in all industries (Tschopp and Huefner (2015)), the results of this study shed light on the mechanics of ESG information perception, which may be of interest for both companies' management and regulatory agencies.[42] By closely understanding the demands and expectations of investors based on the research approach applied in this study, the management could mitigate the problem of asymmetric information with stakeholders[43] and reduce inefficiencies in information transmission, which can result in significant discrepancies between the impression of shareholders and the intended meaning of the management (Barbeito-Caamaño and Chalmeta (2020)).

This paper also offers several reference points for future research settings. Since this study focuses on the US companies and their 10-K reports, the follow-up inquiries could include other regions of the world, with a particular emphasis on European markets due to environmentally focused agenda of European countries, and other information sources such as 10-Q and 8-K reports, analysts' research notes, commentaries in social networks, and news from media outlets. Furthermore, the differences between investors from different regions in terms of their perception of ESG information could be examined, with focus on most impactful ESG words and n-grams. In-depth analysis of industries based on larger samples could provide further evidence on ESG reporting practices with differing sustainability imprints. In addition, the distinction between value relevant and financially immaterial ESG tones could provide further clarification of incentives and motivation of investors (Serafeim and Yoon (2021)), whereas the differentiation between retail and institutional investors could shed light on ESG perception of professional vs amateur market participants. Last but not least, the use of other statistical models from the natural language processing domain, such as Latent Dirichlet Allocation, could further examine the evidence obtained with the linguistic models implemented in this study.

Declaration of Competing Interest

None.

Appendix A. Sample statistics

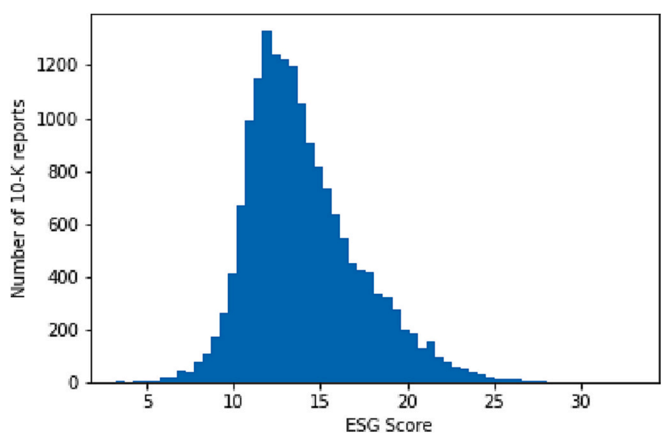

Fig. 4. Distribution of ESG disclosure score in the complete sample. This figure presents the frequency distribution of ESG disclosure score estimated using *idf* -based word weights. The graph depicts the frequency distribution of disclosure score in the complete dataset of 10- K reports that covers the period between January 2013 and December 2019 and includes 17001 observations.

Table 6

Summary statistics of the robustness dataset (2013–2019). This table presents the descriptive statistics for the robustness dataset with 10-K reports of US companies in period between January 2013 to December 2019 including the filing period abnormal stock returns of companies together with the control variables used in regressions. Abnormal return is defined as company's cumulated stock return over a four-day 10-K filing period window $(t-1, t+2)$ in excess of the market return represented by the CRSP value-weighted market index. ESG, Environmental, Social, and Governance $Disclosure_{idf}$ are respectively ESG, Environmental, Social, and Governance disclosure scores based on *idf* term weighting methodology. Earnings announcement return is defined as company's cumulated stock return over a three-day window around the earnings an- nouncement $(t-1, t+1)$ in excess of the market return represented by the CRSP value-weighted market index. Book-to-market is defined as the ratio of the book value of equity to the market value of equity at the end of the quarter before 10-K report filing. Size is defined as the natural logarithm of the market capitalization of company at the end of the quarter before 10-K report filing. Accruals are defined according to Sloan (1996) as one-quarter difference in current assets without cash subtracting change in current liabilities without change in debt included in current liabilities and change in income taxes payable minus

[42] The recent research shows that companies allocate substantial resources to improve their communication practicesin the ESG area, while the regulatory authorities study the reaction of investors to ESG disclosure of companies (Serafeim and Yoon (2021)).

[43] See also Du and Yu (2021).

depreciation and amortization expense. Volatility is defined as a 250-day rolling standard deviation of company's firm-specific (excess) stock return. Shares turnover is defined as the natural logarithm of the number of shares traded over 250 days before the 10-K report filing date divided by the number of shares outstanding on the filing date. ESG, Environmental, Social, and Governance performance are Arabesque S-Ray@ performance scores of a company in the overall ESG category as well as singular pillars available on the 10-K report's filing date. The extreme values of book-to-market, accruals, and volatility variables do not affect the coefficients and their significance in econometric models.

Variable	Observations	Mean	SD	Min.	25%	50%	75%	Max.
Total words	8097	32435	16341	3280	24446	30351	37870	785354
ESG words	8097	1705	754	157	1255	1563	1977	11944
Environmental words	8097	122	179	1	42	67	116	2654
Social words	8097	386	244	11	247	318	437	2476
Governance words	8097	1325	558	143	1001	1236	1540	8887
Abnormal return	8097	0.003	0.065	−0.660	−0.021	−0.000	0.022	1.216
ESG Disclosure$_{idf}$	8097	14.25	3.17	4.68	12.05	13.690	15.94	33.02
Environmental Disclosure$_{idf}$	8097	2.28	2.11	0.09	0.9	1.59	2.76	13.62
Social Disclosure$_{idf}$	8097	4.22	1.75	0.69	3.1	3.77	4.73	13.83
Governance Disclosure$_{idf}$	8097	8.02	1.23	3.47	7.23	7.88	8.7	17.49
Earnings announcement return	8097	0.004	0.078	−0.701	−0.035	0.002	0.039	0.994
Book-to-Market	7769	0.44	0.51	−14.53	0.2	0.38	0.623	8.643
Size	7775	8.45	1.47	2.1	7.460	8.33	9.360	13.54
Accruals	7780	−0.009	0.071	−0.751	−0.022	−0.007	0.003	3.813
Volatility	8069	0.325	1.446	0.085	0.185	0.240	0.338	86.410
Shares turnover	8069	−4.84	0.623	−7.971	−5.214	−4.862	−4.469	−1.331
ESG performance	8097	50.76	8.343	22.91	45.04	51.02	56.83	77.21
Environmental performance	8097	46.93	13.461	22.03	34.35	43.45	57.53	87.89
Social performance	8097	51.08	8.712	22.78	44.41	50.22	57.27	80.93
Governance performance	8097	51.81	12.870	11.33	43.41	52.69	61.18	86.69

Appendix B. ESG dictionary (Baier et al. (2020))

The following pages present the Environmental, Social, and Governance categories of the ESG dictionary proposed by Baier et al. (2020). The dictionary consists of 34 subcategories that include 482 words. The categories include: Biofuels, climate change strategy, emissions management and reporting, access to land, biodiversity management, water, pollution control, waste and recycling, access to medicine, HIV and AIDS, nutrition, product safety, community relations, privacy and free expression, security, diversity, health and safety, ILO core conventions, supply chain labor standards, charity, education, employment, audit and control, board structure, remuneration, shareholder rights, transparency, talent, bribery and corruption, political influence, whistle-blowing system, disclosure and reporting, stakeholder engagement, UNGC compliance.

Table 7
ESG dictionary: Environment.

Topic	Category	Subcategory
Environmental: clean, environmental, epa, sustainability		
	Climate change: Climate, warming	**Biofuels:** biofuels, biofuel
		Climate change strategy: green, renewable, solar, stewardship, wind
		Emissions management and reporting: emission, emissions, ghg, ghgs, greenhouse, atmosphere, emit
	Ecosystem service: agriculture, deforestation, pesticide, pesticides, wetlands	**Access to land:** zoning
		Biodiversity management: biodiversity, species, wilderness, wildlife
		Water: freshwater, groundwater, water
	Environmental management: cleaner, cleanup, coal, contamination, fossil, resource	**Pollution control:** air, carbon, nitrogen, pollution, superfund
		Waste and recycling: biphenyls, hazardous, householding, pollutants, printing, recycling, toxic, waste, wastes, weee, recycle

Table 8
ESG dictionary: Social.

Topic	Category	Subcategory
Social: citizen, citizens, csr, disabilities, disability, disabled, human, nations, social, un, veteran, veterans, vulnerable	**Public health:** children, epidemic, health, healthy, ill, illness, pandemic	**Access to medicine:** childbirth, drug, medicaid, medicare, medicine, medicines
		HIV and AIDS: hiv
		Nutrition: alcohol, drinking
		Product safety: bugs, conformance, defects, fda, inspection, inspections, minerals, standardization, warranty
	Human rights: dignity, discriminate, discriminated, discriminating, discrimination, equality, freedom, humanity, nondiscrimination, sexual	**Community relations:** communities, community
		Privacy and free expression: expression, marriage, privacy
		Security: peace
	Labor standards: bargaining, eeo, fairness, fla, harassment, injury, labor, overtime, ruggie, sick, wage, wages, workplace	**Diversity:** bisexual, diversity, ethnic, ethnically, ethnicities, ethnicity, female, females, gay, gays, gender, genders, homosexual, immigration, lesbian, lesbians, lgbt, minorities, minority, ms, race, racial, religion, religious, sex, transgender, woman, women
		Health and safety: occupational, safe, safely, safety
		ILO core conventions: ilo, labour
		Supply chain labor standards: eicc
	Society: endowment, endowments, people, philanthropic, philanthropy, socially, societal, society, welfare	**Charity:** charitable, charities, charity, donate, donated, donates, donating, donation, donations, donors, foundation, foundations, gift, gifts, nonprofit, poverty
		Education: courses, educate, educated, educates, educating, education, educational, learning, mentoring, scholarships, teach, teacher, teachers, teaching, training
		Employment: employ, employment, headcount, hire, hired, hires, hiring, staffing, unemployment

Table 9
ESG dictionary: Governance.

Topic	Category	Subcategory
Governance: align, aligned, aligning, alignment, aligns, bylaw, bylaws, charter, charters, culture, death, duly, parents, independent	**Corporate governance:** compliance, conduct, conformity, governance, misconduct, parachute, parachutes, perquisites, plane, planes, poison, retirement	**Audit and control:** approval, approvals, approve, approved, approves, approving, assess, assessed, assesses, assessing, assessment, assessments, audit, audited, auditing, auditor, auditors, audits, control, controls, coso, detect, detected, detecting, detection, evaluate, evaluated, evaluates, evaluating, evaluation, evaluations, examination, examinations, examine, examined, examines, examining, irs, oversee, overseeing, oversees, oversight, review, reviewed, reviewing, reviews, rotation, test, tested, testing, tests, treadway
		Board structure: backgrounds, independence, leadership, nomination, nominations, nominee, nominees, perspectives, qualifications, refreshment, skill, skills, succession, tenure, vacancies, vacancy
		Remuneration: appreciation, award, awarded, awarding, awards, bonus, bonuses, cd, compensate, compensated, compensates, compensating, compensation, eip, iso, isos, payout, payouts, pension, prsu, prsus, recoupment,

(*continued on next page*)

Table 9 (*continued*)

Topic	Category	Subcategory
		remuneration, reward, rewarding, rewards, rsu, rsus, salaries, salary, severance, vest, vested, vesting, vests

Topic	Category	Subcategory
		Shareholder rights: ballot, ballots, cast, consent, elect, elected, electing, election, elections, elects, nominate, nominated, plurality, proponent, proponents, proposal, proposals, proxies, quorum, vote, voted, votes, voting **Transparency:** brother, clicking, conflict, conflicts, family, grandchildren, grandparent, grandparents, inform, insider, insiders, inspector, inspectors, interlocks, nephews, nieces, posting, relatives, siblings, sister, son, spousal, spouse, spouses, stepchildren, stepparents, transparency, transparent, visit, visiting, visits, webpage, website **Talent:** attract, attracting, attracts, incentive, incentives, interview, interviews, motivate, motivated, motivates, motivating, motivation, recruit, recruiting, recruitment, retain, retainer, retainers, retaining, retention, talent, talented, talents
	Business ethics: cobc, ethic, ethical, ethically, ethics, honesty	**Bribery and corruption:** bribery, corrupt, corruption, crimes, embezzlement **Political influence:** grassroots, influence, influences, influencing, lobbied, lobbies, lobby, lobbying, lobbyist, lobbyists **Whistle-blowing system:** whistleblower

Topic	Category	Subcategory
	Sustainability management and reporting: announce, announced, announcement, announcements, announces, announcing, communicate, communicated, communicates, communicating, erm, fairly, integrity, liaison, presentation, presentations, sustainable	**Disclosure and reporting:** asc, disclose, disclosed, discloses, disclosing, disclosure, disclosures, fasb, gaap, objectivity, press, sarbanes **Stakeholder engagement:** engagement, engagements, feedback, hotline, investor, invite, invited, mail, mailed, mailing, mailings, notice, relations, stakeholder, stakeholders **UNGC compliance:** compact, ungc

References

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

Atenstaedt, R. (2012). Word cloud analysis of the BJGP. *British Journal of General Practice, 62*, 148.

Aureli, S., Gigli, S., Medei, R., & Supino, E. (2020). The value relevance of environmental, social, and governance disclosure: Evidence from Dow Jones sustainability world index listed companies. *Corporate Social Responsibility and Environmental Management, 27*, 43–52.

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

Barbeito-Caamano, A., & Chalmeta, R. (2020). Using big data to evaluate corporate social responsibility and sustainable development practices. *Corporate Social Responsibility and Environmental Management, 27*, 2831–2848.

Brinckmann, F., & Ender, M. (2019). Impact of CSR-relevant news on stock prices of companies listed in the Austrian traded index (ATX). *International Journal of Financial Studies, 7*, 1–18.

Brogi, M., & Lagasio, V. (2019). Environmental, social, and governance and company profitabil- ity: Are financial intermediaries different? *Corporate Social Responsibility and Environmental Management, 26*, 576–587.

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

Cheema-Fox, A., LaPerla, B. R., Wang, H., & Serafeim, G. (2021). Corporate resilience and response to COVID-19. *Journal of Applied Corporate Finance, 33*, 24–40.

Da Silva, P. P. (2020). Breadth of ESG disclosure and the information content of stock prices. In *CMVM, Portuguese Securities Commission Working Paper* (pp. 1–26).

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

Du, S., & Yu, K. (2021). Do corporate social responsibility reports convey value relevant information? Evidence from report readability and tone. *Journal of Business Ethics, 172*, 253–274.

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

English, D. M., & Schooley, D. K. (2014). The evolution of sustainability reporting. *The CPA Journal*, 26–35.

Eugster, F., & Wagner, A. F. (2021). Earning investor trust: The role of past earnings management. *Journal of Business Finance & Accounting, 48*, 269–307.

Gamble, G. O., Hsu, K., Kite, D., & Radtke, R. R. (1995). Environmental disclosures in annual reports and 10Ks: An examination. *Accounting Horizons, 9*, 34–54.

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

Griffin, P. A. (2003). Got information? Investor response to form 10-K and form 10-Q EDGAR filings. *Review of Accounting Studies, 8*, 433–460.

Henry, E., Jiang, X., & Rozario, A. (2021). *The evolution of environmental discourse: Evidence from conference calls* (pp. 1–27). Stevens Institute of Technology - School of Business Working Papers.

Ho, V. H. (2020). Disclosure overload? Lessons for risk disclosure & ESG reporting reform from the regulation S-K concept release. *Villanova Law Review, 65*, 67–132.

Hummel, K., Mittelbach-Hormanseder, S., Rammerstorfer, M., & Weinmayer, K. (2020). *Stock market reactions and CSR disclosure in the context of negative CSR events* (pp. 1–14). University of Zürich Working Papers.

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

Jin, Y. (2017). Development of word cloud generator software based on Python. *Procedia Engineering, 174*, 788–792.

Jose, A., & Lee, S.-M. (2007). Environmental reporting of global corporations: A content analysis based on web disclosures. *Journal of Business Ethics, 72*, 307–321.

Krüger, P. (2015). Corporate goodness and shareholder wealth. *Journal of Financial Economics, 115*, 304–329.

Landau, A., Rochell, J., Klein, C., & Zwergel, B. (2020). Integrated reporting of environmental, social, and governance and financial data: Does the market value integrated reports? *Business Strategy and the Environment, 29*, 1750–1763.

Li, F. (2010). The information content of forward-looking statements in corporate filings – A Naïve bayesian machine learning approach. *Journal of Accounting Research, 48*, 1049–1102.
Li, F. W., & Sun, C. (2022). Information acquisition and expected returns: Evidence from EDGAR search traffic. *Journal of Economic Dynamics and Control, 141*, 1–20.
Lopez-de-Silanes, F., McCahery, J. A., & Pudschedl, P. C. (2019). ESG performance and disclosure: A cross-country analysis. *European Corporate Governance Institute Law Working Paper, 481*(2019), 1–45.
Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10Ks. *The Journal of Finance, 66*, 35–65.
Loughran, T., & McDonald, B. (2020). Textual analysis in finance. *Annual Review of Financial Economics, 12*, 357–375.
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.
Meier, J.-H., Esmatyar, W., & Frost, R. (2018). The predictive power of the sentiment of financial reports. In *Proceedings of the 14th international conference on ICT in education, Research and Industrial Applications* (pp. 1–13). ICTERI.
Moss, A., Naughton, J. P., & Wang, C. (2020). *The irrelevance of ESG disclosure to retail investors: Evidence from Robinhood* (pp. 1–24). University of Virginia Darden School of Business Working Paper.
Nazari, J. A., Hrazdil, K., & Mahmoudian, F. (2017). Assessing social and environmental performance through narrative complexity in CSR reports. *Journal of Contemporary Accounting and Economics, 13*, 166–178.
Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies, 22*, 435–480.
Sabbaghi, O. (2022). The impact of news on the volatility of ESG firms. *Global Finance Journal, 51*, 1–9.
Schulz, J.-F. (2017). Does greater disclosure of information on CSR performance improve analysts' forecast accuracy? *Journal of Environmental Law and Policy, 40*, 134–160.
Serafeim, G., & Yoon, A. (2021). Stock price reactions to ESG news: The role of ESG ratings and disagreement. *Review of Accounting Studies*, 1–27. Forthcoming.
Serafeim, G., & Yoon, A. (2022). Which corporate ESG news does the market react to? *Financial Analysts Journal, 78*, 59–78.
Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review, 71*, 289–315.
Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance, 62*, 1139–1168.
Tremblay, M. C., Parra, C. M., & Castellanos, A. (2015). Corporate social responsibility reports: Understanding topics via text mining. In *Twenty-first Americas Conference on Information Systems Puerto Rico* (pp. 1–14). AMCIS.
Tschopp, D., & Huefner, R. J. (2015). Comparing the evolution of CSR reporting to that of financial reporting. *Journal of Business Ethics, 127*, 565–577.
Yen, M.-F., Shiu, Y.-M., & Wang, C.-F. (2019). Socially responsible investment returns and news: Evidence from Asia. *Corporate Social Responsibility and Environmental Management, 26*, 1565–1578.