Readability of ESG Topic-Specific Narratives in Sustainability Reports and Stock Price Crash Risk

Yun Hao
E-mail: peterhao.mg12@nycu.edu.tw
Department of Management Science, National Yang Ming Chiao Tung University, TAIWAN

Chia-Chen Chen
E-mail: a7945968@gmail.com
Department of Management Science, National Yang Ming Chiao Tung University, TAIWAN

Kai Wang
E-mail: maciek980707@gmail.com
Department of Management Science, National Yang Ming Chiao Tung University, TAIWAN

Yijie Tseng
E-mail: 081647@mail.fju.edu.tw
Department of Accounting, Fu Jen Catholic University, TAIWAN

Tsung-Kang Chen*
Email: vocterchen@nycu.edu.tw
Department of Management Science, National Yang Ming Chiao Tung University, TAIWAN
Center for Research in Econometric Theory and Applications, National Taiwan University, TAIWAN

* Corresponding author. Address: No. 1001, Dasyue Rd., East Dist., Hsinchu City 300, Taiwan (R.O.C.). Tel: +886-3-5712121 ext 57124.

Readability of ESG Topic-Specific Narratives in Sustainability Reports and Stock Price Crash Risk

Abstract

We explore the association between the readability of ESG (namely environmental, social, and governance) topic-specific narratives in sustainability reports and stock price crash risk using a sample of 3,777 sustainability reports issued by publicly listed companies in Taiwan from 2014 to 2022. Empirical results show that the readability of the E, S, and G topic-specific narratives in sustainability reports are all significantly and negatively related to stock price crash risk when controlling for firm characteristics variables, with the social (S) topic exhibiting the most substantial effect. In addition, we also find that the negative association is significantly enhanced when the tone of the texts is more positive or the firm's ESG performance is better. Finally, the findings remain robust even after addressing endogeneity issues and sample selection bias.

Keywords: ESG topic-specific narrative readability, Stock price crash risk, Sustainability report, Social topic

1. Introduction

This study aims to examine the association between the readability of diverse environmental, social, and governance topic-specific narratives within the sustainability report and the risk of stock price crashes. In contrast to conventional annual financial disclosures, a sustainability report intends to elucidate the non-financial attributes of the company, delineate specific initiatives and prospective practices concerning environmental, social, and corporate governance matters, and disclose the repercussions of climate change on the enterprise itself. Additionally, it provides stakeholders with supplementary information beyond mere financial statements while facilitating the assessment of the efficacy of corporate sustainability policy implementation. This document predominantly encompasses textual and non-quantifiable data (Dhaliwal et al., 2011). It constitutes a vital ongoing medium for sustained dialogue between enterprises and their stakeholders.

Consequently, the report has evolved to become, alongside the annual report, a significant document through which companies disclose their pertinent information, especially concerning non-financial aspects. Given the descriptive and non-financial characteristics of this category of information, textual attributes such as readability, keyword frequency, and tonal quality assume a crucial role in the comprehensive report. Among these attributes, readability is the principal characteristic scrutinized in this study. It signifies readers' difficulty comprehending and assimilating textual content (Harris and Hodges, 1995). Readability is considered an essential criterion for assessing the quality of reports or written discourse, and it serves as a communicative mechanism through which corporations disseminate information.

Moreover, the text's complex vocabulary further exacerbates the audience's reading challenges. An increased frequency of difficult words diminishes reading speed and overall readability. Consequently, this study employs the ratio of complex words to the total word count as a surrogate measure of textual readability.

Moreover, given that the Sustainability Report encompasses the triad of environmental (E), social (S), and corporate governance (G) dimensions, previous literature suggests that the performance metrics associated with these three domains of the the environment (E), social (S), and corporate governance (G) possess varying informational significance. For instance, Marhfor et al. (2017) established that social performance across the three ESG dimensions is intrinsically linked to both performance and pricing information; thus, in contrast to environmental and corporate governance metrics, social performance tends to mitigate external investment risk more effectively due to the asymmetry in the information available to stakeholders. Additionally, Dumitrescu and Zakriya (2021) demonstrated that social performance considerably influences the likelihood of stock price depreciation. Consequently, building upon the ESG lexicon developed by Baier et al. (2020), this investigation further delineates three

categories of textual excerpts on environmental, social, and corporate governance issues from the Sustainability Report to conduct a text analysis. This study also examines the readability of these three categories of text passages to assess their effect on the risk of stock price decline.[1]

The risk of stock price crashes has always been a focal concern in financial markets and among investors, as it can trigger panic and precipitate a sharp decline in market value, with significant disruptions to the overall economic system. Recent research on crash risk has followed the agency theory proposed by Jin and Myers (2006), which suggests that asymmetric information between corporations and stakeholders can lead to crashes, identifying a tendency among management to conceal bad news as a primary predictor of crash risk. Taiwan has experienced several significant stock price crashes, such as the 1997 Asian financial crisis and the 2007 subprime mortgage crisis, severely impacting Taiwan's economy. More recently, Pharmally International Holding Co. Ltd's stock price crash in 2018, primarily due to false statements in financial reports and public disclosures, allowed managers to manipulate stock prices. Therefore, understanding the causes of stock price crashes is crucial for investors and regulatory authorities, as it aids in developing more effective risk management strategies to mitigate the negative impacts of crashes. This study follows the bad news hoarding theory proposed by Basu (1997), Jin and Myers (2006), and Kothari et al. (2009) to infer the relationship between the textual characteristics of sustainability reports and the risk of stock price crashes. As bad news accumulates, its eventual release can cause a stock price crash. Previous research found that when company performance is poor, managers attempt to hide or obscure information by reducing the readability of text disclosures (Li, 2008); companies with poorer environmental performance manipulate environmental information disclosure through word choice (Cho et al., 2010). Therefore, higher readability and a lower proportion of difficult words in sustainability reports suggest a lower likelihood of management concealing bad news, enabling readers to quickly absorb and understand textual information. This speeds up information transmission, reduces information processing costs, and helps mitigate the risk of crashes.

Numerous studies have explored the relationship between corporate social responsibility (CSR) performance, financial performance, profitability, and the risk of stock price crashes. For example, Dumitrescu and Zakriya (2021) examined the impact of ESG performance, finding that only the social dimension of ESG performance significantly negatively correlated with future crash risk, indicating that good social performance can mitigate such risks and reduce the likelihood of stock price crashes.

[1] In this article, the term "category" refers to topic-specific narratives in sustainability reports that address Environmental (E), Social (S), and Governance (G) aspects. These narratives encapsulate key non-financial performance indicators, commitments, and initiatives corporations undertake to foster sustainable practices and transparent governance, making their readability vital for effective stakeholder communication.

However, few studies have investigated the influence of the textual characteristics of sustainability reports on stock price crash risk. Therefore, this study aims to analyze whether the textual features of sustainability reports can provide external investors with incremental information, thereby affecting the company's stock price crash risk, by using the proportion of difficult words as an inverse measure of readability based on the speed and accumulation of information transmission. Additionally, since past literature also shows that the environmental (E), social (S), and governance (G) dimensions of performance in sustainability reports have different informational values, this study further analyzes the readability of text sections on environmental, social, and governance aspects. This helps understand whether investors are more sensitive to textual information on specific elements of ESG and provides deeper insights into whether companies view sustainability reports as an effective communication tool.

This study utilizes sustainability reports published by listed companies in Taiwan from 2014 to 2022 as samples to analyze whether the readability of text sections in these reports' environmental, social, and governance (ESG) dimensions is associated with the risk of stock price crashes. The empirical results indicate a negative correlation between the readability of text sections in all three ESG dimensions and the risk of stock price crashes, with the social dimension having the most significant impact, thus confirming the main hypothesis of this study. Additionally, it was found that when sustainability reports have more positive semantics or when a company's ESG performance is better, the negative impact of ESG text section readability on stock price crash risk is significantly enhanced. Furthermore, the proportion of foreign ownership also significantly strengthens the negative relationship between readability in environmental and social categories and stock price crash risk. Moreover, this study employs the Heckman two-stage sample selection model to address concerns about sample selection bias, and the results indicate that the main findings remain robust after correcting for this bias. Additionally, the study uses instrumental variable regression models to address endogeneity issues, and the empirical results confirm that the conclusions of this study still hold.

The primary contributions of this study are as follows: Unlike previous research exploring the relationship between CSR performance and stock price crash risk (Dumitrescu and Zakriya, 2021), this study focuses on unstructured information in sustainability reports rather than quantitative data on CSR performance. It further distinguishes the text of sustainability reports into environmental, social, and governance dimensions, analyzing the differences in the relationship between the readability of texts in these dimensions and the risk of stock price crashes and understanding the speed and accumulation of information transmission in various types of sustainability reports. Additionally, this study proposes three potential moderating mechanisms in the

relationships mentioned above: the positive semantics of sustainability report texts, ESG performance, and the proportion of holdings by different types of institutional investors. Finally, this research provides recommendations for external investors on the readability, thematic content, emotional tone, and ESG performance of sustainability reports to help them avoid wealth losses caused by stock price crash risks in their future investment strategies.

2. Literature Reviews and Hypotheses Developments

2.1. Characteristics of Sustainability Reports

Sustainability reports lack a widely implemented reporting framework (Tschopp and Huefner, 2015), and there is no unified standard, making CSR reporting multidimensional rather than unidimensional (Kitzmueller and Shimshack, 2012). This complexity hinders horizontal comparisons with other companies, thus limiting the comparability of sustainability reports. Currently, the primary international standards include GRI, SASB, and TCFD, and companies can choose the reporting framework that benefits them the most. Among these, the GRI standards are the most widely communicated, intended for all stakeholders, and the most commonly adopted framework by companies in Taiwan. Even though companies use several indicators to describe their ESG performance, the informativeness of these metrics is not as robust as those in financial reports (Clarkson et al., 2008). This limitation is mainly because companies can disclose different quantitative indicators and decide on different reporting standards, leading to lower informativeness and comparability of these indicators.

Additionally, the content and method of disclosing CSR activities vary significantly among various industries and company sizes, making it difficult for readers to assess a company's ESG performance through comparative disclosure. Moreover, financial reports are compiled for investment purposes, targeting shareholders, financial analysts, and lending institutions. In contrast, sustainability reports are prepared to satisfy a broader range of stakeholders, including suppliers, customers, shareholders, and governments, providing them with relevant information for decision-making.

2.2. Textual Characteristics of Sustainability Reports

2.2.1. Readability

Readability refers to the ease with which readers can comprehend text information; higher readability indicates that the text is more manageable for readers to understand. Rennekamp (2012) found that investors respond more positively and with greater trust to highly readable information, while they react less strongly to poorly readable information, demonstrating that readability indicates whether information can be effectively transmitted. Companies exhibit different readability levels in their

information communication strategies, and investor decision-making varies accordingly (Miller, 2010), suggesting that management can manipulate readability. For instance, good news is typically disclosed quickly and clearly; companies often obscure bad news by reducing the amount of information disclosed or lowering its readability, thereby maintaining information asymmetry and minimizing adverse impacts on the company. This manipulation of readability is a tool for companies to obfuscate information in narrative disclosures. When readability is high, readers do not need to expend much time or effort to understand the essential contents and messages of a report, which not only speeds up the dissemination of information but also enhances the efficiency of information transmission. Thus, avoiding overly technical jargon and complex sentence structures can enhance text readability and the effectiveness of information transmission. However, low readability means that non-professional readers may be unable to effectively use these texts, reducing the effectiveness of the report's dissemination.

Previous literature has explored the readability of documents such as annual reports, key audit matters, corporate social responsibility reports, and Management Discussion and Analysis (MD&A). Li (2008) was the first to use the Fog Index to measure the readability of annual reports and to explore the relationship between report readability and corporate performance, linking the textual characteristics of annual reports with operational performance. This study set the precedent for subsequent scholars to use the Fog Index to measure the readability of annual reports. Li's research found that companies with lower earnings tend to have annual reports that are more difficult to read, suggesting that management attempts to increase the complexity of annual reports, preferring to use complex words to hide or confuse information from investors, thus mitigating the negative impacts faced by the company.

2.2.2. Keywords, Dictionaries, and Sentiment Tone

Keyword analysis in texts relies on predefined word lists, each word associated with specific attributes, allowing researchers to compute the number of words related to each attribute for subsequent analysis. Given the growing interest in ESG topics, there is a notable lack of specialized word lists for in-depth exploration. To address this, Baier et al. (2020) created an ESG dictionary using the 10-K reports of the top 25 companies by market value in the S&P 100 index, which researchers like Ignatov (2023) have since used to examine the relationship between disclosures in the Environmental (E), Social (S), and Governance (G) categories and stock returns. This study found that keywords related to community and health have a strong positive relationship with stock returns, highlighting the importance of ESG-focused lexical resources in financial analysis.

Emotion and tone are crucial features of a report’s narrative style, capturing whether the text is positive, negative, or neutral and providing insight into the management’s views on the company’s current or future outlook. Fisher et al. (2019)

indicated that tone can convey valuable information and can be strategically deployed to influence the reader's perception. For instance, when a company performs well, it may use more positive words to highlight its strengths. The emotional tone is a significant characteristic of text and can be manipulated by company management. Sentiment analysis in texts commonly employs the dictionary method.

In the financial and accounting fields, the LM dictionary (Loughran and McDonald, 2011) is the first dictionary designed explicitly for financial texts tailored to the language used in business disclosures. It measures text sentiment through lists of positive and negative words, containing 354 positive and 2,355 negative words. Before the creation of the LM dictionary, many studies used the Harvard Psychosociological Dictionary to assess the emotional tone of business texts. However, as the Harvard dictionary was based on sociological and psychological fields, Loughran and McDonald noted that some negative words were misclassified in financial contexts, such as 'tax' and 'liabilities,' which are not inherently negative in financial disclosures. This highlights the importance of using domain-specific dictionaries for accurate text analysis.

2.3. Value Relevance of Corporate Social Responsibility Disclosure

Cahan et al. (2016) indicate that the disclosure of corporate social responsibility (CSR) information is value-relevant in capital markets and serves as a critical source of information for investors in making investment decisions. This suggests that investors fully utilize such information to assess a company's profitability and predict future returns. Merkl-Davies and Brennan (2017) apply communication theory to explore CSR reporting, arguing that CSR reports are a form of corporate communication. By publishing these reports, companies provide CSR information to external stakeholders. If the disclosed information is comprehensible to stakeholders, it constitutes effective CSR communication.

Additionally, previous studies have found that the publication of CSR reports often significantly impacts stock markets (Du et al., 2017), attracting investor attention and eliciting market reactions. Du and Yu (2021) further discover that CSR reports with higher readability are associated with higher cumulative abnormal trading volumes and returns. This finding indicates that highly readable CSR reports enhance information transparency and are perceived by investors as reliable signals, stimulating market trading activity. Conversely, CSR reports with lower readability are associated with weaker stock price responses, possibly suggesting that negative information is obscured.

2.4. Stock Price Crash Risk

Stock price crashes undermine investors' trust in a company. Chen et al. (2001) define crash risk as the conditional skewness of return distributions, which represents the

third moment of return distribution. It is considered a proxy for measuring negative return skewness rather than the probability of extreme negative returns. It is regarded as a significant feature of return distributions alongside the mean.

Recent empirical studies on the determinants of crash risk have followed the agency theory framework proposed by Jin and Myers (2006). This framework explains crash phenomena through information asymmetry between managers, who are in a position of informational advantage (information providers), and external stakeholders, who are at an informational disadvantage (information recipients). Existing literature has linked managerial concealment of bad news to stock price crash risk, suggesting that crashes stem from self-serving behavior by internal management in hiding adverse information (Hutton et al., 2009).

2.4.1. Bad News Hoarding Theory

Management's accumulation of bad news primarily stems from the fact that managerial decisions regarding narrative disclosures are not neutral (Bowen et al., 2005). Managers have a certain degree of discretion over the timing of insider information disclosures (Verrecchia, 2001). To protect their interests, managers may strategically conceal bad news and control its flow to the public, allowing the accumulation of negative information to avoid jeopardizing their career prospects or equity wealth due to the release of such news.

Basu (1997) points out that managers typically possess internal information about the company's operations and asset values. When their compensation is tied to profitability, they will likely withhold information that could negatively impact earnings to safeguard their remuneration. Additionally, Kothari et al. (2009) suggest that managers may deliberately suppress bad news while accelerating the release of good news to achieve both long-term and short-term benefits.

However, when the accumulated bad news reaches a critical point where management can no longer conceal it or is unwilling to continue suppressing it, it will be released to the market all at once. This sudden release triggers a crash in market prices, indicating that the ongoing accumulation of bad news amplifies the risk of stock price crashes. Moreover, Mermod and Idowu (2014) find that when disclosing corporate social responsibility (CSR) information, companies often highlight more positive news to enhance their CSR reputation rather than revealing more negative information.

2.5. Hypotheses Developments

2.5.1. Relationship Between Readability of Individual E, S, and G Textual Paragraphs in Sustainability Reports and Stock Price Crash Risk

Previous studies have confirmed that ESG information disclosure has predictive power for crash risk and is value-relevant, indicating that investors incorporate corporate social responsibility (CSR) information into their investment decisions. This makes ESG information an essential tool for risk assessment. Based on the aforementioned theory of bad news accumulation and the value relevance of ESG, this study proposes the following inference: when the readability of sustainability reports is higher and the proportion of complicated words is lower, stakeholders can more quickly comprehend the key content of the text. This accelerates the speed of information transmission and reduces the accumulation of bad news, lowering the likelihood of future stock price crashes.

Conversely, when sustainability reports have lower readability and a higher proportion of difficult words, readers face increased challenges in reading and understanding the content. This slows down the transmission of information or renders it ineffective, thereby increasing the accumulation of bad news and raising the likelihood of future crashes.

Accordingly, this study hypothesizes that the readability of textual sections in sustainability reports related to environmental, social, and governance (ESG) categories is negatively associated with crash risk. Hypothesis 1 is formulated as follows:

Hypothesis 1. The readability of textual sections related to ESG categories in sustainability reports is negatively associated with stock price crash risk.

2.5.2. Importance of the Social Dimension

The study by Marhfor et al. (2017) finds a direct correlation between the social dimension of ESG and price informativeness, implying that the market and investors find it easier to evaluate the costs and benefits of the social dimension compared to the other two dimensions. This, in turn, better reduces the information asymmetry faced by investors. Dumitrescu and Zakriya (2021) explore the relationship between corporate social responsibility (CSR) performance and crash risk, analyzing overall ESG performance and the individual effects of the E, S, and G dimensions on crash risk. Their findings reveal that only the performance in the social dimension significantly affects the likelihood of a crash, suggesting that the social dimension primarily drives the relationship between overall ESG performance and crash risk.

Based on these findings, this study posits that the relationship between the social dimension and stock price crash risk is more potent than the other two dimensions. It further hypothesizes that the social category in sustainability reports holds more significant value and relevance. Hypothesis 2 is formulated as follows:

Hypothesis 2*. The readability of textual sections related to the social category in sustainability reports has the most substantial effect on stock price crash risk among the three ESG dimensions.*

2.5.3. Moderating Effect of Positive Sentiment

Managers communicate messages through positive or negative words (Davis et al., 2015), reflecting their outlook on the company's future performance while potentially influencing or manipulating market participants' expectations. Du and Yu (2021) further investigate the impact of tone on investor behavior, finding that an increase in the proportion of positive words in CSR reports leads to higher cumulative abnormal returns in the market. This indicates that the market reacts to an improvement in tone.

Accordingly, this study hypothesizes that as the tone of textual sections becomes more positive, the use of positive words will be perceived by investors as favorable and reliable signals. As a result, the market and external investors may pay less attention to or struggle to discern the information hidden behind difficult words. This may cause investors to overlook the company's risks, amplifying the accumulation of bad news and increasing the potential risk of future stock price crashes. This would, in turn, strengthen the negative relationship between the readability of ESG textual sections in sustainability reports and stock price crash risk. Based on the above, this study formulates Hypothesis 3 as follows:

Hypothesis 3*. Positive sentiment in textual sections related to ESG categories strengthens the negative relationship between the readability of ESG textual sections in sustainability reports and stock price crash risk.*

2.5.4. Moderating Effect of Institutional Ownership Proportion

Institutional investors differ from retail investors because they manage large-scale capital and can influence a company's operations and strategic direction, playing a crucial role in financial markets. The literature presents varying perspectives on the corporate governance effects of institutional investors. Pound (1992) proposed two hypotheses regarding the role of institutional investors: the Effective Monitor Hypothesis and the Interest Collusion Hypothesis.

Under the Effective Monitor Hypothesis, institutional investors are regarded as effective monitors that enhance corporate performance (Nagel et al., 2015). They exert a positive supervisory effect, improve information flow, and reduce information asymmetry. Stringent monitoring makes it more difficult for management to conceal

unfavorable information about the company (Andreou et al., 2016), thereby reducing the likelihood of bad news accumulation. Consequently, in scenarios where institutional ownership is higher, institutional investors are expected to play a supervisory role, decreasing management's incentives for manipulation. This reduces the accumulation of bad news, thereby lowering the importance of readability in sustainability reports and weakening the negative relationship between the readability of sustainability reports and crash risk.

However, under the Interest Collusion Hypothesis, institutional investors and management may collude by prioritizing their mutual interests over those of minority shareholders (Li et al., 2022). Buchanan et al. (2018) also find that institutional investors may interfere with the disclosure of corporate information, thereby increasing information asymmetry between the company and its stakeholders. In cases of high institutional ownership, institutional investors may collude with management to serve their shared interests, failing to perform their supervisory role. This increases management's incentives for manipulation, raising the likelihood of bad news accumulation. As a result, the importance of readability in sustainability reports becomes more pronounced, strengthening the negative relationship between the readability of sustainability reports and crash risk. Based on these perspectives, this study formulates the following hypotheses:

Hypothesis 4a. *Higher institutional ownership weakens the negative relationship between the readability of ESG textual sections in sustainability reports and stock price crash risk.*

Hypothesis 4b. *Higher institutional ownership strengthens the negative relationship between the readability of ESG textual sections in sustainability reports and stock price crash risk.*

2.5.5. Moderating Effect of ESG Performance

Clarkson et al. (2019) find that companies with strong corporate social responsibility (CSR) performance publish significantly more CSR reports than those with poor CSR performance. Furthermore, companies with superior CSR performance tend to release reports containing more extensive information (Muslu et al., 2014). This suggests that firms with excellent ESG performance are more proactive in publishing sustainability reports, with a higher volume of disclosed ESG information.

Based on this, the study posits that when ESG performance is strong, external investors and the market tend to exhibit higher trust in the company. This reduces their attention to the presence of difficult words in the text and lowers their sensitivity to

unfavorable information. Consequently, this amplifies the effect of bad news accumulation, strengthening the negative relationship between the readability of ESG textual sections in sustainability reports and stock price crash risk. Accordingly, Hypothesis 5 is formulated as follows.

Hypothesis 5. *Better ESG performance strengthens the negative relationship between the readability of textual sections in sustainability reports related to ESG categories and stock price crash risk.*

3. Data and Methodology

3.1. Data Sources and Sample Selection

The sample used in this study is derived from sustainability reports published on the Taiwan Stock Exchange Market Observation Post System, encompassing listed and over-the-counter (OTC) companies in Taiwan. In November 2014, Taiwan introduced a new regulation requiring companies to meet specific criteria or belong to certain industries to prepare CSR reports. Therefore, this study utilizes sustainability reports published from 2014 onward, covering the sample period from 2014 to 2022.

The downloaded sustainability reports, originally in PDF format, were converted to enable text analysis. After excluding reports with failed conversions, those missing financial variables, and financial industry firms, the final sample comprises 3,777 sustainability reports. Data on stock prices, corporate social responsibility performance, and institutional ownership ratios were obtained from the Taiwan Economic Journal (TEJ) database.

As shown in Table 1, the sample distribution indicates a steadily increasing trend in the number of sustainability reports published from 2014 to 2022.

[Insert Table 1 here]

3.2. Data Sources and Sample Selection

3.2.1. Independent Variables

The independent variable in this study is the readability of sustainability reports, measured by the ratio of complex words (DifficultWords_Ratio). The analysis utilizes the Chinese Readability Index Explorer (CRIE), an automated readability analysis system developed by the readability research team at the National Taiwan Normal

University (Sung et al., 2013; Sung et al., 2016). This system defines difficult words based on the frequency statistics from Academia Sinica, where the top 3,000 most frequent words are classified as common. Difficult words are the total number of words not included in this common word list. Previous research has demonstrated the validity of using word frequency to assess readability (Milone, 2009). To account for potential variations in text size across companies, this study uses the proportion of difficult words as a proxy variable for readability.

The textual sections for environmental (E), social (S), and governance (G) categories were extracted using the ESG dictionary developed by Baier et al. (2020). This dictionary contains 491 words, with 64, 151, and 276 words categorized under environment, social, and governance, respectively. Each word in the ESG dictionary was translated into Chinese one-to-many. Subsequently, the text sections containing these keywords were identified, with the paragraph start point defined as the first punctuation mark before the keyword and the endpoint as the first period after the keyword. This process generated separate textual sections for E, S, and G, enabling the calculation of the ratio of difficult words for each category.

The ratio of difficult words for each category was computed as the number of difficult words in the category-specific textual sections divided by the total number of words in those sections. These ratios were denoted as E_DifficultWords_Ratio, S_DifficultWords_Ratio, and G_DifficultWords_Ratio, representing the environment, social, and governance categories, respectively.

3.2.2. Dependent Variables

The dependent variable in this study is **stock price crash risk**, measured using two indicators proposed by Chen et al. (2001) and Hutton et al. (2009). The calculation begins by regressing individual weekly returns of firms against market returns to obtain the residuals (εi,t\varepsilon_{i,t}εi,t) based on the regression model shown in Equation (1).

In this model:

- $r_{i,t}$ represents the unadjusted weekly return for the firm i in week t,
- $r_{m,t}$ denotes the market return of the weighted stock index in week t,
- The regression includes two lead and two lag terms for market returns to account for temporal effects.

The specific weekly return for the firm i in week t ($W_{i,t}$) s is then calculated by adding 1 to the residual ($\varepsilon_{i,t}$) and taking the natural logarithm, as detailed in Equation

(2). This approach ensures that the crash risk indicators reflect firm-specific factors rather than broader market fluctuations. The detailed formulas are as follows:

$$r_{i,t} = \alpha_i + \beta_{1,i} r_{m,t-2} + \beta_{2,i} r_{m,t-1} + \beta_{3,i} r_{m,t} + \beta_{4,i} r_{m,t+1} + \beta_{5,i} r_{m,t+2} + \varepsilon_{i,t} \quad (1)$$

$$W_{i,t} = \ln(1 + \varepsilon_{i,t}) \quad (2)$$

The first crash risk measure employed in this study is the Negative Conditional Skewness (NCSKEW) proposed by Chen et al. (2001). This indicator captures the asymmetry in the distribution of returns, where a more enormous NCSKEW value indicates a higher crash risk. The calculation of NCSKEW is detailed in Equation (3):

$$NCSKEW_{i,w} = -\left[n(n-1)^{\frac{3}{2}} \sum W_{i,t}^3\right] / \left[(n-1)(n-2)\left(\sum W_{i,t}^2\right)^{\frac{3}{2}}\right] \quad (3)$$

Where:

- $NCSKEW_{i,w}$ represents the negative conditional skewness for the firm i in the year w.
- $W_{i,t}$ is the firm-specific weekly return for the firm i in week t, calculated as per Equation (2).
- n is the number of weekly returns available for the firm i in the year w.

This measure quantifies the degree of negative skewness in a firm's return distribution, reflecting the extent to which extreme negative returns dominate, thereby indicating higher stock price crash risk.

The second crash risk measure is **Down-to-Up Volatility (DUVOL)**, as proposed by Chen et al. (2001). This measure compares firm-specific weekly returns to the average weekly return for the year. It divides the returns into two groups: **up weeks** (returns above the average) and **down weeks** (returns below the average). A higher DUVOL value indicates a greater crash risk. The calculation is described in Equation (4):

$$DUVOL_{i,w} = \log\left\{(n_u - 1) \sum_{Down} W_{i,t}^2 / (n_d - 1) \sum_{Up} W_{i,t}^2\right\} \quad (4)$$

Where:

- $DUVOL_{i,w}$ is the down-to-up volatility for the firm i in the year w .

- $W_{i,t}$ is represents the firm-specific weekly return for the firm i in week t, as calculated in Equation (2).
- n_u is the number of up weeks in the year w, and n_d is the number of down weeks in the year w.

3.2.3. Moderating Variables

3.2.3.1. Positive Sentiment

The sentiment of the text in this study is measured using the Augmented NTU Sentiment Dictionary (ANTUSD), developed by Academia Sinica. This is currently the most prominent Chinese sentiment lexicon, comprising 28,799 words. Each word is assigned a corresponding sentiment score ranging from -1 to 1, where -1 indicates the most negative sentiment and 1 indicates the most positive sentiment. The sentiment score of a text is calculated by summing the sentiment scores of all words found in the lexicon and dividing by the total number of such words in the text. The detailed calculation is shown in Equation (5):

$$Sentiment\_Score = \frac{\sum Sentiment\ Scores\ of\ Words\ in\ Text}{Total\ Number\ of\ Words\ in\ Text\ Found\ in\ Lexicon} \quad (5)$$

The result of Equation (5) represents the sentiment score of the text. If the sentiment score (Sentiment_Score Sentiment_Score) is greater than 0, the text is classified as positive; if the score is less than 0, the text is classified as negative.

However, in this study, only one sample from the ESG textual sections of sustainability reports is classified as negative, with all other samples being positive. To address this, the ESG textual sections are divided into two groups based on their sentiment scores:

- Texts with sentiment scores in the top 50th percentile are classified as having stronger positive sentiment, with SentimentPos marked as 1.
- Texts with sentiment scores in the bottom 50th percentile are classified as having weaker positive sentiment, with SentimentPos marked as 0.

3.2.3.2. Institutional Ownership Proportion

This study utilizes the **TEJ Ownership Database** to classify institutional ownership into three categories based on the shareholding ratios of different institutional investors:

1. Foreign Investors' Ownership Ratio ($Foreign\_Investors$)
2. Investment Trust Ownership Ratio ($Investment\_Trust$)
3. Dealers' Ownership Ratio ($Dealer$)

These classifications provide a detailed breakdown of institutional ownership, allowing for a more nuanced analysis of the effects of different types of institutional investors on the variables of interest.

3.2.3.3. ESG Performance

The ESG scores used in this study are sourced from the TESG Sustainability Index provided by the TEJ database. These scores cover publicly listed, over-the-counter (OTC), emerging, and publicly traded companies in Taiwan. The scoring system evaluates three significant categories—environmental, social, and governance—across 16 themes, ultimately producing a quantitative ESG score ranging from 0 to 100, where 100 represents the best performance and zero the worst.

To classify ESG performance, this study uses the annual median ESG score to divide the sample into two groups: **high ESG performance** and **low ESG performance**. If a company's ESG score is in the top 50th percentile, ESG_Lead is assigned a value of 1. If the score is in the bottom 50th percentile, ESG_Lead is assigned a value of 0.

3.2.4. Control Variables

This study adopts control variables based on the crash risk model established by Kim et al. (2011). The control variables are defined as follows:

- **Company Size (SIZE):** The natural logarithm of the company's market capitalization.
- **Return on Assets (ROA):** The net income ratio after tax to total assets.
- **Leverage (LEV):** The ratio of total liabilities to total assets.
- **Market-to-Book Ratio (MB):** The natural logarithm of the ratio of market value to book value of the company.
- **Average Weekly Return (RET):** The arithmetic mean of firm-specific weekly returns for the company i in year t.
- **Standard Deviation of Weekly Returns (SIGMA):** The standard deviation of firm-specific weekly returns for company iii in year t.
- **Change in Turnover (DTURN):** The difference between the average monthly stock turnover in year t and year t-1.
- **Absolute Value of Discretionary Accruals (ABACC):** The absolute value of discretionary accruals, used to control for earnings management, is calculated using the method proposed by Dechow et al. (1995).

To calculate discretionary, the following steps are performed:

$$\frac{TACC_{i,t}}{TA_{i,t-1}} = \alpha \frac{1}{TA_{i,t-1}} + \beta_1 \frac{\Delta SALE_{i,t}}{TA_{i,t-1}} + \beta_2 \frac{PPE_{i,t}}{TA_{i,t-1}} + \varepsilon_{i,t} \quad (6)$$

Where

- $TA_{i,t-1}$: Total assets at the end of the previous year.
- $TACC_{i,t}$: Total accruals for the company.
- $\Delta SALE_{i,t}$: Change in net sales revenue.
- $\text{PPE}_{i,t}$: Property, plant, and equipment of the company.
- $\Delta REC_{i,t}$: Change in accounts receivable.
-

$$DISACC_{i,t} = \frac{TACC_{i,t}}{TA_{i,t-1}} - \hat{\alpha} \frac{1}{TA_{i,t-1}} - \widehat{\beta_1} \frac{\Delta SALE_{i,t} - \Delta REC_{i,t}}{TA_{i,t-1}} + \widehat{\beta_2} \frac{PPE_{i,t}}{TA_{i,t-1}} \quad (7)$$

The absolute value $DISACC_{i,t}$ is taken as ABACC, representing the extent of discretionary accruals. This comprehensive set of control variables ensures that factors influencing crash risk beyond the study's main variables are accounted for.

3.3. Model Specification

Based on the previously defined variables, this study follows the crash risk regression model proposed by Kim et al. (2011), setting the independent and control variables to precede the dependent variable by one year. Since the dataset is structured as **panel data**, the study employs a **panel regression model** to test the hypotheses. This model incorporates **industry and year-fixed effects** and adjusts for heteroskedasticity at the firm level using the approach recommended by Petersen (2009).

The regression model is designed to examine the relationship between the readability of ESG textual sections in sustainability reports and stock price crash risk. The independent variable, **DifficultWords_Ratio**, represents the proportion of difficult words in textual sections for each ESG category:

- $E\_DifficultWords\_Ratio$: Difficult word ratio in environmental sections.
- $S\_DifficultWords\_Ratio$: Difficult word ratio in social sections.
- $G\_DifficultWords\_Ratio$: Difficult word ratio in governance sections.

Stock price crash risk, the dependent variable, is measured using two metrics: **NCSKEW** and **DUVOL**. Separate regression analyses are conducted for each metric.

The structure of the regression model used for all subsequent analyses, including the investigation of readability and crash risk, remains consistent with this framework. This approach ensures robustness and comparability across different analyses.

$$Crash_{i,t} = \alpha + \beta_1 DifficultWords\_Ratio_{i,t-1} + \beta_2 SIZE_{i,t-1} + \beta_3 ROA_{i,t-1} + \beta_4 LEV_{i,t-1} + \beta_5 MB_{i,t-1} + \beta_6 RET_{i,t-1} + \beta_7 SIGMA_{i,t-1} + \beta_8 DTURN_{i,t-1} + \beta_9 ABACC_{i,t-1} + Year\ \ Effect + Industry\ \ Effect + \varepsilon_{i,t} \quad (8)$$

For the model design to validate potential mechanisms, the moderating models for **positive sentiment** and **ESG performance** follow the same structure as Equation (8). Separate regressions are conducted on grouped samples to examine whether the moderating effects exist.

The regression model for testing the moderating effect of **institutional ownership ratios** is shown in Equation (9). This model II_RATIO represents the institutional ownership ratios, which include:

- **Foreign Investors** ($Foreign\_Investors$),
- **Investment Trust** ($Investment\_Trust$)),
- **Dealer** ($Dealer$).

These ownership ratios are tested individually to assess the moderating effects of each type of institutional investor on the negative relationship between the readability of ESG textual sections and stock price crash risk. This model enables a detailed examination of how different institutional investors influence readability and crash risk dynamics across ESG categories.

$$Crash_{i,t} = \alpha + \beta_1 DifficultWords\_Ratio_{i,t-1} + \beta_2 II\_RATIO_{i,t-1} + \beta_3 DifficultWords\_Ratio_{i,t-1} * II\_RATIO_{i,t-1} + \beta_4 SIZE_{i,t-1} + \beta_5 ROA_{i,t-1} + \beta_6 LEV_{i,t-1} + \beta_7 MB_{i,t-1} + \beta_8 RET_{i,t-1} + \beta_9 SIGMA_{i,t-1} + \beta_{10} DTURN_{i,t-1} + \beta_{11} ABACC_{i,t-1} + Year\ \ Effect + Industry\ \ Effect + \varepsilon_{i,t} \quad (9)$$

4. Empirical Analysis

4.1. Descriptive Statistics

Table 2 presents the descriptive statistical analysis of all variables used in this study's regression, including stock price crash risk, the proportion of difficult words in individual ESG textual sections, control variables, and moderating variables. The two

measures of crash risk are **NCSKEW** and **DUVOL**, with higher values indicating greater crash risk. The average values are -0.2884 and 0.0890, respectively, suggesting that the overall sample exhibits a moderately low likelihood of a stock price crash.

The independent variables are the proportions of difficult words in the environmental, social, and governance textual sections, with mean values of 0.3198, 0.3117, and 0.3094 and standard deviations of 0.0265, 0.0246, and 0.0266, respectively. This indicates that, on average, the environmental category has the highest proportion of difficult words. In contrast, the governance category shows the most enormous variability in the proportion of difficult words across companies.

For the moderating variables, the sentiment scores for the environmental (E), social (S), and governance (G) textual sections all have positive mean values, indicating that the tone of these sections in the sustainability reports is generally positive, with minimal use of negative words. A closer look reveals that the minimum sentiment scores for the social and governance categories are 0.0728 and 0.0584, respectively, demonstrating that all textual sections in the sample are positive.

The institutional ownership ratio (**II_RATIO**) ranges from a minimum of 0.0079 to a maximum of 0.9939, indicating significant company variability. The mean ESG performance score is 61.2475, with a standard deviation 7.4072.

Regarding the control variables, the average company size (**SIZE**) is 9.4644, suggesting that the average market value of companies in the sample is approximately NT$12.9 billion. The standard deviation of return on assets (**ROA**) is 7.7290, indicating substantial variability in the profitability of the companies in the sample. The mean values for leverage ratio (**LEV**), market-to-book ratio (**MB**), and the absolute value of discretionary accruals (**ABACC**) are 0.4415, 2.1926, and 0.0765, respectively.

[Insert Table 2 here]

Table 3 presents the Pearson correlation coefficients for all variables used in the regression analysis. It shows that the proportion of difficult words in individual ESG textual sections is positively correlated with both measures of crash risk (**NCSKEW** and **DUVOL**), indicating that a higher proportion of complicated words is associated with greater crash risk. Most of the correlation coefficients for other research variables are below 0.3, suggesting minimal concerns regarding multicollinearity.

[Insert Table 3 here]

4.2. Regression Analysis

This study's sample data is panel data comprising information from multiple companies across several years. Panel regression analyses were conducted to examine the impact of the proportion of difficult words in the environmental, social, and governance textual sections on stock price crash risk while controlling for fixed year and industry effects and adjusting for heteroscedasticity at the firm level. The empirical model is specified as Equation (8).

Table 4 presents the empirical results on the relationship between the readability of individual ESG textual sections in sustainability reports and stock price crash risk. Columns (1) to (3) show that when crash risk is measured by NCSKEW, the proportions of difficult words in the social (S_DifficultWords_Ratio) and governance (G_DifficultWords_Ratio) sections are significantly positively correlated with stock price crash risk, with coefficients of 1.3871 and 1.0387, respectively. This indicates that a higher proportion of complex words (lower readability) in the social and governance sections increases the likelihood of future stock price crashes. In other words, the readability of the social and governance sections is negatively associated with stock price crash risk.

Columns (4) to (6) show that when crash risk is measured by DUVOL, the proportions of difficult words in the environmental, social, and governance sections are all significantly positively correlated with stock price crash risk. This suggests that a higher proportion of complex words (lower readability) in each section increases crash risk. Therefore, Hypothesis 1 of this study is supported across all ESG dimensions.

Hypothesis 2 investigates which ESG dimension's readability has the most potent effect on crash risk. When the crash risk is measured by NCSKEW, a one-standard deviation increase in the proportion of difficult words in the social category increases crash risk by 0.0341 units (1.3871 * 0.0246). In comparison, the same increase in the governance category raises crash risk by 0.0276 units (1.0387 * 0.0266). When the crash risk is measured by DUVOL, a one-standard-deviation increase in the proportions of difficult words in the environmental, social, and governance categories increases crash risk by 0.0142 units (0.5356 * 0.0265), 0.0202 units (0.8213 * 0.0246), and 0.0179 units (0.6737 * 0.0266), respectively. These results demonstrate that the proportion of difficult words in the social category has the most potent effect on stock price crash risk. Moreover, the proportion of difficult words in the social category explains 22.7% of the variation in crash risk as measured by DUVOL (0.0202 / 0.0890).

Thus, Hypothesis 2 is supported, indicating that the readability of the social category's textual sections has the most potent effect on stock price crash risk among the three ESG dimensions.

[Insert Table 4 here]

4.3. Discussion of Moderating Variables

4.3.1. Moderating Effect of Positive Sentiment

Table 5 presents the empirical results of the moderating effect of positive sentiment. The findings indicate that when positive sentiment is strong (Sentiment_Pos = 1), the coefficients for the proportion of difficult words in all ESG categories are significant under both measures of crash risk (NCSKEW and DUVOL). This suggests that a higher proportion of complex words (lower readability) increases stock price crash risk in contexts where textual sentiment is more optimistic. Conversely, when positive sentiment is weak (Sentiment_Pos = 0), the regression coefficients for the proportion of difficult words in all ESG categories are insignificant, indicating that the relationship between difficult words and stock price crash risk becomes insignificant in less positively toned texts.

The results in Table 5 demonstrate that positive sentiment strengthens the negative relationship between the readability of individual ESG textual sections in sustainability reports and stock price crash risk. A potential explanation is that in positively toned texts, investors may pay less attention to or tend to overlook the true implications of difficult words, leading to the accumulation of bad news and an increased likelihood of stock price crashes.

Therefore, Hypothesis 3 is supported, indicating that in strongly positive sentiment contexts, the relationship between the readability of individual textual sections in sustainability reports and stock price crash risk is amplified.

[Insert Table 5 here]

4.3.2. Moderating Effect of Institutional Investor Ownership Proportion

Table 6 presents the empirical results of using the shareholding ratios of three major institutional investors as moderating factors. The findings indicate that:

- Foreign institutional ownership significantly enhances the negative relationship between the readability of the environmental and social categories and stock price crash risk.
- Proprietary trading ownership significantly enhances the negative relationship between the readability of the social category and stock price crash risk.
- Mutual fund ownership does not exhibit a significant moderating effect, meaning it has no notable impact on the negative relationship between readability and crash risk in any category.

Therefore, Hypothesis 4b is partially supported, as different institutional ownership types affect the relationship between readability and stock price crash risk. The effects of foreign institutional ownership and proprietary trading ownership are more pronounced, strengthening the negative relationship between readability and crash risk.

A possible explanation is that higher foreign and proprietary trading ownership ratios imply a greater likelihood of alignment and collusion of interests with management, which could increase managerial incentives for manipulation. This, in turn, heightens the possibility of bad news accumulation, ultimately increasing the risk of a stock price crash.

[Insert Table 6 here]

4.3.3. Moderating Effect of ESG Performance

Table 7 presents the empirical results using ESG performance as a moderating factor. The findings indicate that:

- When ESG performance is better (ESG_Lead = 1), the proportion of difficult words in all ESG categories is significantly positively correlated with stock price crash risk proxies (NCSKEW and DUVOL).
- When ESG performance is poorer (ESG_Lead = 0), the proportion of difficult words in all ESG categories shows no significant correlation with stock price crash risk.

This implies that better ESG performance strengthens the positive (or negative) relationship between the proportion of complex words (readability) in ESG textual sections of sustainability reports and stock price crash risk. Therefore, Hypothesis 5 is empirically supported.

A possible explanation is that ESG performance is a positive signal in the market. When ESG performance is better, investors tend to have greater trust in the company and are thus less sensitive to using complex words in sustainability reports. This reduced sensitivity may exacerbate the accumulation of bad news, ultimately increasing the likelihood of a subsequent stock price crash.

[Insert Table 7 here]

4.4. Robustness Test

4.4.1. Incorporating the Previous Year's Negative Skewness of Returns as a Control Variable

Chen et al. (2001) found that if a company's negative skewness coefficient (**NCSKEW**) in a year is significant, the negative skewness coefficient in year t is also likely to be significant. Accordingly, this study incorporates the serial correlation of the

negative skewness coefficient into the robustness test. Specifically, the study extends Equation (8) by including the prior year's negative skewness coefficient ($\mathbf{NCSKEW_{t-1}}$) as a control variable to examine whether the negative relationship between the readability of ESG textual sections in sustainability reports and stock price crash risk is affected.

Table 8 presents the regression results after including the prior year's negative skewness coefficient ($\mathbf{NCSKEW_{t-1}}$) as a control variable. The empirical results indicate the following:

- When NCSKEW measures crash risk, only the coefficient for the proportion of difficult words in the social category (1.3059) is significant at the 5% level. In contrast, the coefficients for the environmental and governance categories are insignificant.
- When the crash risk is measured by **DUVOL**, the coefficients for the proportions of difficult words in the environmental, social, and governance categories (0.5197, 0.7897, and 0.6361, respectively) are all significant.

These results show that after controlling for the prior year's negative skewness coefficient, the overall regression results remain consistent with the main findings. This demonstrates the robustness of the study's conclusions.

[Insert Table 8 here]

4.4.2. Endogeneity Issue: Instrumental Variable Regression

Since the Ordinary Least Squares (OLS) estimation results may be subject to endogeneity issues, leading to biased estimates that fail to accurately reflect the relationships between variables, instrumental variables (IV) can mitigate endogeneity problems in regression analysis.

In this study, the industry averages of the proportions of difficult words in the environmental, social, and governance sections of sustainability reports are used as instrumental variables. A two-stage regression model is conducted based on this approach. The first-stage regression model is specified in Equation (10), where the instrumental variables for the ESG categories are denoted as E_IV, S_IV, and G_IV, respectively.

The predicted values obtained from the first-stage regression (IV_pred) are then used in the second-stage regression, as specified in Equation (11). In this stage, the estimated values of the proportions of difficult words for each category from the first stage are denoted as E_IV_pred, S_IV_pred, and G_IV_pred, respectively.

$$DifficultWords\_Ratio_{i,t-1} = \alpha + \beta_1 IV_{i,t-1} + \beta_2 SIZE_{i,t-1} + \beta_3 ROA_{i,t-1} + \beta_4 LEV_{i,t-1} + \beta_5 MB_{i,t-1} + \beta_6 RET_{i,t-1} + \beta_7 SIGMA_{i,t-1} + \beta_8 DTURN_{i,t-1} + \beta_9 ABACC_{i,t-1} + Year\ \ Effect + Industry\ \ Effect + \varepsilon_{i,t-1} \quad (10)$$

Where *IV*= $E\_IV,\ S\_IV,\ G\_IV$

$$Crash_{i,t} = \alpha + \beta_1 IV\_pred_{i,t-1} + \beta_2 SIZE_{i,t-1} + \beta_3 ROA_{i,t-1} + \beta_4 LEV_{i,t-1} + \beta_5 MB_{i,t-1} + \beta_6 RET_{i,t-1} + \beta_7 SIGMA_{i,t-1} + \beta_8 DTURN_{i,t-1} + \beta_9 ABACC_{i,t-1} + Year\ \ Effect + Industry\ \ Effect + \varepsilon_{i,t} \quad (11)$$

Where IV_pred=$E\_IV\_pred,\ S\_IV\_pred,\ G\_IV\_pred$

Table 9 presents the results of the two-stage regression under the two stock price crash risk proxies (**NCSKEW** and **DUVOL**). The empirical results indicate that the proportions of difficult words in the environmental, social, and governance categories remain significantly positively correlated with stock price crash risk.

Therefore, after accounting for endogeneity issues, the findings of this study remain robust. This confirms that the proportion of complex words (**DifficultWords_Ratio**) in the textual sections of sustainability reports across all ESG categories is significantly associated with stock price crash risk.

[Insert Table 9 here]

4.4.3. Sample Selection Bias Issue

This study uses companies that publish sustainability reports as the research sample. However, this sampling approach may introduce potential sample selection bias, which could affect the accuracy of the estimation results. To address this concern, the study applies the Heckman two-stage sample selection model (Heckman, 1976) to correct endogeneity issues arising from sample selection bias.

The **first-stage regression** employs a probit model to estimate the probability of an observation being included in the sample. It constructs the **inverse Mills ratio (IMR)** for potential selection bias. The **second-stage regression** incorporates the inverse Mills ratio

into the main model to correct for estimation bias caused by sample selection bias. This is represented by the regression coefficient **LAMBDA** in **Table 10**.

Table 10 presents the regression results of the Heckman two-stage model. Column (1) shows the results of the first-stage probit model, while columns (2)-(3), (4)-(5), and (6)-(7) show the second-stage regression results for the proportions of difficult words in the environmental, social, and governance categories under the two stock price crash risk measures (**NCSKEW** and **DUVOL**), respectively.

Key findings from Table 10 are as follows:

- After correcting for sample selection bias, the coefficients for the proportions of difficult words in the **social** and **governance** categories remain significant when NCSKEW measures crash risk.
- When the crash risk is measured by **DUVOL**, the coefficients for the proportions of difficult words in the **environmental**, **social**, and **governance** categories are all significant.

These results are consistent with the main model's findings, indicating that this study's results are robust even after accounting for sample selection bias.

[Insert Table 10 here]

5. Concluding Remarks

This study investigates the relationship between the textual characteristics of sustainability reports in Taiwan and stock price crash risk. The proportion of complicated words is used as a reverse measure of readability, while stock price crash risk is analyzed using two indicators: the negative skewness coefficient (NCSKEW) and the down-to-up volatility ratio (DUVOL). The empirical results reveal that when crash risk is measured by NCSKEW, the readability of textual sections in the social and governance categories is negatively associated with stock price crash risk. When crash risk is measured by DUVOL, the readability of textual sections in all three categories—environmental, social, and governance—is negatively associated with stock price crash risk.

These findings suggest that a higher proportion of complex words (lower readability) in the textual descriptions of sustainable practices across ESG categories increases the accumulation of bad news, thereby raising the risk of a stock price crash. Among the three categories, the social dimension has the most substantial impact on crash risk, indicating

that unstructured textual information in the social category of sustainability reports is more value-relevant to external investors. This aligns with the view of Dumitrescu and Zakriya (2021), who argue that CSR's social dimension performance is more value-relevant. The primary contribution of this study lies in analyzing stock price crash risk from the perspective of unstructured textual information in ESG categories rather than solely focusing on ESG performance information.

Additionally, the study uses positive sentiment, institutional ownership ratios, and ESG performance as moderating variables. The empirical results indicate that positive sentiment strengthens the negative relationship between the readability of E, S, and G textual sections and stock price crash risk. A possible explanation is that in texts with stronger positive sentiment, investors are less concerned with the implications of difficult words, leading to easier accumulation of bad news and higher crash risk. Among the moderating effects of institutional ownership ratios, foreign institutional ownership and proprietary trading ownership have a more significant moderating effect on the relationship between the readability of social textual sections and crash risk, enhancing the negative relationship between readability and crash risk. When a company's ESG performance improves, the negative impact of the readability of E, S, and G textual sections on stock price crash risk becomes significantly more potent. This could be due to ESG performance acting as a positive market signal, leading to higher trust from investors and reduced sensitivity to difficult words in sustainability reports, which amplifies the accumulation of bad news and increases crash risk.

The conclusions of this study remain robust after accounting for endogeneity issues and sample selection bias. However, there are limitations to this study. The environmental, social, and governance textual sections were processed by extracting paragraphs containing keywords from each ESG category, which might result in a single paragraph being classified into multiple categories. Future research could explore alternative methods for textual categorization. Moreover, the study uses the proportion of difficult words as a proxy for readability, which may not capture other dimensions of readability, such as sentence complexity. Future research is encouraged to adopt natural language processing (NLP) techniques or investigate other textual features to provide a more comprehensive assessment of the overall quality of sustainability reports.

References

Andreou, P. C., Antoniou, C., Horton, J. and Louca, C., 2016. Corporate Governance and Firm-specific Stock Price Crashes. ***European Financial Management***, 22(5), 916-956.

Baier, P., Berninger, M. and Kiesel, F., 2020. Environmental, Social and Governance Reporting in Annual Reports: A Textual Analysis. ***Financial Markets, Institutions & Instruments***, 29(3), 93-118.

Basu, S. 1997. The conservatism principle and the asymmetric timeliness of earnings. ***Journal of Accounting and Economics***, 24: 3-37.

Bowen, R. M., Davis, A. K. and Matsumoto, D. A., 2005. Emphasis on Pro Forma Versus GAAP Earnings in Quarterly Press Releases: Determinants, SEC Intervention, and Market Reactions. ***The Accounting Review***, 80(4), 1011-1038.

Buchanan, B., Cao, C. X. Y. and Chen, C. Y., 2018. Corporate Social Responsibility, Firm Value, and Influential Institutional Ownership. ***Journal of Corporate Finance***, 52, 73-95.

Cahan, S. F., De Villiers, C., Jeter, D. C., Naiker, V. and Van Staden, C. J., 2016. Are CSR Disclosures Value Relevant? Cross-country Evidence. ***European Accounting Review***, 25(3), 579-611.

Chen, J., Hong, H. and Stein, J. C., 2001. Forecasting Crashes: Trading Volume, Past Returns, and Conditional Skewness in Stock Prices. ***Journal of Financial Economics***, 61(3), 345-381.

Cho, C. H., Roberts, R. W. and Patten, D. M., 2010. The Language of US Corporate Environmental Disclosure. ***Accounting, Organizations and Society***, 35(4), 431-443.

Clarkson, P., Li, Y., Richardson, G. and Tsang, A., 2019. Causes and Consequences of Voluntary Assurance of CSR Reports: International Evidence Involving Dow Jones Sustainability Index Inclusion and Firm Valuation. ***Accounting, Auditing & Accountability Journal***, 32(8), 2451-2474.

Clarkson, P. M., Li, Y., Richardson, G. D. and Vasvari, F. P., 2008. Revisiting the Relation Between Environmental Performance and Environmental Disclosure: An Empirical Analysis. ***Accounting, Organizations and Society***, 33(4-5), 303-327.

Davis, A. K., Ge, W. L., Matsumoto, D. and Zhang, J. L., 2015. The Effect of Manager-specific Optimism on the Tone of Earnings Conference Calls. ***Review of Accounting Studies***, 20(2), 639-673.

Dechow, P. M., Sloan, R. G. and Sweeney, A. P., 1995. Detecting Earnings Management. ***Accounting Review***, 193-225.

Dhaliwal, D. S., Li, O. Z., Tsang, A. and Yang, Y. G., 2011. Voluntary Nonfinancial Disclosure and the Cost of Equity Capital: The Initiation of Corporate Social Responsibility Reporting. ***The Accounting Review***, *86*(1), 59-100.

Du, S., Yu, K., Bhattacharya, C. and Sen, S., 2017. The Business Case for Sustainability Reporting: Evidence from Stock Market Reactions. ***Journal of Public Policy & Marketing***, 36(2), 313-330.

Du, S. L. and Yu, K., 2021. Do Corporate Social Responsibility Reports Convey Value Relevant Information? Evidence from Report Readability and Tone. ***Journal of Business Ethics***, 172(2), 253-274.

Dumitrescu, A. and Zakriya, M., 2021. Stakeholders and the Stock Price Crash Risk: What Matters in Corporate Social Performance? ***Journal of Corporate Finance***, 67, 101871.

Fisher, R., van Staden, C. J. and Richards, G., 2019. Watch That Tone an Investigation of the Use and Stylistic Consequences of Tone in Corporate Accountability Disclosures. ***Accounting Auditing & Accountability Journal***, 33(1), 77-105.

Harris, T. L. and Hodges, R. E., 1995. ***The Literacy Dictionary: The Vocabulary of Reading and Writing***. Order Department, International Reading Association, 800 Barksdale Road, PO Box 8139, Newark, DE 19714-8139

Heckman, J. J., 1976. The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. ***In Annals of Economic and Social Measurement***, volume 5, number 4 (pp. 475-492). NBER.

Hutton, A. P., Marcus, A. J. and Tehranian, H., 2009. Opaque Financial Reports, R2, and Crash Risk. ***Journal of Financial Economics***, 94(1), 67-86.

Ignatov, K., 2023. When ESG Talks: ESG Tone of 10-K Reports and its Significance to Stock Markets. ***International Review of Financial Analysis***, 89, 102745.

Jin, L. and Myers, S. C., 2006. R2 Around the World: New Theory and New Tests. ***Journal of Financial Economics***, 79(2), 257-292.

Kim, J. B., Li, Y. H. and Zhang, L. D., 2011. CFOs Versus CEOs: Equity Incentives and Crashes. ***Journal of Financial Economics***, 101(3), 713-730.

Kitzmueller, M. and Shimshack, J., 2012. Economic Perspectives on Corporate Social Responsibility. ***Journal of Economic Literature***, 50(1), 51-84.

Kothari, S. P., Susan S., Wysocki, P.D. 2009. Do managers withhold bad news? ***Journal of Accounting Research***, 47(1): 241-276.

Li, F., 2008. Annual Report Readability, Current Earnings, and Earnings Persistence. ***Journal of Accounting and Economics***, 45(2-3), 221-247.

Li, Z. G., Zhang, T. H., Zhao, X. B. and Zhu, Y. C., 2022. Monitoring or Colluding? Institutional Investors' Heterogeneity and Environmental Information Disclosure Behavior. ***Frontiers in Psychology***, 13, 911901.

Loughran, T. and McDonald, B., 2011. When is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. ***The Journal of Finance***, 66(1), 35-65.

Marhfor, A., Bouslah, K. and M'Zali, B., 2017. Corporate Social Responsibility and Stock Price Informativeness: Evidence from the Canadian Market. ***Unpublished Working Paper***

Merkl-Davies, D. M., & Brennan, N. M. (2017). A Theoretical Framework of External Accounting Communication: Research Perspectives, Traditions, and Theories. ***Accounting, Auditing & Accountability Journal***, 30(2), 433-469.

Mermod, A. Y. and Idowu, S. O., 2014. ***Corporate Social Responsibility in the Global Business World*** (Vol. 258). Heidelberg: Springer.

Miller, B. P., 2010. The Effects of Reporting Complexity on Small and Large Investor Trading. ***The Accounting Review***, 85(6), 2107-2143.

Milone, M., 2009. ***The Development of ATOS: The Renaissance Readability Formula***. Renaissance Learning, Incorporated.

Muslu, V., Mutlu, S., Radhakrishnan, S. and Tsang, A., 2014. Measuring Corporate Social Responsibility Report Quality Using Narratives. ***University of Houston, Houston***.

Nagel, G. L., Qayyum, M. A. and Roskelley, K. D., 2015. Do Motivated Institutional Investors Monitor Firm Payout and Performance? ***Journal of Financial Research***, 38(3), 349-377.

Petersen, M. A., 2008. Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. ***The Review of Financial Studies***, 22(1), 435-480.

Pound, J., 1992. Beyond Takeovers: Politics Comes to Corporate Control. ***Harvard Business Review***, 70(2), 83-93.

Rennekamp, K., 2012. Processing Fluency and Investors' Reactions to Disclosure Readability. ***Journal of Accounting Research***, 50(5), 1319-1354.

Sung, Y.-T., Chen, J.-L., Lee, Y.-S., Cha, J.-H., Tseng, H.-C., Lin, W.-C., Chang, T.-H., & Chang, K.-E., 2013. Investigating Chinese Text Readability: Linguistic Features, Modeling, and Validation. ***Chinese Journal of Psychology***, 55(1), 75–106.

Sung, Y. T., Chang, T. H., Lin, W. C., Hsieh, K. S., & Chang, K. E. 2016. CRIE: An automated analyzer for Chinese texts. ***Behavior Research Methods***, 48(4), 1238-1251.

Tschopp, D. and Huefner, R. J., 2015. Comparing the Evolution of CSR Reporting to that of Financial Reporting. ***Journal of Business Ethics***, 127, 565-577.

Verrecchia, R. E., 2001. Essays on Disclosure. ***Journal of Accounting and Economics***, 32(1-3), 97-180.

Table 1. Sample Distribution

Table of Sample Distribution by Year of Sustainability Report Publication				
Year	Capitalization Less Than 5 Billion	Capitalization Between 5 Billion and 10 Billion	Capitalization Greater Than 10 Billion	Total
2014	97	33	66	196
2015	146	36	70	252
2016	172	72	83	327
2017	197	80	83	360
2018	216	75	85	376
2019	256	86	86	428
2020	308	84	86	478
2021	403	96	84	583
2022	590	95	92	777
Total	2385	657	735	3,777

Note: This table represents the sample distribution from 2014 to 2022.

Table 2. Summary Statistics of Major Variables

Variable	Obs	Mean	Std. Dev.	Min	Max
$NCSKEW$	3,777	-0.2884	1.0209	-2.0558	4.3809
$DUVOL$	3,777	0.0890	0.4064	-1.0966	1.6257
$E\_DifficultWords\_Ratio$	3,777	0.3198	0.0265	0.2198	0.5577
$S\_DifficultWords\_Ratio$	3,777	0.3117	0.0246	0.2188	0.4895
$G\_DifficultWords\_Ratio$	3,777	0.3094	0.0266	0.0938	0.5359
$SIZE$	3,777	9.4644	1.4917	4.9767	16.5848
ROA (%)	3,777	5.9395	7.7290	-49.1000	79.6000
LEV	3,777	0.4415	0.1783	0.0048	0.9787
MB	3,777	2.1926	2.6311	0.2800	76.8300
RET	3,777	0.5688	0.4037	-0.4916	2.4019
$SIGMA$	3,777	1.1097	0.2369	0.4683	2.2428
$DTURN$	3,777	-0.0004	0.2570	-2.2964	4.8157
$ABACC$	3,777	0.0765	0.0750	0.0000	1.5495
$E\_Sentiment\_Score$	3,777	0.1243	0.0152	-0.0798	0.2056
$S\_Sentiment\_Score$	3,777	0.1292	0.0147	0.0728	0.1843
$G\_Sentiment\_Score$	3,777	0.1282	0.0160	0.0584	0.1829
$Foreign\_Investors$	3,777	0.1682	0.1779	0.0000	0.9947
$Investment\_Trust$	3,777	0.0078	0.0189	0.0000	0.2793
$Dealer$	3,777	0.0021	0.0086	0.0000	0.2000
ESG	3,581	61.2475	7.4072	31.9850	83.7300

Note: This table provides the descriptive statistics for each variable, including sample size, mean, standard deviation, minimum, and maximum values. The dependent variables are two stock price crash risk measures: NCSKEW and DUVOL. The main research variables are the proportions of difficult words in the ESG categories, namely E_DifficultWords_Ratio, S_DifficultWords_Ratio, and G_DifficultWords_Ratio. Control variables include company size (SIZE), return on assets (ROA), leverage ratio (LEV), market-to-book ratio (MB), mean of weekly stock returns (RET), standard deviation of weekly stock returns (SIGMA), turnover rate difference (DTURN), and absolute value of discretionary accruals (ABACC). Moderating variables include the sentiment scores for each ESG category (E_Sentiment_Score, S_Sentiment_Score, and G_Sentiment_Score), institutional ownership ratio (II_RATIO), which comprises foreign institutional ownership ratio (Foreign_Investors), investment trust ownership ratio (Investment_Trust), and proprietary trading ownership ratio (Dealer), and ESG performance (ESG).

Table 3. Pearson Correlation Coefficient Table

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
$NCSKEW$ (1)	1.0000												
$DUVOL$ (2)	0.9399	1.0000											
$E\_DifficultWords\_Ratio$ (3)	0.0688	0.0682	1.0000										
$S\_DifficultWords\_Ratio$ (4)	0.0613	0.0652	0.8827	1.0000									
$G\_DifficultWords\_Ratio$ (5)	0.0707	0.0741	0.8327	0.8774	1.0000								
$SIZE$ (6)	0.0033	-0.0166	0.2924	0.3086	0.2590	1.0000							
ROA (7)	0.0408	0.0287	0.0013	0.0116	0.0069	0.2954	1.0000						
LEV (8)	-0.0180	-0.0265	0.1585	0.1501	0.1257	0.1309	-0.2191	1.0000					
MB (9)	-0.0891	-0.0819	-0.0089	0.0164	-0.0021	0.2453	0.3109	-0.0437	1.0000				
RET (10)	-0.4031	-0.3550	-0.0796	-0.0571	-0.0815	0.0444	0.0576	-0.0183	0.2223	1.0000			
$SIGMA$ (11)	-0.1483	-0.1208	-0.0396	-0.0306	-0.0289	-0.0786	-0.0175	-0.0336	0.0831	-0.0066	1.0000		
$DTURN$ (12)	-0.0510	-0.0490	-0.0060	-0.0110	-0.0113	0.0291	0.0837	0.0128	0.0591	0.2282	0.0919	1.0000	
$ABACC$ (13)	-0.0920	-0.0918	0.0262	0.0251	0.0102	0.0567	0.0136	0.0406	0.0731	0.0954	0.0481	0.0319	1.0000

Note: Table 3 presents the Pearson correlation coefficients between variables, based on the total number of sustainability reports from 2014 to 2022. Variable definitions are consistent with those in Table 2.

Table 4. Relationship Between Text Readability of ESG Categories in Sustainability Reports and Stock Price Crash Risk

	(1) *NCSKEW*	(2) *NCSKEW*	(3) *NCSKEW*	(4) *DUVOL*	(5) *DUVOL*	(6) *DUVOL*
E_DifficultWords_Ratio	0.8695			0.5356**		
	(1.37)			(2.08)		
S_DifficultWords_Ratio		1.3871**			0.8213***	
		(2.03)			(2.97)	
G_DifficultWords_Ratio			1.0387*			0.6737***
			(1.73)			(2.76)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3777	3777	3777	3777	3777	3777
Adjusted R^2	0.2421	0.2426	0.2423	0.1872	0.1882	0.1879
Test of Coefficient Difference						
(2) – (1) chi^2(p-value)		2.41			4.07	
(5) – (4) chi^2(p-value)		(0.1202)			(0.0436)	
(2) – (3) chi^2(p-value)		1.02			1.12	
(5) – (6) chi^2(p-value)		(0.3120)			(0.2891)	

Note: NCSKEW and DUVOL are the proxy variables for stock price crash risk and serve as the primary dependent variables in this study. Controls refer to the control variables used in this study, as detailed in Table 2. The empirical results above also account for fixed effects of year and industry, and coefficient difference tests between the two categories were conducted. Values in parentheses represent t-values, where * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** indicates significance at the 1% level.

Table 5. Relationship Between Text Readability of ESG Categories in Sustainability Reports and Stock Price Crash Risk: Moderating Effect of Positive Sentiment

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A	*NCSKEW*	*NCSKEW*	*NCSKEW*	*NCSKEW*	*NCSKEW*	*NCSKEW*
Sentiment_Pos	*1*	*0*	*1*	*0*	*1*	*0*
E_DifficultWords_Ratio	2.5707***	-1.1648				
	(3.08)	(-1.25)				
S_DifficultWords_Ratio			2.5172***	-0.0564		
			(2.87)	(-0.06)		
G_DifficultWords_Ratio					2.0955**	-0.4618
					(2.55)	(-0.54)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1894	1883	1942	1835	1957	1820
Adjusted R^2	0.2463	0.2381	0.2600	0.2295	0.2432	0.2561
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B	*DUVOL*	*DUVOL*	*DUVOL*	*DUVOL*	*DUVOL*	*DUVOL*
Sentiment_Pos	*1*	*0*	*1*	*0*	*1*	*0*
E_DifficultWords_Ratio	1.2539***	-0.3364				
	(3.66)	(-0.85)				
S_DifficultWords_Ratio			1.3658***	0.1316		
			(3.85)	(0.32)		
G_DifficultWords_Ratio					1.0639***	0.0672
					(3.26)	(0.19)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1894	1883	1942	1835	1957	1820
Adjusted R^2	0.1944	0.1813	0.2088	0.1747	0.1868	0.2032

Note: NCSKEW and DUVOL are proxy variables for stock price crash risk and serve as the primary dependent variables in this study. Controls refer to the control variables used in this study, as detailed in Table 2. The SentimentPos variable indicates whether the textual sections of ESG categories are classified as having stronger positive sentiment (i.e., sentiment scores in the top 50th percentile). If so, the value is 1; otherwise, it is 0. The empirical results also control for fixed effects of year and industry. Values in parentheses represent t-values, where * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level.

Table 6. Relationship Between Text Readability of ESG Categories in Sustainability Reports and Stock Price Crash Risk: Moderating Effect of Institutional Ownership Proportion

Panel A. ***Foreign_Investors***	(1) *NCSKEW*	(2) *NCSKEW*	(3) *NCSKEW*	(4) *DUVOL*	(5) *DUVOL*	(6) *DUVOL*
E_DifficultWords_Ratio	0.0153			0.1456		
	(0.02)			(0.42)		
Foreign_Investors	-1.9790**	-1.9049*	-1.2009	-0.8682**	-0.7528*	-0.5258
	(-2.01)	(-1.82)	(-1.29)	(-2.10)	(-1.71)	(-1.31)
E_DifficultWords_Ratio Foreign_Investors*	5.5696*			2.5347**		
	(1.83)			(1.98)		
S_DifficultWords_Ratio		0.4777			0.4517	
		(0.52)			(1.20)	
S_DifficultWords_Ratio Foreign_Investors*		5.4699*			2.2307	
		(1.65)			(1.61)	
G_DifficultWords_Ratio			0.4746			0.4138
			(0.59)			(1.25)
G_DifficultWords_Ratio Foreign_Investors*			3.2650			1.5262
			(1.11)			(1.20)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3777	3777	3777	3777	3777	3777
Adjusted R^2	0.2430	0.2434	0.2428	0.1879	0.1887	0.1881
Panel B. ***Investment_Trust***	(1) *NCSKEW*	(2) *NCSKEW*	(3) *NCSKEW*	(4) *DUVOL*	(5) *DUVOL*	(6) *DUVOL*
E_DifficultWords_Ratio	0.9246			0.5319*		
	(1.33)			(1.89)		
Investment_Trust	-1.5746	-0.3059	0.8446	-1.4560	-1.0097	-0.3724
	(-0.18)	(-0.03)	(0.10)	(-0.38)	(-0.25)	(-0.10)
E_DifficultWords_Ratio Investment_Trust*	0.2193			3.0298		
	(0.01)			(0.26)		
S_DifficultWords_Ratio		1.4657*			0.8244***	
		(1.96)			(2.76)	
S_DifficultWords_Ratio Investment_Trust*		-3.8608			1.6636	
		(-0.13)			(0.13)	
G_DifficultWords_Ratio			1.1359*			0.6902***
			(1.72)			(2.60)
G_DifficultWords_Ratio Investment_Trust*			-7.5709			-0.3586
			(-0.28)			(-0.03)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3777	3777	3777	3777	3777	3777

$Adjusted\ R^2$	0.2423	0.2428	0.2426	0.1872	0.1882	0.1878
	(1)	(2)	(3)	(4)	(5)	(6)
Panel C. *Dealer*	$NCSKEW$	$NCSKEW$	$NCSKEW$	$DUVOL$	$DUVOL$	$DUVOL$
$E\_DifficultWords\_Ratio$	0.7523			0.5113*		
	(1.16)			(1.95)		
$Dealer$	-19.2803	-36.8505**	-6.1428	-3.9506	-13.8522	-1.4067
	(-1.09)	(-2.41)	(-0.29)	(-0.46)	(-1.61)	(-0.13)
$E\_DifficultWords\_Ratio*$ $Dealer$	59.9558			12.4197		
	(1.06)			(0.45)		
$S\_DifficultWords\_Ratio$		1.1711*			0.7393***	
		(1.69)			(2.63)	
$S\_DifficultWords\_Ratio*$ $Dealer$		113.9745**			42.9474	
		(2.24)			(1.49)	
$G\_DifficultWords\_Ratio$			1.0073			0.6664***
			(1.64)			(2.66)
$G\_DifficultWords\_Ratio*$ $Dealer$			18.6252			4.2896
			(0.27)			(0.12)
$Controls$	Yes	Yes	Yes	Yes	Yes	Yes
$Year\ FE$	Yes	Yes	Yes	Yes	Yes	Yes
$Industry\ FE$	Yes	Yes	Yes	Yes	Yes	Yes
$Observations$	3777	3777	3777	3777	3777	3777
$Adjusted\ R^2$	0.2418	0.2427	0.2419	0.1868	0.1882	0.1874

Note: NCSKEW and DUVOL are proxy variables for stock price crash risk and serve as the primary dependent variables in this study. Controls refer to the control variables used in this study, as detailed in Table 2. Foreign_Investors, Investment_Trust, and Dealer represent the proportions of foreign institutional ownership, investment trust ownership, and proprietary trading ownership, respectively. The empirical results also account for fixed effects of year and industry. Values in parentheses represent t-values, where * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level.

Table 7. Relationship Between Text Readability of ESG Categories in Sustainability Reports and Stock Price Crash Risk: Moderating Effect of ESG Performance

Panel A	*NCSKEW*	*NCSKEW*	*NCSKEW*	*NCSKEW*	*NCSKEW*	*NCSKEW*
ESG_Lead	*1*	*0*	*1*	*0*	*1*	*0*
E_DifficultWords_Ratio	2.0540**	-0.1481				
	(2.20)	(-0.17)				
S_DifficultWords_Ratio			2.4427**	0.4544		
			(2.49)	(0.48)		
G_DifficultWords_Ratio					1.7766**	0.5350
					(1.98)	(0.65)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1837	1744	1837	1744	1837	1744
Adjusted R^2	0.2627	0.2295	0.2632	0.2296	0.2623	0.2297
Panel B	*DUVOL*	*DUVOL*	*DUVOL*	*DUVOL*	*DUVOL*	*DUVOL*
ESG_Lead	*1*	*0*	*1*	*0*	*1*	*0*
E_DifficultWords_Ratio	1.1010***	0.0286				
	(2.93)	(0.08)				
S_DifficultWords_Ratio			1.2976***	0.4157		
			(3.22)	(1.08)		
G_DifficultWords_Ratio					1.0517***	0.4080
					(2.91)	(1.20)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1837	1744	1837	1744	1837	1744
Adjusted R^2	0.2080	0.1719	0.2088	0.1724	0.2078	0.1725

Note: NCSKEW and DUVOL are proxy variables for stock price crash risk and serve as the primary dependent variables in this study. Controls refer to the control variables used in this study, as detailed in Table 2. The ESG_Lead variable indicates whether the ESG score of the evaluated company (measured using the TESG rating) is in the top 50th percentile. If so, the value is 1; otherwise, it is 0. The empirical results also account for fixed effects of year and industry. Values in parentheses represent t-values, where * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level.

Table 8. Robustness Test: Incorporating the Previous One-Year Negative Skewness of Returns as a Control Variable

	$NCSKEW$	$NCSKEW$	$NCSKEW$	$DUVOL$	$DUVOL$	$DUVOL$
$E\_DifficultWords\_Ratio$	0.8289			0.5197**		
	(1.35)			(2.07)		
$S\_DifficultWords\_Ratio$		1.3059**			0.7897***	
		(1.98)			(2.92)	
$G\_DifficultWords\_Ratio$			0.9419			0.6361***
			(1.63)			(2.67)
$Controls$	Yes	Yes	Yes	Yes	Yes	Yes
$Year\ FE$	Yes	Yes	Yes	Yes	Yes	Yes
$Industry\ FE$	Yes	Yes	Yes	Yes	Yes	Yes
$Observations$	3777	3777	3777	3777	3777	3777
$Adjusted\ R^2$	0.2482	0.2486	0.2483	0.1930	0.1939	0.1935

Note: NCSKEW and DUVOL are proxy variables for stock price crash risk and serve as the primary dependent variables in this study. Controls include not only the variables listed in Table 2 but also the prior year's negative skewness coefficient (NCSKEW $t-1$). The empirical results also account for fixed effects of year and industry. Values in parentheses represent t-values, where * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level.

Table 9. Robustness Test: Instrumental Variable Regression

Panel A	*E_DifficultWords_Ratio*	*NCSKEW*	*S_DifficultWords_Ratio*	*NCSKEW*	*G_DifficultWords_Ratio*	*NCSKEW*
E_IV	0.9500***					
	(13.74)					
E_IV_pred		5.5681*				
		(1.80)				
S_IV			0.9299***			
			(12.48)			
S_IV_pred				6.7283*		
				(1.88)		
G_IV					0.9700***	
					(12.71)	
G_IV_pred						5.1575*
						(1.72)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3777	3777	3777	3777	3777	3777
Adjusted R^2	0.2710	0.2424	0.2610	0.2426	0.2317	0.2424
Panel B	*E_DifficultWords_Ratio*	*DUVOL*	*S_DifficultWords_Ratio*	*DUVOL*	*G_DifficultWords_Ratio*	*DUVOL*
E_IV	0.9500***					
	(13.74)					
E_IV_pred		2.7260**				
		(2.25)				
S_IV			0.9299***			
			(12.48)			
S_IV_pred				2.6979*		
				(1.93)		
G_IV					0.9700***	
					(12.71)	
G_IV_pred						2.3423*
						(1.92)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3777	3777	3777	3777	3777	3777
Adjusted R^2	0.2710	0.1873	0.2610	0.1871	0.2317	0.1872

Note: This table presents the results of the two-stage regression model incorporating instrumental variables, where E_IV, S_IV, and G_IV are the instrumental variables, and E_IV_pred, S_IV_pred, and G_IV_pred are the estimated values of the respective instrumental variables. NCSKEW and DUVOL serve as proxy variables for stock price crash risk and represent the primary dependent variables in this study. Controls refer to the control variables listed in Table 2. The empirical results also account for fixed effects of year and industry. Values in parentheses represent t-values, where * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level.

Table 10. Robustness Test: Heckman Two-Stage Sample Selection Model

		E		*S*		*G*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Report	*NCSKEW*	*DUVOL*	*NCSKEW*	*DUVOL*	*NCSKEW*	*DUVOL*
E_DifficultWords_Ratio		0.8721	0.5366**				
		(1.38)	(2.09)				
S_DifficultWords_Ratio				1.3916**	0.8231***		
				(2.04)	(2.98)		
G_DifficultWords_Ratio						1.0274*	0.6693***
						(1.71)	(2.73)
LAMBDA		0.0817	0.0327	0.0824	0.0331	0.0783	0.0305
		(1.01)	(1.03)	(1.02)	(1.05)	(0.97)	(0.96)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13611	3777	3777	3777	3777	3777	3777
Adjusted/Pseudo R^2	0.3452	0.2420	0.1871	0.2425	0.1882	0.2422	0.1878

Note: This table presents the results of robustness tests using the Heckman two-stage sample selection model. Report indicates whether the company published a sustainability report in the given year (1 for published, 0 for not published). LAMBDA represents the inverse Mills ratio. NCSKEW and DUVOL serve as proxy variables for stock price crash risk and represent the primary dependent variables in this study. Controls refer to the control variables listed in Table 2. The empirical results also account for fixed effects of year and industry. Values in parentheses represent t-values, where * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. ◦