Environmental Specificity Information Disclosure of Heavily Polluting Companies and Stock Price Crash Risk

Mei Li*

School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China

*Corresponding author: 1721973573@qq.com

Abstract. Using Natural Language Processing (NLP) techniques, this study empirically examines the impact of environmental specificity information disclosure on stock price crash risk, employing a sample of A-share listed companies in heavily polluting industries from 2013 to 2022. The study finds that environmental specificity information disclosure by heavily polluting companies significantly mitigates future stock price crash risk. Mechanism tests reveal that such disclosure reduces crash risk through pathway: improving corporate reputation. Heterogeneity analysis demonstrates that the inhibitory effect of environmental specificity information on crash risk is particularly prominent among firms characterized by low analyst coverage and weak government environmental regulation. This research verifies that environmental specificity information disclosure enhances information efficiency in heavily polluting enterprises, thereby providing empirical evidence for improving capital market resource allocation efficiency and optimizing environmental information disclosure systems.

Keywords: Environmental Specificity Information; Stock Price Crash Risk; information transparency.

1. Introduction

The capital market serves as the core hub for driving high-quality economic development. Effective capital markets provide critical support for corporate technological innovation and industrial upgrading by optimizing resource allocation and expanding direct financing channels. China has continuously advanced capital market reforms, enhancing service efficiency to the real economy through policies such as the comprehensive registration-based IPO system and the new "National Nine Guidelines." By the end of February 2025, the number of domestic listed companies reached 5,404 with a total market capitalization of RMB 87.11 trillion. However, compared with mature capital markets, China's stock market persistently suffers from stock price crash risk—a phenomenon that not only undermines the resource allocation function of capital markets but also threatens financial stability. Stock price crashes refer to sudden, severe declines in individual stock prices within short timeframes. Their unpredictability and instant-trigger characteristics can cause sharp devaluation of equity assets, leading to rapid wealth erosion for investors [1]. Such crashes may trigger investor panic selling, market liquidity droughts, and credit chain fractures. These effects not only damage individual firms' financing capacity but may also amplify risk contagion throughout the financial system. As cautioned in the State Council's *Several Opinions on Further Strengthening Capital Market Risk Prevention and Control*: "localized risks may evolve into systemic risks."

Simultaneously, accelerated progress toward the "dual carbon" goals and the development of the green financial system has elevated environmental issues to a core focus of regulatory oversight and public scrutiny. Within this context, the transmission pathways through which corporate environmental risks impact capital markets have been significantly amplified. In securities markets, stock price crashes triggered by environmental compliance violations or ecological incidents occur frequently. As a critical nexus linking corporate environmental responsibility with capital market pricing mechanisms, the risk-alert function of environmental information disclosure is increasingly prominent in policy practice. Regulations such as the Ministry of Ecology and Environment's Measures for the Management of Mandatory Corporate Environmental Information Disclosure further reinforce corporate environmental accountability and transparency requirements. Nevertheless, despite existing disclosure mandates, listed companies in China exhibit persistent shortcomings including: the absence of standardized frameworks, untimely reporting, and insufficient content coverage. Some firms treat environmental disclosure merely as a compliance formality, resorting to vague and perfunctory narratives to satisfy mandatory requirements. Conversely, disclosures incorporating entity-specific details—such as personnel names, timelines, locations, and organizational references—substantially enhance authenticity and informational value. Existing research defines such granular data as specificity information [2], which improves reliability and provides incremental decision-useful insights [2,3]. Building on this foundation, this study investigates: Could environmental specificity information disclosure provide incremental information for heavily polluting firms, thereby mitigating stock price crash risk through capital market channels?

This study utilizes a sample of A-share listed companies in heavily polluting industries. Applying Natural Language Processing techniques, we measure the specificity information content within environmental disclosure sections of annual reports and ESG reports to construct an objective environmental specificity information index. Based on information asymmetry theory and corporate reputation theory, we investigate and empirically test the impact of environmental specificity information on stock price crash risk and its transmission channels in heavily polluting firms. Results indicate that environmental specificity information disclosure suppresses stock price crash risk through pathway: improving corporate reputation. Furthermore, we examine contextual heterogeneity in the relationship between environmental specificity information and crash risk across different external environments. The study reveals that when external information intermediaries are deficient (i.e., under low analyst coverage), environmental specificity information more effectively generates incremental information, resulting in significantly stronger crash risk reduction. In contexts of weak government environmental regulation, such information substitutes for institutional supervision gaps to more substantially mitigate crash risk.

The contributions of this study are threefold. First, whereas prior research predominantly measures corporate environmental disclosure through constructed disclosure indices [4,5,6,7], this paper applies Natural Language Processing techniques to investigate environmental information disclosure via objective indicators derived from textual features. Second, we expand the literature on precursors of stock price crashes. Departing from existing studies examining the effects of financial information disclosure [8], ESG report publication [9], annual report readability [10], or textual similarity [11] on crash risk, we validate the inhibitory effect of environmental specificity information content on crash risk. This offers new insights for firms seeking to mitigate crash risk and investors utilizing environmental data in capital markets. Finally, we demonstrate environmental information disclosure's compensatory function for regulatory shortcomings, providing empirical evidence and policy implications for enhancing environmental disclosure frameworks.

2. Theoretical Analysis and Hypotheses Development

Grounding in agency theory, information manipulation theory posits that corporate insiders strategically conceal negative news by exploiting informational advantages. When accumulated negative information reaches a critical threshold and can no longer be suppressed, its sudden release triggers sharp stock price declines [12]. The accumulation and abrupt disclosure of adverse news resulting from information opacity constitute the primary driver of stock price crash risk [13]. Currently, environmental disclosures by Chinese listed companies suffer from inconsistent standards, ambiguous boundaries, and incomplete reporting [14]. Firms retain substantial discretion over what and how to disclose environmental information. Some entities resort to vacuous declarations such as: "Our company adheres to green development principles, continuously optimizes environmental management systems, enhances environmental performance, and actively contributes to building a Beautiful China." Such narratives lack substantive content. Conversely, other firms provide granular, entity-specific information incorporating temporal, locational, organizational, and quantitative details: "In 2023, we retrofitted the circulating water systems in high-pressure air separation stations and sintering facilities at Phase I of Qujing Defang Base, achieving daily electricity savings exceeding

30,000 kWh." It is evident that such granular, entity-specific environmental specificity information conveys greater incremental insights to report users, enabling external stakeholders to more effectively evaluate corporate environmental performance. It enables external stakeholders to better evaluate corporate environmental performance through: contextual precision (time/location references), operational verifiability (quantified outcomes), entity identifiability (project/organizational details).

Environmental specificity information disclosure by heavily polluting enterprises mitigates information asymmetry, preventing stock price crashes triggered by concentrated releases of concealed negative news. Traditional environmental disclosures often emphasize macro-level statistics and vague descriptions—termed "compliance statements" due to unverifiable authenticity stemming from poor traceability and entity linkage. In contrast, environmental specificity information enhances transparency through entity anchoring, temporal referencing, and relational mapping, shifting disclosure quality from mere existence to verifiability [15,16]. This reduces informational ambiguity and elevates decision usefulness, enabling investors to more accurately assess corporate environmental performance. Given their high environmental risk exposure, heavily polluting sectors face intensified challenges including regulatory pressures, public trust deficits, investor risk aversion, and sustainability demands. By disclosing specificity information—concretized through verifiable environmental actions—these firms provide substantive, high-transparency data to investors. This facilitates precise evaluation of environmental risks and sustainable development capabilities [17], effectively alleviating stakeholder skepticism and transforming environmental risks into competitive advantages. Crucially, such disclosure preempts sudden stock collapses caused by concentrated negative events (e.g., regulatory violations, pollution incidents). Simultaneously, as critical non-financial signals of organizational legitimacy, environmental specificity information improves verifiability. This allows investors to holistically evaluate current operations and future prospects [18], promoting rational investment decisions and thereby reducing crash risk from irrational trading behaviors.

Conversely, environmental specificity information disclosure by heavily polluting enterprises generates a signaling effect that enhances corporate reputation and investor confidence, thereby mitigating stock price crash risk. As economic entities, firms prioritize production-oriented activities. According to the financial capacity hypothesis proposed by Preston and O'Bannon (1997), companies allocate resources to core operations before engaging in non-productive initiatives [19]. Firms incapable of sustaining core operations cannot execute environmental or social responsibilities [3]. Substantive environmental disclosures signal robust operational capacity to undertake beyond-compliance environmental initiatives. Such information conveys positive operational signals to investors, increasing investment attractiveness [20]. Furthermore, disclosing environmental specificity information demonstrates a firm's commitment to mitigating production-driven environmental impacts through concrete actions. This cultivates an eco-conscious corporate image aligned with sustainable development. Such reputation capital provides competitive advantages and acts as a reputational buffer during negative events: environmental legitimacy insulates against investor overreactions, reducing crash risk.

Collectively, we propose:

H1: Environmental specificity information reduces stock price crash risk in heavily polluting companies

Environmental specificity information may reduce stock price crash risk in heavily polluting enterprises through reputational enhancement pathways. Compared to generic disclosures, specificity information—characterized by concreteness, traceability, and action orientation—more effectively builds stakeholder trust, strengthens brand value, and mitigates greenwashing suspicion. By disseminating differentiated, verifiable, and substantive environmental data, firms demonstrate unique environmental accountability to investors, consumers, regulators, and the public. This establishes an eco-conscious corporate identity that signals genuine environmental commitment [21]. Furthermore, detailed environmental disclosure generates a dual signaling effect. First, as survival constitutes the fundamental organizational imperative, the capacity to disclose specificity information signals robust operational viability—a prerequisite for undertaking environmental responsibilities. This conveys financial resilience to investors, elevating corporate reputation. Second, accumulated reputation capital serves as a buffer against overreactions during adverse events: stakeholders' preexisting trust attenuates panic selling, thereby suppressing crash risk. Thus, we hypothesize:

H2: Environmental specificity information reduces stock price crash risk in heavily polluting companies through enhanced corporate reputation.

3. Date and Research Design

3.1. Sample and Data

This study utilizes an initial sample of China's A-share listed companies in heavily polluting industries spanning 2013 to 2022. After excluding firms with ST or ST* delisting warnings and observations with missing data, the final sample consists of 5,640 firm-year observations. Financial and stock market data were sourced from the CSMAR database, while annual reports and ESG reports in PDF format were obtained from the CNINFO website.

Heavily polluting industries were defined according to Pan [22], encompassing: coal mining and washing, petroleum and natural gas extraction, ferrous metal mining and processing, non-ferrous metal mining and processing, textile manufacturing, leather/fur/feather/footwear production, paper and paper products, petroleum processing/coking/nuclear fuel processing, chemical raw materials and products, chemical fiber manufacturing, rubber and plastic products, non-metallic mineral products, ferrous metal smelting and rolling, non-ferrous metal smelting and rolling, and electricity/heat production and supply.

3.2. Variable Definitions

3.2.1 Explanatory variable

The explanatory variable of this paper is the environmental specificity information content (*Especificity*). Specifically, annual reports and ESG reports in PDF format were downloaded from the CNINFO website using web crawlers. Python's third-party module *pdfplumber* was then employed to extract textual content from these corporate reports. First, an environmental keyword lexicon was constructed through seed words expanded via word embeddings. Following Bozanic et al. [23], regular expressions were applied to analyze annual report sentences sequentially based on environmental keywords: sentences containing $\geqslant 1$ environmental keyword were classified as environmental disclosures, while others were excluded. All environmental disclosures were aggregated to form the annual environmental corpus for each firm. Second, referencing Hope et al. [2], Named Entity Recognition (NER) technology was implemented using Baidu's open-source NLP tool *PaddleNLP* to process environmental corpora, identifying seven entity categories: Person, Organization, Location, Time, Money, Number (including ordinals), and Date. Finally, the frequency ratio of these entity words to total word count was calculated to construct the corporate environmental specificity information metric (*Especificity*).

$$Especificity = \text{The total frequency of specificity entities in environmental corpus / Total word count of environmental corpus} \quad (1)$$

3.2.2 Explained variable

The explanatory variable in this paper is stock price crash risk. Following prior studies, stock price crash risk is measured using two indicators: the negative coefficient of skewness (*NCSKEW*) and the down-to-up volatility (*DUVOL*) [24,25]. The computational procedures for these metrics are specified below.

First, for each year, the weekly firm-specific return is calculated by estimating the regression model in Equation (2) using weekly individual stock returns and market returns. In Equation (2), $r_{i,t}$

denotes the return of firm *i* in week *t*, while $r_{M,t}$ represents the value-weighted market return in week *t*. The residual $\varepsilon_{i,t}$ captures the portion of individual stock returns unexplained by market fluctuations, i.e., the weekly firm-specific return. Following established methodology, we transform $\varepsilon_{i,t}$ as: $W_{i,t}=ln(1+\varepsilon_{i,t})$.

$$r_{i,t}=\alpha_i+\beta_1 r_{M,t-2}+\beta_2 r_{M,t-1}+\beta_3 r_{M,t}+\beta_4 r_{M,t+1}+\beta_5 r_{M,t+2}+\varepsilon_{i,t} \tag{2}$$

After calculating the weekly firm-specific returns ($W_{i,t}$), these values are substituted into Equations (3) and (4) to compute the *NCSKEW* and the *DUVOL*, respectively.

In Equation (3), *n* denotes the number of trading weeks for stock *i* in year *t*. The higher the *NCSKEW* value, the more pronounced the negative skewness of stock returns, indicating greater stock price crash risk.

$$NCSKEW_{i,t}=-[n(n-1)^{3/2}\sum W_{i,t}^3]/[(n-1)(n-2)(\sum W_{i,t}^2)^{3/2}] \tag{3}$$

In Equation (4), n_u denotes the number of weeks in year *t* where the weekly return of stock *i* exceeds the annual mean return (Up weeks), while n_d represents the number of weeks where the weekly return falls below the annual mean return (Down weeks). A higher *DUVOL* value indicates greater left-skewness in the return distribution, signaling higher stock price crash risk.

$$DUVOL_{i,t}=log\{[(n_u-1)\sum_{Down} W_{i,t}^2]/[(n_d-1)\sum_{Up} W_{i,t}^2]\} \tag{4}$$

3.2.3 Control variables

Following Kim [26] and Shi [25], we control for: Firm size (*Size*), Return on assets (*Roa*), Leverage ratio (*Lev*), Firm age (*Eage*), Growth rate (*Growth*), Proportion of independent directors (*Independent*), Firm age (*Firmage*), Tobin's Q (*TobinQ*), Managerial shareholding ratio (*Mngshare*), State-owned enterprise (*Soe*), Corporate sustainability capacity (*ESG*), Earnings per share (*Earning*_per), Volatility of weekly returns (*Sigma*), and Average weekly return (*Ret*). Year and industry fixed effects were also included. See Table 1 for complete variable definitions.

Table 1. Definition of variables

Variable type	variable	notation	Variable Definition
Explanatory variable	Environmental specificity information content	Especificity	The total frequency of specificity entities in environmental corpus / Total word count of environmental corpus
Explained variable	Negative coefficient of skewness	NCSKEW	Calculate according to the formula
	Down-to-up volatility	DUVOL	Calculate according to the formula
Control variables	Firm size	Size	Natural logarithm of firm's total assets.
	Return on assets	Roa	Net profit after tax / Total assets
	Leverage ratio	Lev	Total liability / Total assets
	Firm age	Eage	Sample year - The year when the firm was established
	Growth rate	Growth	Growth rate of operating income
	Proportion of independent directors	Independent	The number of independent directors / The total number of board members
	Tobin's Q	TobinQ	Market value of equity / Book value of equity
	Managerial shareholding ratio	Mngshare	Management shareholding/Total share capital
	State-owned enterprise	Soe	Assign 1 if the enterprise is state-owned, otherwise assign 0
	Corporate sustainability capacity	ESG	Huazheng ESG Rating
	Earnings per share	Earning_per	After-tax profit / Total share capital
	Volatility of weekly returns	Sigma	The standard deviation of the weekly return rate
	Average weekly return	Ret	Average weekly rate of return
	Year	Year	Year dummy variable
	Industry	Ind	Industry dummy variable

3.3. Model

According to the previous analysis and hypotheses, combined with the selection of variables, the regression model (5) is constructed to test hypothesis H1.

$$NCSKEW_{i,t+1}/DUVOL_{i,t+1} = \alpha_0 + \alpha_1 Especificity_{i,t} + \alpha_2 Controls_{i,t} + \sum Year + \sum Ind + \varepsilon_{i,t} \quad (5)$$

$NCSKEW_{i,t+1}$ and $DUVOL_{i,t+1}$ denote the stock price crash risk in the year following annual report; $Especificity_{i,t}$ represents the environmental specificity information content in the report of firm; $Controls_{i,t}$ signify control variables; $Year_{i,t}$ and $Ind_{i,t}$ represent the year and industry, respectively; $\varepsilon_{i,t}$ denotes the residual.

$$NCSKEW_{i,t+1}/DUVOL_{i,t+1} = \alpha_0 + \alpha_1 Especificity_{i,t} + \alpha_2 Controls_{i,t} + \sum Year + \sum Ind + \varepsilon_{i,t} \quad (6)$$

$$Rep_{i,t+1} = \alpha_0 + \alpha_1 Especificity_{i,t} + \alpha_2 Controls_{i,t} + \sum Year + \sum Ind + \varepsilon_{i,t} \quad (7)$$

$$NCSKEW_{i,t+1}/DUVOL_{i,t+1} = \alpha_0 + \alpha_1 Especificity_{i,t} + \alpha_2 Rep_{i,t+1} \alpha_3 Controls_{i,t} + \sum Year + \sum Ind + \varepsilon_{i,t} \quad (8)$$

the model (6) is constructed to test hypothesis H2. $Rep_{i,t+1}$ denotes the corporate reputation in the year following the annual report.

4. Empirical Results

4.1. Summary Statistics

As shown in the descriptive statistics in Table 2, the mean value of environmental specificity information content (*Especificity*) is 0.0534, with minimum and maximum values of 0.0092 and 0.1248, respectively, indicating substantial variation in environmental information disclosure practices across sample firms. The mean values of *NCSKEW* and *DUVOL* are -0.3284 and -0.2212, while their standard deviations are 0.7313 and 0.4815, respectively. This suggests significant heterogeneity in stock price crash risk among sample firms, with some exhibiting substantial crash risk exposure.

Regarding the control variables, firm size (*Size*) has a mean value of 22.4621 with a standard deviation of 1.3737. The leverage ratio (*Lev*) shows a mean of 0.4338 and median of 0.4342, consistent with prior literature. Other metrics also fall within reasonable ranges, indicating that our sample is representative of the target population.

Table 2. Summary statistics of main variables

Variable	N	Mean	Sd	Min	Median	Max
Especificity	5640	0.0534	0.0192	0.0092	0.0519	0.1248
NCSKEW	5640	-0.3284	0.7313	-3.2413	-0.2954	2.3869
DUVOL	5640	-0.2212	0.4815	-1.6794	-0.2166	1.5418
Size	5640	22.4621	1.3737	19.7034	22.2253	26.5988
Roa	5640	0.0434	0.0566	-0.2424	0.0381	0.2875
Lev	5640	0.4338	0.1980	0.0338	0.4342	0.9849
Eage	5640	18.5988	5.4265	4.0000	19.0000	35.0000
Growth	5640	0.1657	0.4069	-0.5629	0.0914	5.3008
Independent	5640	0.3715	0.0507	0.3333	0.3333	0.5714
TobinQ	5640	1.7929	1.1165	0.7781	1.4499	11.4217
Mngshare	5640	0.0698	0.1543	0.0000	0.0002	0.9665
Soe	5640	0.4151	0.4928	0.0000	0.0000	1.0000
ESG	5640	73.4585	5.5445	53.5300	73.8800	86.7400
Earning_per	5640	0.4181	0.6507	-2.1907	0.2849	5.7357
Sigma	5640	0.0622	0.0253	0.0193	0.0564	0.2074
Ret	5640	0.0035	0.0090	-0.0220	0.0024	0.0572

4.2. Analysis of Main Effects Regression Results

The main regression results indicate that environmental specificity information content (*Especificity*) exhibits statistically significant negative relationships with both *NCSKEW* and *DUVOL* at the 1% level. This suggests that disclosure of environmental specificity information by heavily polluting firms effectively reduces stock price crash risk in China's capital market. These findings support H1. The complete regression outputs are presented in Table 3.

Table 3. Environmental specificity information content and stock price crash risk

	(1)	(2)
VARIABLES	NCSKEW	DUVOL
Especificity	-1.6931***	-0.9486***
	(-3.111)	(-2.662)
Size	-0.0060	-0.0127*
	(-0.521)	(-1.687)
Roa	-0.4385	-0.0472
	(-1.493)	(-0.243)
Lev	0.0307	0.0477
	(0.441)	(1.037)
Growth	0.0398	0.0098
	(1.644)	(0.575)
Independent	0.0214	0.0649
	(0.111)	(0.501)
Eage	-0.0029	-0.0016
	(-1.390)	(-1.148)
TobinQ	0.0272***	0.0131*
	(2.814)	(1.882)
Mngshare	-0.0580	-0.1043**
	(-0.832)	(-2.221)
Soe	-0.0066	-0.0155
	(-0.276)	(-0.979)
ESG	-0.0005	-0.0003
	(-0.273)	(-0.278)
Earning_per	0.0631***	0.0253*
	(2.797)	(1.700)
Sigma	-2.3635***	-1.1025***
	(-3.885)	(-2.739)
Ret	8.5959***	4.9543***
	(5.031)	(4.277)
Constant	0.0297	0.1567
	(0.105)	(0.840)
Observations	5,640	5,640
R-squared	0.054	0.056
Year Ind	YES	YES
F	6.185	4.361
r2_a	0.0476	0.0497

Constant represents the intercept term; robust standard errors are reported in parentheses.

4.3. Mechanism Test of Corporate Reputation

Following Guo's methodology [27], corporate reputation (*Rep*) is measured as the natural logarithm of one plus the annual sum of positive media coverage in online and print publications.

Table 4 reports regression results indicating a marginally significant positive relationship between environmental specificity information content (*Especificity*) and corporate reputation (*Rep*) at the 10% level. This suggests that disclosing environmental specificity information enhances corporate reputation. When introducing *Rep* into the main regression model, it shows a significant negative association with stock price crash risk at the 1% significance level. These results demonstrate that environmental specificity information content reduces stock price crash risk by improving corporate reputation. H2 is thereby supported.

Table 4. Mechanism testing of corporate reputation

VARIABLES	(1) NCSKEW	(2) DUVOL	(3) Rep	(4) NCSKEW	(5) DUVOL
Especificity	-1.5510***	-0.8768**	0.9337*	-1.4661***	-0.8058**
	(-2.807)	(-2.421)	(1.724)	(-2.664)	(-2.237)
Rep				-0.0910***	-0.0760***
				(-6.526)	(-8.091)
Size	-0.0077	-0.0134*	0.3940***	0.0282**	0.0165*
	(-0.655)	(-1.752)	(35.570)	(2.168)	(1.955)
Roa	-0.5201*	-0.1117	0.4859*	-0.4759	-0.0748
	(-1.709)	(-0.559)	(1.658)	(-1.565)	(-0.375)
Lev	0.0499	0.0592	0.0235	0.0521	0.0610
	(0.705)	(1.266)	(0.357)	(0.740)	(1.317)
Growth	0.0422*	0.0115	0.0414	0.0459*	0.0147
	(1.724)	(0.671)	(1.555)	(1.873)	(0.851)
Independent	0.0693	0.0945	0.6856***	0.1317	0.1466
	(0.355)	(0.723)	(3.660)	(0.679)	(1.130)
Eage	-0.0038*	-0.0021	-0.0040*	-0.0041*	-0.0024*
	(-1.763)	(-1.499)	(-1.914)	(-1.931)	(-1.715)
TobinQ	0.0314***	0.0164**	0.1098***	0.0414***	0.0247***
	(3.220)	(2.316)	(9.526)	(4.184)	(3.460)
Mngshare	-0.0810	-0.1209**	0.1959***	-0.0632	-0.1060**
	(-1.151)	(-2.535)	(3.003)	(-0.901)	(-2.228)
Soe	-0.0091	-0.0159	0.0035	-0.0088	-0.0156
	(-0.378)	(-0.991)	(0.147)	(-0.366)	(-0.980)
ESG	0.0003	0.0001	0.0050***	0.0007	0.0005
	(0.134)	(0.094)	(2.811)	(0.375)	(0.404)
Earning_per	0.0680***	0.0286*	0.0997***	0.0771***	0.0361**
	(2.932)	(1.875)	(4.141)	(3.308)	(2.359)
Sigma	-2.5187***	-1.1525***	0.2408	-2.4968***	-1.1342***
	(-4.066)	(-2.812)	(0.399)	(-4.048)	(-2.789)
Ret	8.7191***	5.1062***	9.0030***	9.5383***	5.7906***
	(5.048)	(4.358)	(5.292)	(5.521)	(4.950)
Constant	-0.0059	0.1252	-5.4072***	-0.4979*	-0.2858
	(-0.021)	(0.661)	(-19.214)	(-1.680)	(-1.459)
Observations	5,499	5,499	5,499	5,499	5,499
R-squared	0.056	0.057	0.478	0.063	0.069
Year Ind	YES	YES	YES	YES	YES
F	6.514	4.681	180.0	8.935	8.924
r2_a	0.0492	0.0510	0.474	0.0564	0.0627

Constant represents the intercept term; robust standard errors are reported in parentheses.

4.4. Robustness Tests

4.4.1 Eliminate the samples of abnormal years

The samples from 2015 were removed to eliminate interference with the disaster of the Chinese stock market in 2015. The *Especificity* coefficient remains negatively significant. The results are shown in Table 5 as follows.

Table 5. Egression after excluding the samples from 2015

VARIABLES	(1) NCSKEW	(2) DUVOL
Especificity	-1.8297***	-0.9847***
	(-3.213)	(-2.647)
Size	-0.0041	-0.0105
	(-0.338)	(-1.327)
Roa	-0.4342	-0.0527
	(-1.410)	(-0.261)
Lev	0.0215	0.0375
	(0.292)	(0.772)
Growth	0.0442*	0.0096
	(1.836)	(0.560)
Independent	0.0228	0.0503
	(0.113)	(0.374)
Eage	-0.0035	-0.0020
	(-1.596)	(-1.402)
TobinQ	0.0355***	0.0173**
	(3.309)	(2.287)
Mngshare	-0.0717	-0.1020**
	(-1.035)	(-2.171)
Soe	-0.0134	-0.0209
	(-0.542)	(-1.267)
ESG	-0.0003	-0.0003
	(-0.163)	(-0.264)
Earning_per	0.0693***	0.0282*
	(3.002)	(1.860)
Sigma	-2.6204***	-1.3058***
	(-4.071)	(-3.073)
Ret	8.4896***	5.0142***
	(4.581)	(4.011)
Constant	0.0094	0.1428
	(0.032)	(0.735)
Observations	5,197	5,197
R-squared	0.053	0.053
Year Ind	YES	YES
F	6.558	4.384
r2_a	0.0467	0.0461

Constant represents the intercept term; robust standard errors are reported in parentheses.

4.4.2 Replace the measurement of explained variable

As a robustness check, we re-estimate the regressions using alternative market-adjusted crash risk measures. While the main regressions employ negative skewness (*NCSKEW*) and down-to-up volatility (DUVOL) computed based on float market capitalization weighting, this sensitivity analysis adopts total market capitalization weighting. The coefficient on *Especificity* remains negative and statistically significant. These results are reported in Table 6.

Table 6. Egression after replacing the measurement of explained variable

VARIABLES	(1) NCSKEW	(2) DUVOL
Especificity	-1.6075***	-0.9786***
	(-2.950)	(-2.733)

Size	-0.0033	-0.0091
	(-0.281)	(-1.203)
Roa	-0.4406	-0.0373
	(-1.506)	(-0.193)
Lev	0.0244	0.0420
	(0.351)	(0.921)
Growth	0.0408*	0.0112
	(1.688)	(0.650)
Independent	0.0347	0.0544
	(0.179)	(0.420)
Eage	-0.0030	-0.0016
	(-1.425)	(-1.109)
TobinQ	0.0290***	0.0159**
	(2.989)	(2.305)
Mngshare	-0.0590	-0.0706
	(-0.899)	(-1.595)
Soe	-0.0134	-0.0215
	(-0.564)	(-1.359)
ESG	-0.0006	-0.0003
	(-0.310)	(-0.229)
Earning_per	0.0647***	0.0219
	(2.881)	(1.463)
Sigma	-2.4708***	-1.2098***
	(-4.051)	(-3.026)
Ret	8.8834***	5.2430***
	(5.199)	(4.585)
Constant	-0.0427	0.0651
	(-0.151)	(0.349)
Observations	5,640	5,640
R-squared	0.054	0.057
Year Ind	YES	YES
F	6.569	4.754
r2_a	0.0474	0.0505

Constant represents the intercept term; robust standard errors are reported in parentheses.

4.4.3 Propensity Score Matching Method (PSM)

Given the substantial discretion in environmental information disclosure, firms with superior environmental performance are more likely to voluntarily disclose such information, potentially introducing endogeneity concerns. While the main regressions mitigate this issue by controlling for ESG ratings, we explicitly follow the approach of Liu et al. [28] and Huang et al. [29] to conduct propensity score matching (PSM) for additional endogeneity testing. Firms are divided into treatment and control groups based on their environmental specificity information content levels. Using all control variables as covariates, we implement a nearest-neighbor matching method with a 1:1 ratio via Logit regression. Table 7 presents the post-matching regression results, showing that the coefficient on *Especificity* remains negative and statistically significant.

Table 7. result of PSM

	(1)	(2)
VARIABLES	NCSKEW	DUVOL
Especificity	-1.6971***	-0.9530**
	(-2.752)	(-2.331)
Size	0.0028	-0.0082

	(0.218)	(-0.970)
Roa	-0.3165	0.1147
	(-0.944)	(0.521)
Lev	-0.0378	0.0191
	(-0.470)	(0.358)
Growth	0.0483*	0.0210
	(1.852)	(1.149)
Independent	0.1792	0.1968
	(0.838)	(1.350)
Eage	-0.0035	-0.0023
	(-1.440)	(-1.428)
TobinQ	0.0323***	0.0153*
	(2.715)	(1.784)
Mngshare	-0.0096	-0.0743
	(-0.125)	(-1.385)
Soe	-0.0202	-0.0218
	(-0.736)	(-1.201)
ESG	-0.0002	-0.0003
	(-0.086)	(-0.188)
Earning_per	0.0509**	0.0165
	(2.115)	(1.042)
Sigma	-2.3202***	-1.1225**
	(-3.324)	(-2.416)
Ret	7.7970***	4.1154***
	(3.964)	(3.073)
Constant	-0.2148	0.0254
	(-0.691)	(0.123)
Observations	4,282	4,282
R-squared	0.058	0.059
Year Ind	YES	YES
F	5.021	3.440
r2_a	0.0495	0.0508

Constant represents the intercept term; robust standard errors are reported in parentheses.

5. Further Analysis

5.1. Heterogeneity Analysis of Analyst Coverage

The effective functioning of information intermediary mechanisms in capital markets significantly improves the informational ecology of corporate disclosures. As quintessential information intermediaries, financial analysts provide critical information channels for external stakeholders. When market information structures reach a benign equilibrium, analysts—acting as professional information decoders—inject incremental information flows into the market through dual pathways: information screening and private information acquisition [30]. This process effectively counteracts the informational deficiency arising from homogeneous and generic environmental disclosures.

Conversely, when firms have limited disclosure channels (i.e., under low analyst coverage), environmental disclosures by heavily polluting enterprises become the primary—or even sole—source for investors to assess environmental performance. Consequently, environmental risk evaluations become heavily anchored to firm-supplied disclosures. Thus, the mitigating effect of environmental specificity information on stock price crash risk manifests more prominently in subsamples with lower analyst coverage.

Analyst coverage (*Coverage*) is measured as the natural logarithm of one plus the number of securities analysts following a listed firm [31]. We split the sample into high- and low-coverage groups based on the median value. Columns (1) and (3) in Panel A of Table 8 show statistically insignificant coefficients on *Especificity* in regressions explaining *NCSKEW* and *DUVOL* for the high analyst coverage subsample. In contrast, Columns (2) and (4) reveal that *Especificity* exhibits a significantly negative coefficient on *NCSKEW* at the 1% level and on *DUVOL* at the 5% level in the low analyst coverage subsample.

Table 8. Grouping regression of analyst coverage

	(1)	(2)	(3)	(4)
VARIABLES	NCSKEW	NCSKEW	DUVOL	DUVOL
Especificity	-1.0464	-2.2025***	-0.7589	-1.0629**
	(-1.410)	(-2.762)	(-1.519)	(-2.081)
Size	0.0168	-0.0559***	0.0080	-0.0384***
	(1.053)	(-2.757)	(0.742)	(-2.970)
Roa	-0.2979	-0.5879	-0.1645	0.1392
	(-0.798)	(-1.181)	(-0.600)	(0.456)
Lev	0.0079	0.0711	0.0119	0.0832
	(0.080)	(0.742)	(0.173)	(1.347)
Growth	0.0107	0.0575*	-0.0081	0.0183
	(0.262)	(1.899)	(-0.281)	(0.861)
Independent	0.3106	-0.1340	0.2200	-0.0302
	(1.160)	(-0.474)	(1.189)	(-0.163)
Eage	-0.0036	-0.0025	-0.0015	-0.0022
	(-1.279)	(-0.785)	(-0.766)	(-1.077)
TobinQ	0.0304**	0.0112	0.0180*	0.0079
	(2.293)	(0.741)	(1.870)	(0.739)
Mngshare	-0.1294	-0.0144	-0.1107*	-0.1020
	(-1.504)	(-0.124)	(-1.857)	(-1.338)
Soe	-0.0562	0.0276	-0.0513**	0.0010
	(-1.612)	(0.828)	(-2.143)	(0.046)
ESG	-0.0005	-0.0017	-0.0007	-0.0005
	(-0.181)	(-0.638)	(-0.412)	(-0.308)
Earning_per	0.0434*	0.0698	0.0257	0.0179
	(1.685)	(1.435)	(1.422)	(0.610)
Sigma	-1.0032	-3.0988***	-0.5139	-1.4206**
	(-1.163)	(-3.583)	(-0.847)	(-2.538)
Ret	7.2353***	9.4940***	4.6923***	5.2983***
	(3.145)	(3.541)	(3.023)	(2.933)
Constant	-0.6490	1.3163***	-0.3632	0.7902**
	(-1.620)	(2.703)	(-1.312)	(2.543)
Observations	2,688	2,952	2,688	2,952
R-squared	0.071	0.055	0.075	0.057
Year Ind	YES	YES	YES	YES
F	3.227	3.503	2.637	2.843
r2_a	0.0575	0.0427	0.0620	0.0455

Constant represents the intercept term; robust standard errors are reported in parentheses.

5.2. Heterogeneity Analysis of Government Environmental Regulation

Government environmental regulation refers to statutes, policies, standards, and measures enacted by governmental authorities to protect ecosystems, reduce pollution, and conserve resources. Its core objective is internalizing ecological externalities through legislative mandates and administrative interventions. Due to heterogeneous industrial structures and economic development levels across Chinese regions, regulatory stringency exhibits significant interprovincial variation.

Under stringent regulatory regimes, heavily polluting enterprises face binding compliance costs. To meet governmental environmental requirements, they implement substantive measures to reduce emissions and enhance resource efficiency [32]. In such contexts, even with informational gaps in firm-specific environmental disclosures, investors form stable environmental risk assessments through institutional safeguards. Consequently, the risk-pricing function of idiosyncratic information becomes attenuated by regulatory assurance mechanisms.

Conversely, in weak regulatory environments, polluting firms retain greater discretion over environmental investments, resulting in heightened cross-firm environmental risk heterogeneity. Investors thus heavily anchor their risk assessments to firm-supplied idiosyncratic information. Environmental specificity disclosures thereby provide incremental informational value—enhancing transparency and mitigating stock price crash risk. Voluntary environmental specificity disclosures partially substitute for regulatory voids by exerting compensatory functions [33]. Therefore, the inhibitory effect of environmental specificity disclosures on crash risk is significantly stronger for firms facing weak environmental regulation, whereas it becomes statistically indistinguishable for those subject to stringent regulatory regimes.

Government environmental regulatory stringency (*RegStringency*) is measured by the frequency of environmental regulation-related keywords in provincial government work reports [34]. We partition the sample into high- and low-stringency groups based on the median value. Columns (1) and (3) in Panel A of Table 9 report statistically insignificant coefficients on *Especificity* in both *NCSKEW* and *DUVOL* regressions for the high regulatory stringency subsample. Conversely, Columns (2) and (4) reveal significantly negative coefficients on *Especificity* at the 1% level for *NCSKEW* and at the 5% level for *DUVOL* in the low regulatory stringency subsample.

Table 9. Grouping regression of government environmental regulation

VARIABLES	(1) NCSKEW	(2) NCSKEW	(3) DUVOL	(4) DUVOL
Especificity	-0.7596	-2.5638***	-0.5622	-1.2932**
	(-1.015)	(-3.238)	(-1.145)	(-2.477)
Size	-0.0176	0.0017	-0.0238**	-0.0036
	(-1.048)	(0.105)	(-2.176)	(-0.340)
Roa	-0.8963**	0.0795	-0.3852	0.3396
	(-2.117)	(0.193)	(-1.425)	(1.205)
Lev	-0.0768	0.1370	-0.0155	0.1054
	(-0.768)	(1.397)	(-0.238)	(1.599)
Growth	0.0181	0.0653*	0.0004	0.0200
	(0.579)	(1.756)	(0.019)	(0.771)
Independent	-0.2519	0.2593	-0.1947	0.2860
	(-0.887)	(0.968)	(-1.030)	(1.597)
Eage	-0.0026	-0.0035	-0.0016	-0.0018
	(-0.911)	(-1.086)	(-0.865)	(-0.883)
TobinQ	0.0291**	0.0281**	0.0150	0.0142
	(2.098)	(2.074)	(1.519)	(1.444)
Mngshare	0.0508	-0.2010*	-0.0892	-0.1281*
	(0.560)	(-1.843)	(-1.439)	(-1.783)
Soe	0.0030	-0.0027	-0.0156	-0.0065
	(0.089)	(-0.078)	(-0.695)	(-0.285)
ESG	-0.0008	-0.0005	0.0001	-0.0011
	(-0.325)	(-0.169)	(0.071)	(-0.605)
Earning_per	0.0892***	0.0317	0.0510**	-0.0041
	(2.963)	(0.918)	(2.570)	(-0.183)
Sigma	-3.2237***	-1.5358*	-1.7004***	-0.5099
	(-3.751)	(-1.770)	(-2.964)	(-0.894)

Ret	9.8410***	7.0510***	5.5625***	4.2297**
	(4.118)	(2.866)	(3.475)	(2.510)
Constant	0.4553	-0.2856	0.5097*	-0.1226
	(1.096)	(-0.734)	(1.868)	(-0.473)
Observations	2,802	2,838	2,802	2,838
R-squared	0.062	0.059	0.061	0.066
Year Ind	YES	YES	YES	YES
F	3.796	3.614	2.991	2.536
r2_a	0.0492	0.0470	0.0485	0.0535

Constant represents the intercept term; robust standard errors are reported in parentheses.

6. Summary

This study employs natural language processing (NLP) techniques to extract environmental specificity information from annual reports and ESG disclosures of A-share listed heavily polluting enterprises during 2013-2022, examining its capital market implications. Results demonstrate that environmental specificity disclosure reduces stock price crash risk—greater disclosure corresponds to lower crash risk exposure.

Mediation tests reveal that environmental specificity disclosure mitigates crash risk through pathway: improving corporate reputation.

Further analyses indicate this informational effect is more pronounced in firms characterized by: lower analyst coverage and weaker government environmental regulation.

Based on the research conclusion of this paper, the following important inspirations can be drawn:

First, enterprises should regard environmental information disclosure as a strategic imperative rather than mere regulatory compliance. Through proactive and transparent disclosure, companies can reduce investor uncertainty while building positive social legitimacy and enhancing brand value. This contributes not only to short-term price stability but also establishes foundations for long-term sustainable development. Meanwhile, when formulating information disclosure strategies, enterprises should fully consider external environmental factors. In markets with low analyst coverage, intensify disclosures to fill information gaps; In regions with weak environmental regulation, demonstrate social accountability through beyond-compliance reporting.

Secondly, regulatory authorities should improve the mandatory and standardized framework for environmental information disclosure. It is suggested that the regulatory authorities formulate classified and graded disclosure standards for heavily polluting industries, requiring enterprises to quantitatively disclose core indicators such as pollutant emissions, environmental governance investment, and carbon footprint. Moreover, by establishing an environmental information quality verification system, the modifying components of environmental information disclosure can be identified to enhance the credibility and comparability of disclosed data. By leveraging artificial intelligence technology, an enterprise environmental data anomaly monitoring system is constructed to cross-verify the disclosed information with satellite remote sensing monitoring data, supply chain energy consumption data, etc. Corresponding measures are implemented for enterprises suspected of "greenwashing". Meanwhile, the quality of environmental information disclosure should be incorporated into the green credit assessment system, bond issuance review standards and tax preferential policy considerations, and enterprises should be guided to disclose proactively through the capital market pricing mechanism. Give full play to the complementary effect between the disclosure behavior of enterprises and the intensity of environmental regulations.

Finally, give full play to the decision-making usefulness of environmental information of heavily polluting enterprises and improve the efficiency of resource allocation in the capital market. The research conclusion of this paper indicates that the characteristic information in the environmental information disclosure of heavily polluting enterprises can provide incremental information, which is of reference significance for investors to judge the future development and stock price stability of enterprises. Accordingly, improve the environmental information interpretation support system for small and medium-sized investors, require listed companies to add a visual summary module for environmental information disclosure in their annual reports, and force investment advisory institutions to conduct environmental risk stress test demonstrations when promoting stocks in heavily polluting industries.

References

[1] YE Yanyi, GAO Haoyue, YANG Xiaoguang. The Risk of Stock Price Surging: An Empirical Study Based on China's Stock Market. Chinese Journal of Management Science, 2019, 27(7): 11-22.

[2] HOPE O, HU D, LU H. The benefits of specific risk-factor disclosures. Review of Accounting Studies,2016,21(4):1005-1045.

[3] LI Sihai, LI Zhen. Research on the Signaling Effect of Characteristic Information Content in Corporate Social Responsibility Reports: Analysis Based on Natural Language Processing Technology. China Industrial Economics, 2023(1): 114-131.

[4] HUANG Jun, ZHOU Chunna. Equity Structure, Management Behavior and Their Impact on Environmental Information Disclosure: Empirical Evidence from Heavily Polluting Industries in Shanghai Stock Market. China Soft Science, 2012(1): 133-143.

[5] DUAN Cunru, WANG Linjie, ZHOU Xiaoxi, et al. Environmental Information Disclosure and Green Product Market Information Asymmetry: Mechanism and Empirical Test. Statistics & Information Forum, 2022(12): 66-77.

[6] ZHANG Z, SU Z, WANG K, et al. Corporate environmental information disclosure and stock price crash risk: Evidence from Chinese listed heavily polluting companies. Energy Economics, 2022, 112: 106116.

[7] LI Zhengda, LI Feng, ZHAO Yating. The Financing Effect of Environmental Information Disclosure: Evidence from Heavily Polluting Enterprises. Audit & Economy Research, 2024, 39(1): 117-127.

[8] MAMUN M. A, BALACHANDRAN B, DUONG H. N. Powerful CEOs and Stock Price Crash Risk. Journal of Corporate Finance,2020,62.

[9] XI Longsheng, WANG Yan. Corporate ESG Information Disclosure and Stock Price Crash Risk. Economic Problems, 2022(8): 57-64.

[10] LIU Huiqin, SHI Xianwang. Annual Report Text Readability and Stock Price Crash Risk. Journal of Investment Research, 2022, 41(7): 129-148.

[11] SONG Xinbei, CHEN Ying, LU Dong, et al. Information Environment, Incremental Information Disclosure of Listed Companies, and Capital Market Pricing Efficiency: Research Based on MD&A Text Similarity. Nankai Business Review, 2024(8): 1-26.

[12] JIN L, MYERS S C. R2 around the world: New theory and new tests. Journal of Financial Economics, 2006, 79(2): 257-292.

[13] HUTTON A P, MARCUS A, TEHRANIAN H. Opaque Financial Reports, R2 and Crash Risk. Journal of Financial Economics,2008,94(1):67-86.

[14] LI Xiaoqi, HU Yanglin, SHI Wei. Top-level Institutional Design of ESG Reporting in China: A Preliminary Study. Securities Market Herald, 2022(4): 35-44.

[15] TIETENBERG T, 1998. Disclosure Strategies for Pollution Control. Environmental and Resource Economics, 11(3): 587-602.

[16] BOIRAL O, 2013. Sustainability reports as simulacra? A counter-account of A and A+ GRI reports. Accounting, Auditing & Accountability Journal, 26(7): 1036-1071.

[17] CORMIER D, LEDOUX M, MAGNAN M, 2011. The informational contribution of social and environmental disclosures for investors. Management Decision, 49(8): 1276-1304.

[18] ROBINSON, S., and M. Eilert. The Role of Message Specificity in Corporate Social Responsibility Communication. Journal of Business Research,2018,90(C):260-268.

[19] PRESTON, L. E., and D. P. O'BANNON. The Corporate Social-financial Performance Relationship: A Typology and Analysis. Business and Society,1997,36(4):419-429.

[20] MORALES-RAYA M, MARTIN-TAPIA I, ORTIZ-DE-MANDOJANA N, 2018. To Be or to Seem: The Role of Environmental Practices in Corporate Environmental Reputation. Organization & Environment, 32: 108602661775315.

[21] WANG Zhongbing, JIN Xiaochao. Research on the Correlation Between Carbon Information Disclosure and Enterprise Value. Macroeconomics, 2013(1): 86-90.

[22] PAN Ailing, LIU Xin, QIU Jinlong, et al. Can Green Mergers and Acquisitions Under Media Pressure Promote Substantive Transformation of Heavily Polluting Enterprises?. China Industrial Economics, 2019(2): 174-192.

[23] BOZANIC Z, ROULSTONE D T, Van Buskirk A. Management earnings forecasts and other forward-looking statements. Journal of Accounting and Economics,2018,65(1):1-20.

[24] WU Xiaohui, GUO Xiaodong, QIAO Zheng. Institutional Investor Grouping and Stock Price Crash Risk. China Industrial Economics, 2019(2): 117-135.

[25] SHI Yong, XU Mingyuan, LIU Yongqi, et al. Audit Report Information Content and Stock Price Crash Risk. China Soft Science, 2023(11): 201-212.

[26] KIMJ B, WANG Z, ZHANG L.CEO overconfidence and stock price crash risk. Contemporary accounting research,2016,33(4):1720-1749.

[27] GUO Wenwei, HUANG Zicong, HE Jie. Confucian Culture and Corporate Climate Change Risk Disclosure: Text Analysis and Machine Learning. China Journal of Economics, 2024, 11(2): 170-204.

[28] LIU Jianqiu, YIN Guangying, WU Jinghua. Corporate Social Responsibility Report Tone and Asset Mispricing. Accounting Research, 2022(5): 131-145.

[29] HUANG Pingping, LI Sihai. Corporate Social Responsibility Report Tone and Stock Price Crash Risk. Audit & Economy Research, 2020, 35(1): 69-78.

[30] DANG C, FOERSTER S, LI Z (Frank), 2021. Analyst talent, information, and insider trading. Journal of Corporate Finance, 67: 101803.

[31] LIU Gang, TANG Yin, YIN Jianling. The Impact Mechanism of Corporate Social Responsibility on Performance from the Perspective of Analyst Attention: A Study Based on Machine Learning and Text Analysis. China Soft Science, 2024(2): 156-166.

[32] PORTER M E, LINDE V D. Towards a new conception of environment-competitiveness relationship. The Journal of Economic Perspectives,1995,9 (4):97-118.

[33] WILLIAMSON O E. The new institutional economics: Taking stock, looking ahead. Journal of Economic Literature, 2000, 38(3): 595-613.

[34] CHEN Shiyi, CHEN Dengke. Haze Pollution, Government Governance, and High-quality Economic Development. Economic Research Journal, 2018, 53(2): 20-34.