International Journal of
Financial Studies

MDPI

Article

Data Asset Disclosure and Stock Price Crash Risk: A Double Machine Learning Study of Chinese A Share Firms

Junzuo Zhou [1], Zhaoyang Zhu [2], Huimeng Wang [3], Yuki Gong [4,5], Yuge Zhang [6,*] and Frank Li [4,6,*]

[1] Academic Affairs Office, Guangdong University of Finance and Economics, Guangzhou 510320, China; zhoujunzuo@gdufe.edu.cn
[2] School of Management, Xi'an Jiaotong University, Xi'an 710049, China; zzy123456@stu.xjtu.edu.cn
[3] College of Arts and Social Sciences, Australian National University, Canberra, ACT 2601, Australia; katherine.wang@anu.edu.au
[4] College of Computing, Georgia Institute of Technology, Atlanta, GA 30332, USA; yuki.so@gatech.edu
[5] College of Business and Economics, Australian National University, Canberra, ACT 2601, Australia
[6] College of Economics, Ocean University of China, Qingdao 266100, China
* Correspondence: zyg9039@stu.ouc.edu.cn (Y.Z.); sso31@gatech.edu (F.L.)

Abstract

In the digital economy, data assets have become key drivers of firm competitiveness and market stability. This study examines the association between data asset information disclosure and stock price crash risk. Using annual reports of Chinese A-share listed firms from 2010 to 2023, we construct a Data Asset Information Disclosure Index through textual analysis. A double machine learning framework is employed to flexibly control for high-dimensional confounders, and the results indicate that greater disclosure is associated with lower crash risk across multiple specifications. Generalized random forest analysis further highlights heterogeneous relationships, with disclosures on both internally used and transactional data assets showing stronger negative associations with crash risk. Mechanism evidence suggests that disclosure may facilitate information dissemination, strengthen investor confidence, and improve analyst forecast accuracy. The association is more pronounced in firms with weaker corporate governance, higher reporting transparency, more competitive industries, and in regions with less developed digital economies. An industry spillover pattern is also observed, whereby one firm's disclosure is linked to reduced crash risk among peers. Overall, this study contributes to the literature on data asset disclosure and corporate risk management by providing empirical evidence from a major emerging market and by highlighting the potential relevance of enhanced transparency for digital governance and capital market resilience.

Keywords: data asset information disclosure; stock price crash risk; double machine learning

check for updates

Academic Editor: Indranil SenGupta

Received: 23 October 2025
Revised: 23 November 2025
Accepted: 27 November 2025
Published: 2 December 2025

Citation: Zhou, J., Zhu, Z., Wang, H., Gong, Y., Zhang, Y., & Li, F. (2025). Data Asset Disclosure and Stock Price Crash Risk: A Double Machine Learning Study of Chinese A Share Firms. *International Journal of Financial Studies*, *13*(4), 229. https://doi.org/10.3390/ijfs13040229

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In the digital economy era, data assets have become a critical component of firms' core competitiveness globally. Internationally, data is increasingly recognized as a vital resource, with many countries advancing regulations to integrate data into financial reporting. In 2022, the International Accounting Standards Board (IASB) launched a project to revise its disclosure standards to accommodate data as an asset, acknowledging its growing economic significance. Similarly, the European Union and other international organizations are exploring frameworks to better account for data assets in corporate financial statements, recognizing their potential to drive innovation and long-term growth. These

efforts underscore the importance of clear standards for the disclosure and valuation of data resources.

In emerging markets like China, data assets are increasingly crucial for firms' competitiveness. In 2024, China's data production reached 41.06 zettabytes, with the digital economy contributing around 10% to GDP. However, challenges remain in recognizing and disclosing data assets. Despite these challenges, China has been proactive in addressing the recognition and disclosure of data assets. In response to global developments, the Ministry of Finance in China officially recognized data as an asset in 2023, aligning with the ongoing revisions by the IASB to establish clearer frameworks for data asset disclosure. This shift prompts important questions about how data asset disclosure influences market stability, especially in emerging markets.

Stock price crash risk, a key indicator of market volatility, has gained significant attention in recent years (Zhang et al., 2016). It not only causes substantial losses for investors but also weakens a firm's financing ability, threatening its operations and long-term growth. Moreover, a stock price crash can trigger a panic sell-off, impacting the entire market. Research indicates that information asymmetry is a key driver of stock price crash risk. When management conceals unfavorable information, the market may temporarily maintain an inflated price, but once this information is revealed, stock prices can become volatile and crash (Wang et al., 2024).

In this context, data assets play a crucial role in reducing stock price crash risk. As a novel category of assets, their quality and transparency in disclosure help markets assess a firm's true value more accurately. Unlike traditional physical assets, data assets have low marginal costs and significant economies of scale, allowing firms to create value with minimal resource input. Moreover, as data assets integrate into the capital market through mechanisms such as data ownership and creditization, their role in financial resource allocation becomes increasingly critical. High-quality data asset disclosure can mitigate stock price volatility by reducing market information asymmetry, fostering investor confidence, and curbing irrational sell-offs. Understanding how data asset disclosure affects stock price crash risk is therefore essential for investors, management, and policymakers alike.

Existing research provides a foundation for understanding data asset information disclosure, exploring its concept (Collins & Lanz, 2019; T. Xu et al., 2023), accounting methods (Jackson & Luu, 2023), and economic effects, such as alleviating financing constraints (He et al., 2024) and promoting operational efficiency (Hu et al., 2022). However, how data asset information disclosure impacts capital market stability remains underexplored.

International disclosure theories suggest that high-quality information improves market transparency (Bleck & Liu, 2007), with evidence linking better disclosure to reduced stock price crash risk (Kim & Zhang, 2014). Improved information environments and voluntary disclosure mitigate bad news accumulation, with non-financial disclosures like ESG performance (Bae et al., 2021; Murata & Hamori, 2021) similarly enhancing market stability. These findings hold across both developed and emerging markets, highlighting information transparency as a universal driver of market stability. This study extends this literature by examining data asset disclosure as a new digital-driven disclosure category. Unlike standardized financial or ESG information, data asset disclosure features lower standardization and weaker investor recognition, particularly in emerging markets like China, where digital growth outpaces regulatory development. By testing its effect on crash risk, we provide comparative insights into how this emerging disclosure category functions across institutional contexts, thereby advancing disclosure theory in the digital economy.

The main contributions of this study are as follows: (1) By empirically investigating the linkage between data asset information disclosure and stock price crash risk, this paper offers a novel perspective on the economic consequences of data asset information

disclosure—distinct from traditional "information disclosure–market risk" research. Unlike well-recognized, highly salient disclosure indicators (e.g., financial or ESG metrics) that attract broad investor attention and prompt clear reactions, data asset information disclosure is a less obvious, emerging category. General investors often lack clarity on its value, relevance and measurement, leading to uncertainty about incorporating it into decisions. This study fills this gap by identifying data asset information disclosure as a new, understudied factor of stock price crash risk. (2) This study develops a comprehensive theoretical framework that elucidates the mediating mechanisms and contextual conditions under which data asset disclosure influences stock price crash risk.

2. Theoretical Analysis and Research Hypotheses

2.1. The Impact of Data Asset Information Disclosure on the Stock Price Crash Risk

To systematically explain how data asset information disclosure mitigates stock price crash risk, we integrate signaling theory, behavioral finance theory, and the Resource-Based View (RBV) into a coherent framework. We address the "source of signal credibility", "behavioral response to signals", and "substantive value of the signal carrier (data assets)", respectively, rather than treating these elements as isolated components.

Rooted in the RBV, data assets are not ordinary resources but core intangible assets that generate sustained competitive advantage (He et al., 2024; Ghasemaghaei & Calic, 2020). Unlike physical assets with diminishing marginal returns, data assets have low marginal cost and scale economies: once collected and processed, they can support multiple business scenarios without proportional cost increases, helping firms stabilize operational performance and reduce firm-specific risks that may trigger stock crashes (Hu et al., 2022). This competitive value gives data asset disclosure substantive weight—when firms reveal details like data governance capabilities and application scenarios, they are not releasing "superficial information" but evidence of long-term profitability, distinguishing such disclosure from symbolic claims (e.g., vague "digital transformation" statements) and laying the groundwork for credible signaling.

In information-asymmetric capital markets, managers hold more firm value information than external investors (Kim & Zhang, 2014). Signaling theory notes that high-quality firms can stand out by sending costly, verifiable signals, and data asset disclosure fits this model. Comprehensive disclosure requires investments in data governance systems and third-party audits—costs that low-quality firms with weak data capabilities cannot bear—while investors can cross-validate disclosed content against real performance or industry benchmarks. This credible signaling directly tackles the core driver of stock crashes: the accumulation of hidden bad news (Friske et al., 2023). When firms proactively disclose the value of their data assets, insiders have less incentive to hoard negative information (e.g., data security risks, inefficient utilization), as the market already has transparency into the firm's digital capabilities, preventing the "sudden release of accumulated bad news" that often triggers crashes.

Even with credible signals, investor cognitive biases (e.g., overreaction, herding) can amplify volatility (Fu et al., 2021). Behavioral finance theory explains how data asset disclosure acts as a "behavioral anchor" for investor expectations. In opaque environments, investors rely on speculative information (e.g., rumors about digital failures) and overreact to negatives, leading to panic selling. Data asset disclosure, by contrast, provides concrete, verifiable details about a firm's digital value, anchoring expectations to its actual performance. Confident investors, supported by transparent data, are also less likely to herd: during market downturns, they recognize the firm's long-term advantages and avoid irrational sell-offs, thus reducing synchronized price collapses (Li et al., 2022).

These three theories reinforce each other: the RBV confirms the substantive value of data assets (making disclosure worthwhile), signaling theory ensures that this value is credibly transmitted (alleviating information asymmetry), and behavioral finance theory explains how transparency stabilizes investor behavior (curbing irrational volatility). Without the RBV, data asset disclosure would lack substance; without signaling theory, its value would stay hidden; without behavioral finance, transparency might not translate to stability. Based on the above analysis, the article proposes hypothesis H1:

H1: *Data asset information disclosure is negatively associated with stock price crash risk.*

2.2. Mechanism Analysis of How Data Asset Information Disclosure Affects Stock Price Crash Risk

2.2.1. Information Transmission Mechanism

The core logic of data asset information disclosure, mitigating stock price crash risk, lies in its role as an effective information transmission bridge. As Kim and Zhang (2014) emphasized, intensified information asymmetry enables managers to conceal adverse information, and the sudden release of accumulated bad news triggers stock price crashes. Data asset information disclosure directly addresses this dilemma by revealing key details of data collection, processing, application, and value creation—supplementing both financial and non-financial decision-making information for external investors.

This disclosure reduces information asymmetry through dual channels: on the one hand, it provides market participants with a comprehensive perspective to assess firm operations and growth potential, enabling stock prices to more accurately reflect intrinsic value and reducing volatility caused by information gaps (Xie et al., 2023); on the other hand, it raises the cost of managerial opportunism, curbing the deliberate concealment or delayed disclosure of negative information related to data assets and preventing the accumulation of bad news (Luo, 2025). Additionally, proactive disclosure by leading firms forms a demonstration effect, fostering industry-wide transparency and stabilizing market expectations.

In summary, data asset information disclosure alleviates information asymmetry between insiders and outsiders, restricts managerial opportunistic behaviors, and blocks the key channel of bad news accumulation—thus reducing stock price crash risk.

2.2.2. Investor Confidence Mechanism

As developed in our baseline hypothesis, data asset disclosure mitigates crash risk by alleviating information asymmetry and preempting the buildup of hidden bad news. We further argue that this effect is channeled, in part, through the reinforcement of investor confidence—a mechanism grounded in behavioral finance.

In opaque information environments, investor anxiety and uncertainty prevail, which amplify cognitive biases such as overreaction and herding (Fu et al., 2021). Transparent disclosure of data assets—by quantifying scale, application, and value—serves as a verifiable signal that not only clarifies a firm's fundamental standing but also actively cultivates investor confidence in its governance and future prospects (Li et al., 2022). Confident investors are less prone to interpret ambiguous news negatively and more likely to maintain stable holdings during periods of market stress.

This sustained commitment reduces coordinated sell-offs and dampens volatility, thereby directly countering the liquidity shocks and sentiment-driven crashes often precipitated by collective panic. Thus, by strengthening investor confidence, data asset disclosure stabilizes shareholder behavior and contributes to a lower probability of stock price crash risk.

Based on the above analysis, we argue that data asset disclosure enhances investor confidence in a company's governance and long-term potential, alleviates information

asymmetry and cognitive biases, thereby reducing market overreaction and the risk of stock price crashes.

2.2.3. Analyst Forecast Accuracy Mechanism

We further propose that data asset information disclosure reduces stock price crash risk by enhancing the accuracy of analyst forecasts, a channel that bridges the perspectives of market efficiency and behavioral finance.

In capital markets characterized by information asymmetry, analysts play a critical role in information interpretation and dissemination. High-quality disclosure of data assets—encompassing their scale, application, and strategic relevance—provides analysts with a more concrete foundation for evaluating a firm's intrinsic value and growth trajectory (X. L. Gong & Liu, 2023). This reduces reliance on speculative or incomplete information, thereby lowering forecast dispersion and improving forecast consensus.

Improved forecast accuracy, in turn, stabilizes market expectations and mitigates two key sources of crash risk. First, it limits the scope for extreme earnings surprises, which are often triggers of abrupt price corrections. Second, by offering reliable earnings guidance, analysts help anchor investor judgments, reducing their susceptibility to cognitive biases such as overreaction to ambiguous news—a behavioral tendency that can amplify sell-offs. Consistent with this, prior research confirms that greater forecast accuracy dampens irrational volatility and lowers crash risk (Yao et al., 2024).

Thus, by equipping analysts with richer and more reliable information, data asset disclosure enhances the informational efficiency of prices and restrains the behavioral overreactions that precipitate crashes

Based on the above analysis, this paper proposes the following hypothesis:

H2: *Data asset information disclosure is negatively associated with corporate stock price crash risk through (a) information transparency, (b) investor confidence, and (c) analyst forecast accuracy.*

While we propose that data asset disclosure is generally beneficial, prior literature also raises two important counterarguments that merit consideration. A first perspective concerns the possibility of selective disclosure and misleading information. Some critics argue that data asset disclosure may increase stock price crash risk when firms reveal information strategically or provide incomplete details. Because managers often possess substantial discretion over the content of disclosure, firms may highlight only favorable aspects of their data assets while omitting critical information such as data security vulnerabilities or operational inefficiencies. Such selective reporting can mislead investors and heighten information asymmetry (Kim & Zhang, 2014). When erroneous or incomplete information shapes market consensus, stock prices may deviate from their fundamental value (Epstein & Schneider, 2008). Once the concealed information eventually surfaces, prices may adjust sharply, triggering a crash. Thus, although disclosure is intended to alleviate information asymmetry, it may in certain contexts exacerbate volatility by offering misleading signals that distort investor expectations.

A second perspective concerns the risk of information overload. Excessive or poorly structured disclosure may overwhelm investors and reduce their ability to process information effectively. This issue is more pronounced when disclosures are unstructured, low in quality, or insufficiently relevant, which can create confusion and impair decision-making. Our study, however, focuses on structured and verifiable data asset disclosure that is specific and directly related to the firm's value creation. Such structured disclosures provide clear and actionable information that reduces uncertainty and improves the efficiency of investor information processing. By presenting coherent and relevant insights, they avoid the confusion associated with information overload and support more accurate assessments

of firm value. Moreover, high-quality and well-organised disclosures filter out market noise by emphasising value-relevant information rather than contributing to excessive informational complexity.

These competing views do not contradict our theoretical framework but instead highlight the boundary conditions for effective disclosure. To prevent selective reporting and information overload, disclosures must be non-proprietary, structured, and verifiable, avoiding the release of sensitive details while ensuring clarity and relevance for investors. These boundary conditions are built into our measurement approach, which focuses on non-proprietary and observable data asset details, such as application scenarios, rather than proprietary information or vague qualitative descriptions.

3. Research Design

3.1. Model Construction

This paper aims to examine the impact of data asset information disclosure on stock price crash risk. Current research in this field predominantly employs traditional linear regression models (e.g., OLS) for empirical analysis. This study utilizes a double machine learning (DML) framework to examine the relation between data asset information disclosure and stock price crash risk, addressing key limitations of OLS and enhancing the reliability of results (Yin et al., 2025; Ling et al., 2024).

In functional form flexibility, OLS inappropriately assumes a linear relationship between data asset information disclosure (DA) and stock price crash risk—though their actual link is often nonlinear (e.g., marginal crash risk reductions may diminish at a disclosure threshold or vary by firm governance). DML avoids this via machine learning (e.g., gradient boosting) to flexibly model complex relationships between controls and outcome/treatment variables, preventing misspecification bias. For confounding variables, OLS relies on manual selection and assumes observable, correctly specified key confounders, but unobservable factors (e.g., corporate culture, implicit data governance) behind crash risk cause endogeneity; DML mitigates omitted-variable concerns by automatically capturing high-dimensional, nonlinear confounding via algorithmic feature selection, cutting omitted variable bias and improving the robustness of the estimated associations.

Economically, OLS estimates a homogeneous average treatment effect (ATE) that hides heterogeneous responses (e.g., stronger effects for weakly governed firms), while DML not only isolates DA's net effect for a more reliable ATE but also enables heterogeneous treatment effect (HTE) estimation (e.g., via generalized random forests)—aligning with intuition that firms with different internal governance and external environments react differently to disclosure, as data transparency may substitute weak supervision or complement competitive pressures. These advantages of DML have been corroborated in prior business-related empirical studies: Yin et al. (2025) leveraged DML to explore the relation between artificial intelligence and climate risk, and Ling et al. (2024) adopted the framework to analyze transportation infrastructure upgrading and green development efficiency, both demonstrating its robustness in capturing complex relationships for economic outcomes.

Based on this, this paper refers to the classic research paradigm of Bodory et al. (2022) and employs a double machine learning partial linear model for empirical testing, with the specific configuration as follows:

$$Y_{it+1} = \theta_0 D_{it} + g(X_{it}) + U_{it} \tag{1}$$

$$E(\varepsilon_{it}|D_{it}, X_{it}) = 0 \tag{2}$$

Y_{it+1} is the explained variable stock price crash risk, where t represents the firm and t denotes the year. D represents the explanatory variable, which is the level of disclosure

of corporate data assets. The coefficient θ_0 is the regression coefficient of the primary focus in this paper. X_{it} denotes the set of control variables, which may encompass all confounding factors simultaneously influencing both the outcome variable Y_{it+1} and the treatment variable D_{it}. The specific effect function $g(X_{it})$ of these confounding factors remains unknown. Machine learning algorithms are employed to estimate this, yielding $\hat{g}(X_{it})$. $\hat{\theta}_0$ represents the error term, with a conditional mean of 0.

If we directly employ machine learning algorithms to estimate $g(X_{it})$ in the model, the result may obtain a biased estimate $\hat{\theta}_0$ due to regularization bias during large-scale data processing. To circumvent this issue, we further construct an auxiliary regression model:

$$D_{it} = m(X_{it}) + V_{it} \tag{3}$$

$$E(V_{it}|X_{it}) = 0 \tag{4}$$

The function form of $m(X_{it})$ is unknown, with V_{it} representing the error term and having a conditional mean of 0. Machine learning algorithms are employed to estimate $\hat{m}(X_{it})$ and construct residual estimates $\hat{V}_{it} = D_{it} - \hat{m}(X_{it})$. Subsequently, the same algorithms are used to estimate $\hat{g}(X_{it})$ in the main regression, with outcome $Y_{it} - \hat{g}(X_{it}) = \theta_0 D_{it} + U_{it}$. Treating $\hat{V}_{it}$ as an instrumental variable for D_{it}, regression is performed to obtain coefficient estimates as follows:

$$\hat{\theta}_0 = (1/n \sum_{i\in I, t\in T} \hat{V}_{it} D_{it})^{-1} \, 1/n \sum_{i\in I, t\in T} \hat{V}_{it}(Y_{it} - \hat{g}(X_{it})) \tag{5}$$

During the regression process, the convergence rate of $g(X_{it})$ and $m(X_{it})$ toward their true functions $\hat{g}(X_{it})$ and $\hat{m}(X_{it})$ will influence the convergence rate of $\hat{\theta}_0$. By performing two machine learning estimations, we can not only eliminate the impact of D_{it} caused by the confounding variable set X_{it} but also accelerate the convergence of $\hat{\theta}_0$, thereby obtaining more accurate estimates with a finite sample size. To enhance the stability and reliability of model estimation, this paper employs five-fold cross-validation to process regression samples. Concurrently, resampling tests are utilized for robustness verification.

3.2. Variable Definition

3.2.1. Dependent Variable: Stock Price Crash Risk

Following the methodology of Kim and Zhang (2016) and Y. Gong et al. (2025), this paper employs the Negative Return Skewness (*NCSKEW*) and the Down-to-Up Volatility Ratio (*DUVOL*) to measure stock price crash risk. Higher values of these indicators indicate greater exposure to stock price crash risk. Considering the time-sensitive impact of data asset information disclosure on stock prices, both *NCSKEW* and *DUVOL* indicators are calculated for the $t + 1$ period.

3.2.2. Independent Variable: Data Asset Information Disclosure (*DA*)

Quantitative measurement of corporate data asset information disclosure remains a relatively new field in both academia and practice. Consequently, methods for measuring data asset information disclosure indicators face several limitations. Some studies directly utilize existing financial data for calculations (Shi et al., 2025). While this approach offers readily available data and a straightforward computational process, the resulting indicators may contain excessive noise, failing to accurately reflect the true state of data assets and leading to unstable empirical results. Furthermore, most studies posit that data assets embody critical strategic decisions and future outlooks, information that is more readily reflected in annual reports. However, existing methods may omit significant terms during seed word construction and lack objective policy documents or research reports as textual foundations (Hu et al., 2022; Qian et al., 2025). This results in incomplete sets of

synonyms for seed words, potentially introducing subjective biases and causing substantial discrepancies between the word sets and actual conditions.

To address these issues, this study draws on the latest research, including the work of Luo (2025) and He et al. (2024), and employs a comprehensive measurement methodology to evaluate the extent of data asset information disclosure for Chinese A-share listed firms from 2010 to 2023. This approach is designed to provide a more objective and comprehensive data asset indicator by integrating legal and regulatory texts related to data assets, thereby enhancing the accuracy and completeness of the measurement process.

The first step involves identifying seed words and constructing a regulatory corpus. Four core terms—"information", "network", "digital", and "data"—were selected as seed words, based on the conceptual definition of data assets and their direct relevance to data asset characteristics. To ensure objectivity, a corpus of legal and regulatory texts was compiled from national and local levels, covering data security, fundamental data systems, data ownership, and data asset accounting treatment (e.g., Cybersecurity Law, Data Security Law, Interim Measures for Accounting Treatment of Corporate Data Resources). This corpus served as the textual foundation for expanding the seed word set.

The second step applies the Word2Vec neural network model using the Skip-Gram architecture. The regulatory corpus was tokenized and converted into high-dimensional word vectors. Cosine similarity between these vectors was calculated to expand the seed word set: terms with a similarity threshold ≥ 0.7 (validated via expert review to avoid irrelevance) were retained, while terms with negative connotations (e.g., "non-digital", "data leakage risk") were manually excluded to ensure the expanded set accurately reflects data asset-related content.

The final step calculates the frequency of the expanded keyword set within annual reports of A-share listed firms between 2010 and 2023. Total frequency was normalized by dividing by 100—this step compresses index values into a manageable range and simplifies coefficient interpretation (a 1-unit increase in *DA* corresponds to a 100-fold rise in keyword frequency, making economic significance more intuitive). The normalized value is the core *DA* index, with higher values indicating more comprehensive disclosure.

To capture heterogeneous effects of different data asset types, we further classified data assets into transactional data assets (*DDA*) and operational data assets (*ODA*) based on their usage scenarios:

Transactional Data Assets (*DDA*): Reflect market-oriented data activities (e.g., "data transaction", "data service platform", "data circulation"), with keywords screened to represent data monetization and external value exchange.

Operational Data Assets (*ODA*): Reflect internal data application (e.g., "data-driven decision-making", "digital operations", "data management"), with keywords focusing on data support for business processes and efficiency improvement.

The frequency of *DDA* and *ODA* keywords was calculated and normalized using the same method as the core *DA* index. For detailed keywords, similarity thresholds, and text preprocessing steps (e.g., stopword removal, tokenization), refer to Appendix A.

While our initial analysis relied on an automated text mining approach, we recognized the potential biases inherent in this method. To address these concerns, we implemented additional verification steps. Specifically, we manually validated a representative subset of the data and compared the automated results with expert judgments. This process involved expert coders reviewing and categorizing the same data, ensuring that the classifications were consistent with the automated system's output. We also conducted a benchmark comparison with established methods in the field to verify that our results aligned with those obtained through traditional manual approaches. The results of the expert coding

were found to be highly consistent with the automated analysis, further validating the reliability of our indicators.

Considering the concern that our index might overemphasize "digital speak" over substantive disclosure—despite refinements like incorporating legal texts—we validated it using two correlations, following F. Xu et al. (2025). First, we linked the data asset disclosure index to annual report disclosure quality, as higher-quality reporting suggests less reliance on rhetoric. Second, we correlated it with firm-level digital inputs and outcomes, assuming that substantive disclosure should reflect actual digital activities. Both correlations showed significant positive associations, confirming that the index captures meaningful content beyond superficial terms.

3.2.3. Control Variables

In selecting control variables, this paper draws upon the research of Kim and Zhang (2016) and Zhang et al. (2016). Based on the corporate financial condition, Based on corporate financial health, governance practices, and firm characteristics, the following control variables influencing stock crash risk were chosen: Firm Size (*Size*), Leverage Ratio (*Lev*), Return on Assets (*ROA*), Cash Flow Ratio (*Cashflow*), Fixed Assets Ratio (*FIXED*), Book-to-Market Ratio (*BM*), Revenue Growth Rate (*Growth*), Board Size (*Board*), CEO Duality (*Dual*), Top 1 Shareholding Ratio (*Top1*), Big Four Auditor (*Big4*), Stock Return (*Ret*), Stock Trading Turnover (*Dturn*), Stock Return Volatility (*Sigma*). Furthermore, to comprehensively capture potential nonlinear relationships among variables, this paper incorporates quadratic terms of control variables to enhance model fit. Additionally, to further control for industry and time-period differences, industry fixed effects and time fixed effects are introduced. Variable definitions are presented in Table 1.

Table 1. Variable Definitions.

Variable Type	Variable Name	Symbol	Definition
Dependent Variable	Stock Price Crash Risk	*NCSKEW*	Proxy for stock price crash risk, measured by the negative conditional skewness of firm i's weekly returns in year $t + 1$.
		DUVOL	Proxy for stock price crash risk, measured by the down-to-up volatility ratio of firm i's weekly returns in year $t + 1$.
Independent Variable	Data Asset Information Disclosure	*DA*	Index of data asset information disclosure, measured through text analysis.
Control Variables	Firm Size	*Size*	Natural logarithm of total assets.
	Stock Return	*Ret*	Average weekly stock return of the firm.
	Stock Turnover	*Dturn*	Average annual excess turnover rate of the firm's stock.
	Stock Volatility	*Sigma*	Standard deviation of weekly stock returns.
	Leverage Ratio	*Lev*	Total liabilities/Total assets.
	Return on Assets	*ROA*	Net profit/Total assets at year-end.
	Cash Flow Ratio	*Cashflow*	Net cash flow from operating activities/Total assets.
	Fixed Assets Ratio	*FIXED*	Net fixed assets/Total assets.
	Book-to-Market Ratio	*BM*	Book value/Market value.
	Revenue Growth Rate	*Growth*	(Current year revenue/Previous year revenue) − 1.
	Board Size	*Board*	Natural logarithm of the number of board members.
	CEO Duality	*Dual*	Equals 1 if the chairman and CEO are the same person, and 0 otherwise.
	Ownership Concentration	*Top1*	Number of shares held by the largest shareholder/Total number of shares
	Big Four Auditor	*Big4*	Equals 1 if audited is a Big Four accounting firm, and 0 otherwise.

3.3. Sample Selection and Data Sources

This study selected data from A-share listed companies between 2010 and 2023 as the initial sample. To ensure sample accuracy and validity, the following screening criteria were applied: (1) Exclusion of ST, *ST, and PT-classified companies, as well as those in the financial and insurance sectors; (2) Exclusion of samples with debt ratios exceeding 100%; (3) Samples with anomalous or missing data for relevant variables were excluded; (4) Annual observations with fewer than 30 trading weeks per year were excluded. Following these procedures, annual observations for 30,157 enterprises were ultimately obtained. Following the aforementioned procedures, annual observational data for 30,157 enterprises were ultimately obtained. To mitigate the potential bias and impact of extreme values on the test results, we winsorize continuous variables at the 1% and 99% levels. Data concerning the level of digital economic development were sourced from the China Urban Statistical Yearbook, whilst other financial-related data were drawn from the CSMAR database, Wind database, and Choice database.

4. Empirical Findings and Analysis

4.1. Descriptive Statistics

The descriptive statistics for key variables are presented in Table 2. The indicators reflecting stock price crash risk, *FNCSKEW* and *FDUVOL*, yielded mean values of −0.3313 and −0.2138, respectively, with medians of −0.2868 and −0.2140. These findings broadly align with existing literature, indicating substantial variation in corporate stock price crash risk metrics. The minimum and maximum values for the data asset information disclosure indicator (*DA*) are 0.04 and 3.68, respectively, demonstrating considerable disparity. This indicates significant variations in the level of data asset information disclosure across different enterprises, reflecting differing stages in their data assetization processes.

Table 2. Descriptive Statistics.

Variable	N	Mean	Std	Median	VIF
FDUVOL	30,157	−0.2138	0.4765	−0.2140	
FNCSKEW	30,157	−0.3313	0.7319	−0.2868	
DA	30,157	0.3443	0.4297	0.2300	1.11
Size	30,157	22.2663	1.2885	22.0705	2.23
Lev	30,157	0.4239	0.2024	0.4176	1.68
ROA	30,157	0.0413	0.0634	0.0388	1.65
Cashflow	30,157	0.0477	0.0688	0.0464	1.35
FIXED	30,157	0.2086	0.1573	0.1768	1.18
Growth	30,157	0.1683	0.3912	0.1081	1.12
Board	30,157	2.1220	0.1970	2.1972	1.12
Dual	30,157	0.2788	0.4484	0.0000	1.07
Top1	30,157	0.3406	0.1483	0.3190	1.18
Big4	30,157	0.0604	0.2383	0.0000	1.15
Dturn1	30,157	1.7310	1.3046	1.3759	1.97
BM	30,157	0.6236	0.2543	0.6207	1.93
Sigm	30,157	0.0628	0.0238	0.0582	2.15
Ret	30,157	0.0022	0.0091	0.0011	1.62

Additionally, we conducted a variance inflation factor (VIF) test to assess multicollinearity among the variables. The results show that the maximum VIF value across all variables is 2.23, which is well below the commonly accepted threshold of 10. This indicates that there is no significant multicollinearity issue in the model, ensuring the reliability of the subsequent regression results.

4.2. Analysis of Benchmark Regression Results

To ensure robustness of results, this paper first employs an OLS regression model to analyze the impact of data asset information disclosure on stock price crash risk. Findings reveal that data asset information disclosure is significantly negatively correlated with stock price crash risk. For the double machine learning (DML) model, we adopt a 1:4 sample split ratio and gradient boosting algorithm to improve predictive accuracy. Models (5) and (6) further introduce quadratic terms for control variables, allowing for the capture of nonlinear relationships. Table 3 reports these results. The coefficients for *DA* in these models are −0.0379 and −0.0527, both significant at the 1% level. This means that a one-unit increase in the data asset disclosure index is associated with a significant negative correlation with stock price crash risk by 3.79% in Model (5) and 5.27% in Model (6). These results are consistent with our theoretical expectation that data asset information disclosure is significantly negatively correlated with stock price crash risk. The negative coefficients indicate that greater transparency—by revealing verifiable information about a firm's data assets—helps stabilize investor expectations. This transparency reduces the likelihood of irrational market reactions, such as panic selling or speculative price movements, which are common triggers of stock price crashes. In sum, these findings support our hypothesis (H1).

Table 3. Benchmark Regression Results.

Variable	(1) Ols *FDUVOL*	(2) Ols *FNCSKEW*	(3) Machine Learning (DML) *FDUVOL*	(4) DML *FNCSKEW*	(5) DML *FDUVOL*	(6) DML *FNCSKEW*
DA	−0.0388 ***	−0.0542 ***	−0.0372 ***	−0.0525 ***	−0.0379 ***	−0.0527 ***
	(0.008)	(0.013)	(0.009)	(0.013)	(0.009)	(0.013)
Constant	−0.1125	−0.6492 ***	0.0001	0.0001	−0.0001	0.0001
	(0.071)	(0.108)	(0.003)	(0.004)	(0.003)	(0.004)
Control Variable First-Order Term	YES	YES	YES	YES	YES	YES
Control Variable Second-Order Term	NO	NO	NO	NO	YES	YES
Time Fixed Effect	YES	YES	YES	YES	YES	YES
Industry Fixed Effect	YES	YES	YES	YES	YES	YES
N	30,157	30,157	30,157	30,157	30,157	30,157

Notes: standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4.3. An Examination of the Individual Average Effect of Data Asset Information Disclosure on Stock Price Collapse Risk

Benchmark regression analysis reveals that data assets information disclosure exerts a significant negative influence on the stock price crash risk, meaning that higher levels of disclosure correlate with lower crash risks. However, while benchmark regression analysis provides overall trends and patterns, it may fail to fully reveal significant variations in disclosure effectiveness across different companies or time periods. Such discrepancies may stem from firm-specific characteristics—such as governance structures, market positioning, or industry-specific traits—which could cause identical disclosure levels to yield differing impacts across firms or for the same firm at different junctures. To explore this idiosyncratic effect more deeply, this paper introduces the Generalized Random Forests (GRF) model. As a cutting-edge method in machine learning, GRF offers considerable advantages in capturing data heterogeneity. Employing the GRF model enables a more precise assessment of the idiosyncratic effect of information disclosure on stock price crash risk.

Figure 1 visually illustrates the distribution of individual intervention effects of data asset information disclosure on stock price crash risk when the number of trees is 1000 and 4000, respectively. The findings reveal a certain degree of differentiation in the average individual disposal effect. Although a minority of firms exhibit disposal effects exceeding 0, the majority concentrate within the range of 0 to −0.1. This outcome aligns with the

negative impact observed in the benchmark regression, indicating that enhanced quality of data asset information disclosure generally exerts a positive influence in mitigating the stock price crash risk.

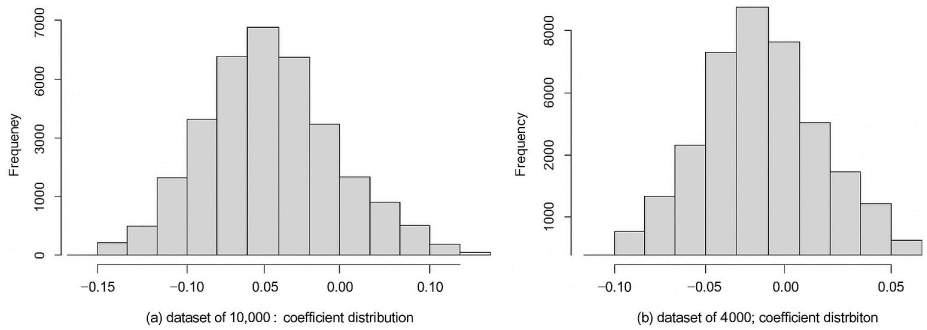


Figure 1. Distribution of Individual Disposition Effects of Data asset information disclosure on Share Price Crash Risk (Upper *FDUVOL*, Lower *FNCSKEW*).

4.4. Distinguishing Between Different Types of Data Asset Information Disclosure

To better understand the impact of data asset information disclosure on stock price crash risk, this paper distinguishes between two types of disclosure: Operational Data Asset Disclosure (*ODA*) and Data for Direct Application (Self-use Data Assets) (*DDA*). *ODA* focuses on the market transactions and value realization of data assets, while *DDA* emphasizes the internal use of data to enhance operational efficiency and decision-making. By categorizing these types, we can more precisely evaluate their effects on stock price crash risk.

The results in Table 4 indicate that both types of data asset disclosure are significantly negatively correlated with stock price crash risk at the 1% level. Disclosing operational data assets enhances market recognition of their value, while disclosing self-use data assets improves transparency regarding internal operations, which stabilizes investor expectations and reduces market concerns. This transparency helps mitigate stock price crash risk.

Table 4. Testing the Impact of Different Types of Data asset information disclosure on Stock Price Crash Risk.

Variable	(1) *FDUVOL*	(2) *FNCSKEW*	(3) *FDUVOL*	(4) *FNCSKEW*
ODA	−0.0448 ***	−0.0611 ***		
	(0.010)	(0.015)		
DDA			−0.1417 ***	−0.2558 ***
			(0.053)	(0.082)
Constant	−0.0001	0.0001	−0.0001	0.0001
	(0.003)	(0.004)	(0.003)	(0.004)
Control Variable First-Order Term	YES	YES	YES	YES
Control Variable Second-Order Term	YES	YES	YES	YES
Time Fixed Effect	YES	YES	YES	YES
Industry Fixed Effect	YES	YES	YES	YES
N	30,157	30,157	30,157	30,157

Notes: standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4.5. Robustness Test

4.5.1. Adjusting the Research Sample

Given that the capital market in 2015 may have been severely impacted by the financial crisis or the stock price crash, the sample observations for that year were excluded to eliminate potential distortions in stock price crash risk analysis caused by these extreme market events.

Given that municipalities directly under the central government may exhibit distinct characteristics in data asset information disclosure and stock price crash risk due to policy preferences and unique economic structures, this paper excludes samples from the four major municipalities directly under the central government to minimize the interference of regional heterogeneity with findings. Results indicate that regardless of the sub-sample regression method employed, all test outcomes remain substantively consistent with the benchmark regression, with *DA*'s regression coefficient consistently significant at the 1% level.

4.5.2. Eliminate the Impact of Outliers

This paper ensures the robustness of results by eliminating the influence of outliers, specifically removing the lowest 5% and highest 3% of observations to mitigate potential distortion of regression outcomes by extreme values. Following re-regression, the negative correlation between data asset information disclosure and stock price crash risk remains consistently present, indicating that the research findings are robust and unchanged by the treatment of outliers.

4.5.3. Consider the Region–Time Interaction Fixed Effects

Given that both the year and the region play significant roles across multiple dimensions—including government governance, economic development, and socio-cultural factors—they often exert profound influences on various economic and social phenomena. In particular, cities within the same province or different regions within the same year may face similar macroeconomic environments, policy orientations, and social transitions, all of which could exert a joint effect on the research subjects. Therefore, this paper incorporates an interaction fixed effect for year and province and conducts a re-examination. The findings indicate that the regression results remain robust even after incorporating this stronger control effect.

4.5.4. Reset the Dual Machine Learning Model

To ensure that the configuration of dual machine learning models does not mislead research findings, this study implemented a series of measures to guarantee the robustness of conclusions. Firstly, the sample partitioning ratio was adjusted from the initial 1:4 to 1:2, 1:3, and 1:7 to assess the potential impact of varying sample sizes on outcomes. Additionally, we replaced the algorithm, substituting gradient boosting with logistic regression, random forests, neural networks, support vector machines, and elastic net models to assess the impact of different predictive methods on the results. Despite these changes, the regression outcomes remained consistent, further confirming the reliability and robustness of the findings.

4.5.5. Reset the Main Regression Model

In benchmark regression analysis, this paper employs dual machine learning techniques to construct partial linear models for evaluating the relationship between data asset information disclosure and stock price crash risk. However, the model construction process may be subject to a degree of subjective judgment. To validate the robustness of research conclusions, this paper further utilizes dual machine learning to construct a more generalized interaction model, thereby exploring the potential impact of model specification on research findings. The primary and auxiliary regressions analyzed are modified as follows:

$$Y_{it+1} = g(D_{it}, X_{it}) + U_{it} \tag{6}$$

$$D_{it} = m(X_{it}) + V_{it} \tag{7}$$

Interactive models require binary discrete variables as independent variables. Therefore, when conducting robustness tests on the relationship between data asset information disclosure and stock price crash risk, this paper first converts continuous data asset information disclosure variables into binary form. Specifically, based on the median for each year and industry, companies with data asset information disclosure levels above the median are coded as 1, while those below the median are coded as 0, thereby creating a binary variable. By introducing an interactive model, we can more precisely capture the nonlinear impact of data asset information disclosure on stock price crash risk and test the robustness of the benchmark regression model's results. The regression results show that after introducing the interactive model, the regression coefficient for data asset information disclosure remains significantly negative, and the benchmark results remain unchanged.

4.5.6. Control the Influence of Potential Variables

Excluding potential policy interference. When examining the impact of data asset information disclosure on stock price crash risk, the conclusions of this paper may be subject to external interference, particularly from macro-policy shifts that could influence both data asset information disclosure and stock price performance. Examples include the Broadband China Strategy, reducing information asymmetry, and public data openness policies, enhancing data transparency and accessibility. To exclude the potential effects of these policy changes, the study controlled for the impact of these two policies within the model. Research findings indicate that, after accounting for potential policy interference, the regression coefficient for data asset information disclosure remains significantly negative. This further confirms the mitigating effect of data asset information disclosure on stock price crash risk.

Excluding alternative explanations. Digital strategy commitments and digital transformation may influence stock price crash risk. These factors collectively influence corporate stock performance and market stability. Therefore, to exclude potential confounding effects from digital transformation and digital strategy commitments, this study controlled for digital transformation and digital strategy commitment indicators in the regression. Test results demonstrate that, after controlling for these alternative variables, data asset information disclosure remains effective in mitigating financial market risks and suppressing corporate stock crash risk.

4.5.7. Replace Variable

Firstly, the method for measuring the dependent variable—stock price crash risk—was altered. The crash risk indicators were recalculated using the market-adjusted market capitalization average method (*NCSKEW_Mdeq*; *DUVOL_Mdeq*) and the market-adjusted total market capitalization average method (*NCSKEW_Mdos*; *DUVOL_Mdos*), then substituted into the original model for regression. The results remained unchanged. Secondly, the method for measuring the explanatory variable of data asset information disclosure was altered. The benchmark regression applied a scaling process by dividing the data asset information disclosure indicator by 100. In response, this paper further subjected the original indicator to natural logarithm transformation, with the results remaining unchanged. Concurrently, the seed vocabulary was expanded. 'Data assets,' 'data resources,' and 'data elements' were designated as seed terms, from which high-similarity words were screened using deep learning similarity functions. Subsequently, the frequency of these seed terms and their synonyms within the listed companies' annual financial reports was extracted from the database. Their proportion relative to the total report word count was calculated to serve as a metric for data asset information disclosure. To mitigate potential bias from

excessively large indicators, the metrics were log-transformed and divided by 100. The regression results remained unchanged.

4.5.8. Endogeneity Issues

Although this study mitigates potential bidirectional causality issues by using a t + 1-period stock price crash risk indicator and including relevant variables, there may still be reverse causality, where stock price crash risk influences a firm's decision to disclose data asset information. For example, firms facing higher stock price crash risk may have a stronger incentive to disclose more data asset information to stabilize investor confidence. On the other hand, if stock prices are stable, firms may reduce disclosure to minimize costs or protect proprietary information. To address this potential reverse causality, the study adopts a dual machine learning-based instrumental variables (IV) model, as detailed below:

$$Y_{it+1} = \theta_0 D_{it} + g(X_{it}) + U_{it} \tag{8}$$

$$IV_{it} = m(X_{it}) + V_{it} \tag{9}$$

Among these, IV_{it} serves as the instrumental variable for the data asset information disclosure indicator. This paper adopts Lewbel's (1997) approach by generating the instrumental variable through the cube of the deviation between the original data asset information disclosure indicator and its industry mean. The advantage of this method lies in its avoidance of seeking a conventional exogenous instrumental variable, instead creating the instrumental variable through a transformation of the endogenous variable itself. The instrumental variable captures the degree of deviation in each firm's data asset information disclosure level relative to other firms within its industry. This deviation is likely correlated with the firm's own disclosure behavior, thereby satisfying the relevance principle. Furthermore, as a deviation from the industry mean, this instrumental variable eliminates unobservable factors common across the industry, making it less susceptible to direct influence from stock price crash risk. Moreover, as it represents a relative deviation within the industry, it is unlikely to incorporate other unobserved factors affecting stock price crash risk, thus satisfying the requirement of exogeneity.

In addition to the Lewbel-type instrumental variable, we introduce the number of executives with a digital background as an additional instrumental variable. This is because executives with a digital background are more likely to understand and prioritize data assets as key strategic resources. Their expertise could influence the firm's data asset disclosure practices. Therefore, the number of executives with a digital background should be positively correlated with the level of data asset disclosure, satisfying the relevance condition. Importantly, while the number of executives with a digital background affects the firm's data disclosure behavior, it is unlikely to have a direct effect on stock price crash risk. The mechanism through which this variable operates is indirect—by influencing how data assets are disclosed, it affects market transparency and reduces information asymmetry, which can lower stock price crash risk. This ensures that the executive digital background variable satisfies the exogeneity condition, as it affects data asset disclosure without directly affecting stock price crash risk.

Collectively, all results from the instrumental variable validity tests and the subsequent causal effect estimation align in showing a significant negative correlation, which verifies that higher data asset information disclosure is associated with a reduction in stock price crash risk.

To further address selection bias arising from observable confounding factors, we conduct propensity score matching (PSM): the continuous DA index is converted into a binary variable (1 for above the annual industry median, 0 otherwise), and 1:1 nearest

neighbor matching (caliper = 0.05) is performed using covariates including firm size, leverage, profitability, and corporate governance indicators. Post-matching covariate balance is achieved, and regression results on the matched sample show the binary *DA* indicator is significantly negative, confirming that the conclusion that data asset disclosure reduces stock price crash risk remains robust to observable confounding and selection bias.

4.5.9. Median-Based Resampling Test

In the dual machine learning model, a sample partitioning approach was employed to estimate the treatment coefficient. Specifically, the sample was divided into training and test sets according to a predetermined ratio. The model was then fitted using the training set, with predictions made on the test set. To examine the impact of specific sub-sample divisions on estimation outcomes within finite samples, this study conducted resampling tests with 101 iterations (reps = 101). By estimating across different sub-sample divisions, the robustness of key findings was verified. Results indicate that after resampling tests, primary conclusions remained stable, demonstrating that model estimation is unaffected by particular sample divisions and thereby confirming the robustness of research findings.

5. Additional Analysis

5.1. Conduction Mechanism Verification

Based on the theoretical analysis presented earlier, we hypothesize that data asset information disclosure may mitigate the risk of corporate stock price crashes by enhancing information transparency, investor confidence, and analyst forecast accuracy. To test this hypothesis, this section employs a mediation model to validate the proposed mechanisms. The results of the mediation analysis are presented in Table 5.

Table 5. Mechanism Testing Results.

	Information Transparency			**Investor Confidence**	**Analyst Forecast Accuracy**		
Variable	**(1)** ***ASY1***	**(2)** ***ASY2***	**(3)** ***TRANS***	**(4)** ***IC***	**(5)** ***FERR1***	**(6)** ***FERR2***	**(7)** ***FERR3***
DA	−0.019 ***	−0.025 ***	0.014 ***	0.018 ***	0.162 ***	0.001 *	0.003 **
	(0.003)	(0.004)	(0.002)	(0.004)	(0.054)	(0.001)	(0.001)
Constant	0.000	−0.000	−0.000	0.000	−0.004	−0.000	−0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.013)	(0.000)	(0.000)
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Industry Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Control Variable First Order Terms	YES	YES	YES	YES	YES	YES	YES
Control Variable Second Order Terms	YES	YES	YES	YES	YES	YES	YES
N	30,157	30,157	29,916	26,890	16,455	16,239	16,455

Notes: standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5.1.1. Enhance Information Transparency

Regarding the measurement of information transparency indicators. On the one hand, this paper employs principal component analysis to integrate the liquidity ratio (*LR*), illiquidity ratio (*ILL*), and Return Reversal Measure (*GAM*) to construct an information asymmetry index reflecting information asymmetry. Higher index values indicate less disclosed information and greater information opacity. On the other hand, this paper constructs a comprehensive information transparency index based on five indicators: analyst coverage (*ANALYST*), disclosure score (*DSCORE*), earnings quality (*DD*), analyst earnings forecast accuracy (*ACCURACY*), and Big Four audit (*BIG4*). TRANS is derived by calculating the average of the sample percentile ranks for these variables. If data for certain variables is missing, TRANS is the average of the percentile ranks for the remaining

variables. A higher TRANS value indicates greater corporate information transparency. Regression results are presented in Table 5, Columns (1), (2), and (3). Columns (1) and (2) show the regression coefficients of the information asymmetry index ASY, constructed based on liquidity ratio indicators, for DA, both of which are significantly negative at the 1% level. Column (3) presents the regression coefficient for *DA* on the comprehensive information transparency index *TRANS*, which is significantly negative at the 1% level. These regression results indicate that data asset information disclosure enhances corporate information transparency. As corporate data asset information disclosure increases, market participants gain access to more comprehensive information. This not only enhances the market's ability to identify and respond to risks but also promotes overall market transparency. Consequently, it stabilizes market expectations and reduces market volatility.

5.1.2. Increase Investor Confidence

This paper constructs an investor confidence index by referencing existing research, selecting three indicators: annual price-to-book ratio (*PB*), price-to-earnings ratio (*P/E*), and average annual turnover rate (*TO*). These metrics, respectively, reflect a company's valuation, profitability, and market activity from different perspectives, serving as crucial benchmarks for investors to assess corporate value and market prospects. Through principal component analysis, these indicators are synthesized into an Investor Confidence Index that reflects market sentiment and corporate characteristics. This index measures investor confidence in market and corporate prospects while analyzing its impact on corporate investment behavior. The specific formula is as follows:

$$Investor\ Confidence\ (IC) = 0.8775 \times PE + 0.82 \times PB + 0.6290 \times TO \tag{10}$$

As shown in the formula, all principal component coefficients constituting the IC are positive values, consistent with expectations. The investor confidence index is calculated through standardization. This index quantifies investors' perceptions of market and corporate prospects, enabling analysis of how such confidence influences corporate investment decisions and behaviors. The regression results are presented in Table 5, Column (3). The regression coefficient for *DA* is significantly positive at the 1% level, fully demonstrating the positive impact of data asset information disclosure on investor confidence. When companies disclose their data assets fully and accurately, investors gain a more comprehensive understanding of the company's data capabilities, data governance capacity, and data-driven business models. This transparency reduces information asymmetry between companies and investors, lowering investors' information search costs and risk perception. Simultaneously, transparent and detailed data asset information disclosure enhances investors' trust and sense of belonging toward the company, boosting investor confidence. This stabilizes market expectations and effectively reduces the risk of stock price crashes.

5.1.3. Analyst Prediction Accuracy

This paper employs analyst forecast accuracy as the measurement standard. By calculating the discrepancy between analysts' projected future earnings and the company's actual reported earnings, and then normalizing this discrepancy (by dividing it by actual earnings, initial stock price, or total assets per share), we derive the analyst forecast accuracy metrics FERR1, FERR2, and FERR3. Standardization reflects the precision of analyst forecasts while eliminating the influence of varying company sizes and stock price levels, enabling fairer and more meaningful comparisons. In regression analysis, this study further regresses the analyst forecast accuracy metrics over $t + 1$ periods. This approach is primarily based on the following consideration: analyst forecasts are not merely reactions to current information but, more importantly, projections of future market conditions. By

examining $t + 1$-period forecast accuracy, we can assess whether analysts effectively predict future earnings based on current-period disclosures, thereby validating their forecasting capability. Estimation results are presented in columns (5), (6), and (7) of Table 5. The regression coefficients for *DA* are significantly positive at the 1%, 10%, and 5% significance levels, respectively, indicating that data asset information disclosure enhances the accuracy of analyst forecasts. By improving forecast accuracy, reducing information asymmetry, diminishing stock price synchrony, stabilizing market expectations, and encouraging investor focus on long-term corporate value, data asset information disclosure effectively mitigates the risk of stock price crashes.

In summary, data asset information disclosure mitigates the risk of corporate stock price crashes by enhancing information transparency, investor confidence, and the accuracy of analyst forecasts. Furthermore, to ensure the robustness of the mechanism test, this paper employs a dual machine learning model for resampling tests. The results remain unchanged, validating H2.

5.2. Heterogeneity Test Based on Internal and External Perspectives

Previous research has found that corporate data asset information disclosure effectively reduces the stock price crash risk by enhancing information transmission, solidifying investor confidence, and improving the accuracy of analyst forecasts. This paper further explores the heterogeneous factors influencing the efficacy of these mechanisms from both internal and external perspectives.

5.2.1. Corporate Governance Level

Companies with higher governance standards typically have well established governance structures, such as independent boards of directors, clear separation of CEO and chairman roles, and robust internal control mechanisms. These governance practices ensure the accuracy, reliability, and timeliness of information disclosures. This enhanced transparency helps reduce stock price crash risk by minimizing the potential for hidden bad news and mitigating market overreaction to unexpected events (Thai et al., 2025; Rasheed et al., 2025). As a result, in firms with strong governance mechanisms, the impact of data asset information disclosure on crash risk is relatively limited, as these firms already have robust systems in place to ensure the reliability of their disclosures. In contrast, companies with weaker governance structures—such as those with concentrated ownership, combined CEO-chairman roles, or a lack of independent oversight—often face challenges in ensuring the quality of their data asset disclosures. These firms may suffer from inadequate internal controls or misleading information, making data asset disclosure more impactful in reducing stock price crash risk. By improving the quality and timeliness of disclosures through digital tools, these firms can significantly enhance transparency, reduce information asymmetry, and lower the risk of stock price crashes.

Based on this, this paper constructs a comprehensive indicator for corporate governance levels. It selects executive compensation and the proportion of executive shareholding to represent corporate incentive mechanisms. The proportion of independent directors and board size represent the supervisory role of the board of directors. The proportion of institutional investor holdings and the degree of equity checks and balances represent the supervisory role of the equity structure. Whether the chairman and general manager positions are combined determines the extent of the general manager's decision-making power. Using the above seven indicators, the first principal component obtained through principal component analysis is used as the comprehensive indicator reflecting corporate governance levels. Based on the median corporate governance level, the sample is divided into two groups: low and high governance. The regression results, shown in Panel A

of Table A2 (Appendix B), indicate that *DA* exhibits a statistically significant negative coefficient at the 1% level for both stock price crash risk indicators. Notably, the regression coefficient is larger and more statistically significant for the low governance group compared to the high governance group. The findings indicate that the mitigating effect of data asset information disclosure on stock price crash risk is more pronounced in firms with low corporate governance levels.

5.2.2. Quality of Information Disclosure

The quality of information disclosure in annual reports is crucial for mitigating stock price crash risk. High-quality disclosures narrow the information gap between corporate insiders and external investors, reduce managers' incentives to conceal negative information, and minimize the risk of unexpected "bad news" causing sudden market reactions (Bakhiet, 2024). Enhanced disclosure quality ensures that investors have access to verifiable "hard news" rather than unverifiable "soft news", which is subject to interpretation and can lead to greater uncertainty in the market. Shareholders and external investors generally prefer information that can be validated, as this allows for better post-event accountability and constrains the scope for management's voluntary disclosures.

Based on this, this paper argues that when a company's information disclosure governance is strong, the mitigating effect of data asset information disclosure on stock price crash risk becomes more pronounced. This is because there exists a complementary relationship between the quality of annual report disclosure and data asset information disclosure, particularly when annual report disclosure governance already incorporates requirements for disclosing data asset information. As the primary channel for listed companies' information disclosure, improvements in the quality of annual reports also enhance the transparency and accuracy of data asset information disclosure. Specifically, annual report disclosures encompass not only financial data but also multiple dimensions, including corporate operational status, market environment, and management strategies—all of which inherently include data asset-related information. Consequently, when companies provide high-quality data asset information disclosures in their annual reports, this not only elevates the overall disclosure quality of the annual report but also renders data asset information disclosures more comprehensive and reliable. This further strengthens the relationship between annual report disclosure quality and reduced stock price crash risk.

This study measures the quality level of annual report disclosure using audit opinions, simplifying audit opinions into a binary variable of 0 and 1. An audit opinion of 1 indicates higher disclosure quality in the annual report, signifying that the auditor believes the report's preparation fully complies with applicable accounting standards. An audit opinion of 0 indicates lower disclosure quality in the annual report. The regression results are shown in Panel B of Table A2. *DA* is not significant in the sample with audit opinion 0, but is significantly negative in the sample with audit opinion 1. The findings indicate that the mitigating effect of data asset information disclosure on stock price crash risk is more pronounced in companies with higher annual report disclosure quality, highlighting the critical role of high-quality disclosure in reducing market risk.

5.2.3. Level of Industry Competition

From an agency theory perspective, the intensity of industry competition reshapes the agency relationship between managers and external investors, influencing the impact of data asset information disclosure on stock price crash risk. In highly competitive industries, managers face greater pressure to maintain market share and secure investor support, which heightens information asymmetry (Benkraiem et al., 2022; Zheng et al., 2024). This increased pressure incentivizes managers to distort or withhold data asset

information, risking investor confidence and triggering stock price crashes. To mitigate this, managers are motivated to adopt transparent data asset disclosures, aligning their interests with investors and reducing information asymmetry. Such disclosure limits managerial discretion to hide negative information, as external scrutiny makes opacity costlier, while also enhancing the credibility of managers' claims about the firm's digital capabilities. By providing verifiable information, firms bridge the information gap, reducing hidden bad news accumulation and lowering stock price crash risk. Therefore, expect that in industries with intense competition, the effect of data asset information disclosure on reducing stock price crash risk will be more pronounced, as the competitive environment amplifies the agency conflict, making transparent disclosure a more critical tool for aligning managerial behavior with investor interests.

This paper uses the industry-adjusted Lerner index to quantify the competitive intensity within a firm's industry. Specifically, the competitive intensity within an industry is derived by calculating the difference between an individual firm's Lerner index and the industry-average Lerner index. The formula for a firm's Lerner index is: (Operating Revenue—Operating Costs—Selling Expenses—Administrative Expenses)/Operating Revenue. To derive the industry Lerner index, this study calculates a weighted average of each company's revenue against the total annual revenue of all listed firms within the industry. A higher Lerner index indicates less intense competition within the industry. Based on the median Lerner Index, the sample is divided into two groups: high and low industry competition levels. The regression results, shown in Panel C of Table A2, indicate that *DA* is not significant in the sample with low industry competition but is significantly negative in the sample with high industry competition. The findings suggest that in a highly competitive environment, firms can effectively reduce the stock price crash risk by improving the quality of information disclosure. This protects investor interests, maintains market stability, and consequently lowers the stock price crash risk.

5.2.4. The Development Level of the Digital Economy

The level of development in the digital economy significantly influences the stock price crash risk. In regions with a more advanced digital economy, enterprises can leverage sophisticated digital technologies to manage and utilize data assets more effectively, enhancing the transparency and precision of decision-making. This technological advantage not only optimizes internal corporate controls but also improves adaptability to external environmental changes, reducing market information asymmetry and thereby mitigating the stock price crash risk. As the digital economy deepens, enterprises gain enhanced capabilities in information circulation, data processing, and market forecasting. These improvements directly contribute to share price stability, enabling firms to navigate market volatility with greater resilience. This paper anticipates that in regions with advanced digital economies, the development of the digital economy and the disclosure of data asset information may exhibit a degree of substitutability in mitigating the share price crash risk.

To validate this conclusion, this paper employs an assessment methodology to measure the development level of the digital economy. The evaluation comprehensively examines the state of digital economic advancement through two dimensions: the prevalence of internet access and the inclusiveness of digital finance. For measuring internet development at the city level, the following four key indicators were selected: broadband internet users per hundred people, the proportion of urban employment in the information technology services sector, telecommunications service volume per capita, and mobile telephone users per hundred people. The digital finance development index utilized the China Digital Inclusive Finance Index. Through principal component analysis, the data from these five indicators were standardized and reduced in dimension to derive a comprehensive

regional digital economy development index. Regions were categorized into high and low digital development tiers based on sample medians. Regression results (Panel D of Table A2) indicate that *DA* exhibits statistically significant negative regression coefficients at the 1% level for both stock price crash risk indicators. Notably, the mitigating effect of data asset information disclosure on crash risk is more pronounced in regions with lower digital development levels compared to those with higher development levels. The findings indicate that in regions with lower digital economic development, data asset information disclosure plays a more pronounced role in mitigating stock price crash risk. Conversely, this effect is relatively weaker in regions with higher digital economic development, revealing a substitution effect between digital economic level and data asset information disclosure in alleviating stock price crash risk.

5.3. Extension Study: The Radiation Effects of Data Asset Information Disclosure

Corporate behavior influences other firms within the economic network, and the impact of digital transformation on stock price crash risk may extend beyond individual firms to their peers. This paper investigates whether data asset information disclosure also has a spillover effect within industries, reducing stock price crash risk for other firms. First, from a competition perspective, when one firm discloses data asset information, it raises industry transparency standards, prompting others to follow suit to remain competitive, thereby enhancing sector-wide transparency and efficiency. Second, resource dependence theory suggests that data asset disclosure helps firms build relationships with stakeholders, securing valuable resources. As a result, one firm's disclosure may encourage others to disclose in order to remain competitive in resource acquisition.

To examine the spillover effects of data asset information disclosure, this study constructed a stock price crash risk indicator for other firms within the same industry. Table 6 reports the regression results, revealing that the regression coefficients for *DA* are significantly negative at both the 10% and 1% significance levels. These findings support the spillover effect of data asset information disclosure on stock price crash risk, indicating that a firm's disclosure of data assets not only reduces its own crash risk but also exerts a restraining influence on the crash risk of other firms within the same industry. These findings indicate that transparent management of data assets can generate positive spillover effects within industries, thereby enhancing sector-wide stability and competitiveness.

Table 6. Testing the Radiation Effects of Data Asset Information Disclosure.

Variable	(1) *FDUVOL*	(2) *FNCSKEW*
DA	−0.0011 *	−0.0039 ***
	(0.001)	(0.001)
Constant	0.0001	0.0000
	(0.000)	(0.000)
Control variables quadratic terms	YES	YES
Time fixed effects	YES	YES
Industry fixed effects	YES	YES
Control variables linear terms	YES	YES
N	30,150	30,150

Notes: standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

6. Discussion

In this study, we examine the impact of data asset information disclosure on stock price crash risk in the context of Chinese A-share listed firms. Our findings suggest that greater data asset information disclosure significantly reduces stock price crash risk, which has

both economic and behavioral implications. This section discusses these results, compares them with prior literature, and explores their broader implications, particularly in the Chinese context.

6.1. Economic and Behavioral Interpretation

From an economic perspective, our findings are consistent with signaling theory, which posits that transparency reduces information asymmetry between firms and market participants. By disclosing data assets, firms reduce the uncertainty surrounding their true value, enabling investors to make more informed decisions. This process is especially crucial in reducing stock price crash risk, as the disclosure of previously hidden or ambiguous information can prevent sudden and drastic corrections in stock prices. Our results show that higher data asset information disclosure decreases the likelihood of stock price crashes by providing investors with more accurate assessments of corporate value, ultimately stabilizing market expectations.

Behaviorally, investor sentiment plays a critical role in the relationship between data asset information disclosure and stock price crash risk. Increased transparency helps alleviate investor anxiety and reduces panic-driven selling. According to behavioral finance theories, when investors have access to clear, reliable information, they are less likely to overreact to negative news, which helps mitigate herd behavior and irrational market responses. This aligns with our findings that higher levels of disclosure improve investor confidence and reduce the likelihood of sudden price drops driven by market panic.

The heterogeneous effects further reflect economic and behavioral rationality. On the one hand, distinguishing between transactional data assets (focused on market transactions and value realization) and operational data assets (focused on internal operational efficiency and decision-making optimization) reveals consistent risk-mitigating effects—both types of data asset information disclosure significantly reduce stock price crash risk. Transactional data asset information disclosure clarifies the market value and profit potential of data, strengthening investor recognition of the firm's external value creation capacity; operational data asset information disclosure demonstrates internal management efficiency and data-driven operational capabilities, easing market concerns about operational stability. On the other hand, for firms with weaker corporate governance, data asset information disclosure acts as an "external supervision substitute": where internal controls fail to constrain managerial opportunism, transparent data disclosure exposes inefficiencies to market scrutiny, limiting managers' ability to hide poor performance. In highly competitive industries, data disclosure becomes a strategic tool to gain investor trust—peer transparency pressures force firms to improve their own disclosure quality, raising industry-wide standards. In regions with underdeveloped digital economies, data asset information disclosure compensates for weak digital infrastructure, while in digitally advanced regions, infrastructure already reduces information asymmetry, leading to a substitution effect that diminishes the marginal impact of disclosure.

6.2. Comparison with Prior Research

Prior studies have consistently validated the core framework that high-quality disclosure mitigates stock price crash risk by alleviating information asymmetry. This logic has been corroborated across diverse disclosure contexts: Kim and Zhang (2014) demonstrated that financial reporting opacity enables the accumulation of hidden bad news, as managers exploit information asymmetry to withhold adverse information—ultimately triggering sudden stock price collapses when such news is revealed. Extending this to non-financial disclosure, Bae et al. (2021) found that ESG disclosure reduces crash risk by addressing financial constraints and enhancing transparency, which stabilizes investor expectations and

curbs irrational overreactions to negative events. At a theoretical level, Bleck and Liu (2007) further underscored that market opacity (exacerbated by inadequate disclosure) creates a "veil" for managers to mask true firm performance, leading to more frequent and severe asset price crashes—reinforcing the critical role of transparency in risk mitigation. Even in emerging disclosure domains like sustainability, Lin and Wu (2023) confirmed that climate risk disclosure mitigates crash risk by reducing information asymmetry, highlighting the generalizability of the "disclosure-transparency-risk reduction" framework.

While our study aligns with the established "disclosure-transparency-risk reduction" framework, it makes two key contributions to the literature. First, we broaden the concept of critical disclosures by introducing data asset information disclosure, which is particularly relevant in today's digital economy but remains underexplored. Unlike traditional financial disclosures, ESG reports, or climate risk disclosures—each focusing on past performance, non-financial obligations, or specific sustainability risks—data asset disclosure provides vital insights into a company's digital infrastructure, data governance practices, and future growth potential. This disclosure type, which highlights intangible data assets, is increasingly critical for evaluating firms in a data-driven economy but has been largely absent from traditional transparency models. Second, we extend the mechanistic understanding of how disclosure reduces crash risk. While previous studies primarily focus on alleviating information asymmetry to curb hidden bad news or irrational overreactions, our findings reveal three complementary mechanisms: data asset disclosure not only improves information transparency but also boosts investor confidence and enhances analyst forecast accuracy. These mechanisms collectively reinforce each other, addressing both informational and behavioral drivers of crash risk.

In sum, our study validates the generalizability of the disclosure framework to emerging data assets while expanding it by uncovering the multifaceted role of data asset disclosure in mitigating crash risk. This provides new evidence that the core logic of disclosure-driven risk reduction remains robust, even as business models shift toward data-centric approaches.

6.3. Data Asset Disclosure in China: Institutional Drivers and Broader Implications

The Chinese setting is critical to interpreting the findings, and the results carry conditional generalizability to other markets.

China's institutional environment shapes the data asset disclosure's crash risk-mitigating effect in unique, non-universal ways. Unlike mature economies dominated by institutional investors (who access private information via research), China's market is retail-driven: retail investors make up a large share of trading and lack tools to process complex data, making concrete data asset disclosure (e.g., details on data-driven efficiency) especially critical for stabilizing expectations. China also frames data as a strategic national resource via the "Digital China" policy—so disclosing data assets aligns with national goals, boosting credibility—a dynamic absent in markets where data is not tied to national strategy. Additionally, China's data asset regulations remain formative: while data was formally recognized as an asset recently, no uniform disclosure standards exist, creating firm-level variation that lets us identify causal effects. This regulatory uncertainty is temporary and China-specific, though—mature markets with established data rules lack this variation, so our findings on "disclosure intensity" cannot be directly transplanted.

These traits explain why disclosure's risk-mitigating effect is more visible in China, such as stronger impacts in regions with less developed digital economies (tied to China's uneven infrastructure, a gap not seen in markets with unified digital systems). Importantly, while disclosure's core mechanisms (reducing information asymmetry, anchoring expectations) are universal for digital-era markets, their strength and execution are bounded by

China's context—we do not claim they work identically elsewhere, only that they offer an "analysis template" for markets with similar traits.

Thus, our single-country dataset limits cross-jurisdictional implications to conditional references: other markets can use our analytical framework (e.g., studying how investor structure shapes disclosure effects) to craft policies, but they cannot directly adopt China's practices without adapting to their own contexts (e.g., investor makeup, regulatory maturity). This prevents overgeneralization—our results inform, but do not dictate, global data asset disclosure design.

6.4. Theoretical Contribution

This study makes the following three theoretical contributions:

First, it extends financial risk theory by moving beyond traditional disclosure paradigms (e.g., financial reports, ESG disclosures) to account for digital assets. Existing financial risk theories, while encompassing various information types, overlook how intangible digital resources with unique economic attributes (low marginal cost, scale economies) reshape risk dynamics. By embedding data asset information disclosure into stock price crash risk frameworks, this study shows that data transparency—via its role in stabilizing operations through data-driven efficiency and governance—acts as a new digital-era risk–mitigation pathway. Highlighting data assets' distinct value in modern market stability, it expands financial risk theory to incorporate the idiosyncrasies of digital factors.

Second, it integrates multi-theory logic for digital signals. Prior research treats information economics and behavioral finance as isolated mechanisms, failing to explain their interaction in non-standard digital contexts. To address this, the study builds a synergistic framework. Rooted in the Resource-Based View (RBV), it ensures that data assets' substantive value provides disclosure credibility. Signaling theory then explains how this value becomes a costly, verifiable signal to reduce information asymmetry. Meanwhile, behavioral finance shows how the signal anchors investor expectations, curbing biases like overreaction. This integrated logic unifies "signal credibility" and "behavioral responses," offering a holistic view of digital disclosure's market-stabilizing effects.

Third, it refines heterogeneous disclosure with digital contexts. Existing theories neglect how external and internal contexts shape digital signals' efficacy. The study refines this by identifying three boundary conditions. In terms of corporate governance, strong governance enhances data disclosure and the risk-mitigating effect. Regarding industry competition, high competition amplifies the signal value of data asset disclosure, as rivals' benchmarking validates authenticity. When it comes to digital economy development, disclosure complements digital infrastructure in developed regions and substitutes for it in underdeveloped ones. These findings add nuance to heterogeneous disclosure theory, emphasizing the "context-signal-effect" interplay.

7. Conclusions

7.1. Conclusions and Implications

This study constructs a text-based index of data asset information disclosure using a sample of Chinese A-share listed firms from 2010 to 2023, and employs a double machine learning (DML) framework to examine its causal effect on corporate stock price crash risk.

The empirical findings demonstrate that:

1. Data asset information disclosure significantly mitigates stock price crash risk. When distinguishing between self-use and transactional data asset information disclosures, both types exhibit a pronounced negative impact on crash risk.

2. Mechanism analyses reveal that data asset information disclosure reduces crash risk through three channels: the information-transmission mechanism, the investor-confidence mechanism, and the analyst-forecast-accuracy mechanism.
3. Heterogeneity analyses show that the mitigating effect of data asset information disclosure is stronger among firms with weaker corporate governance and higher annual report transparency, as well as in industries with more intense market competition and regions with lower digital economy development levels.
4. Furthermore, the results reveal a significant industry spillover effect—firms' data asset information disclosures not only reduce their own crash risk but also lower the crash risk of peer firms within the same industry.

Collectively, these findings extend the literature on data asset information disclosure and market risk, providing empirical evidence for how data transparency contributes to capital market stability. The results also offer practical implications for firms, regulators, and investors.

- Implications for Enterprises

First, strengthen data asset governance: enhance the internal management, valuation, and control of data assets to ensure the accuracy and completeness of disclosures—especially for firms with weak corporate governance, where standardized data governance can make disclosure a reliable substitute for internal supervision. Second, improve disclosure transparency by proactively releasing detailed information on data collection, processing, utilization, and value creation; for firms in highly competitive industries, such transparency can serve as a strategic tool to gain investor trust, while those in digitally underdeveloped regions can use disclosure to compensate for insufficient digital infrastructure. Third, develop data talent: invest in cultivating and recruiting data management and analytics professionals to strengthen data-driven decision-making, which not only supports high-quality disclosure but also enhances corporate resilience and market valuation by unlocking the value of data assets.

- Implications for Regulators and Policymakers

Drawing on the study's findings that data asset information disclosure significantly mitigates stock price crash risk by enhancing information transparency, boosting investor confidence, and improving analyst forecast accuracy—with more pronounced effects for firms with weak corporate governance and in regions with underdeveloped digital infrastructure—regulators and policymakers should prioritize developing executable, scenario-specific data asset disclosure standards: first, issue industry-tailored guidelines to clarify core dimensions for data classification and boundary conditions for valuation methods, while establishing a tiered qualification system for third-party auditors to ensure these entities possess the necessary expertise to audit both the rationality of data asset valuation and the legitimacy of data ownership, with audit results formally linked to key corporate activities that require regulatory oversight. To address potential implementation barriers and encourage firm compliance, design a quantifiable graded incentive mechanism that ties disclosure quality to tangible policy preferences, while complementing this with strict accountability measures to address misconduct such as disclosure fraud, ensuring the integrity of the incentive system. For regions where digital infrastructure remains underdeveloped, launch targeted pilot programs in sectors with high digital reliance, set clear indicators to evaluate pilot effectiveness, and establish a dynamic adjustment mechanism to revise implementation details based on emerging challenges (such as disputes over valuation standards or risks to data security) —all to accumulate actionable experience that can guide the gradual expansion of disclosure requirements to a broader range of firms and regions.

- Implications for Capital Market Participants

For capital market participants, investors should first enhance their understanding of data assets by integrating disclosure information into investment analysis and risk assessment frameworks: prioritize firms that clearly disclose the value contribution of operational data assets (e.g., data-driven operational efficiency) and the market potential of transactional data assets (e.g., data monetization channels), as these details reduce uncertainty and crash risk. Second, financial analysts should enhance their data analytics capabilities: master methods to evaluate data asset value (e.g., linking data word frequency to revenue growth) and improve the accuracy of earnings forecasts, which in turn stabilizes market expectations. Finally, promote information sharing and interdisciplinary collaboration among investors, analysts, and data experts—for example, through industry forums or digital platforms—to foster data-driven investment strategies and improve overall market efficiency, especially in retail-dominated markets where individual investors often lack professional data analysis capabilities.

- Implications for Auditors

Investors and analysts should systematically incorporate firms' Data Asset Information Disclosure Index into their valuation and risk assessment models. They should pay particular attention to the disclosure practices of firms in regions with lower digital infrastructure and those in highly competitive industries, as these firms exhibit a greater marginal benefit from transparency, presenting potential alpha opportunities. The documented spillover effect also suggests that analysts should monitor the data disclosure trends of industry leaders, as this can serve as a leading indicator for the stability of the entire sector.

- Development of Disclosure Standards and Lessons for Other Jurisdictions

Data asset information disclosure standards should evolve to reflect the growing importance of data as a financial asset. Policymakers must move beyond traditional financial reporting frameworks and create specific guidelines for the recognition, valuation, and reporting of data assets. Drawing from China's 2023 regulatory reform, other jurisdictions can develop similar frameworks for data asset information disclosure. However, it is important to recognize that while China's institutional context provides a valuable reference, these findings are not universally applicable without further empirical testing in different regions with varied institutional and regulatory environments.

Other jurisdictions, particularly those with more developed data regulations, such as the U.S. and Europe, can take lessons from China's approach, particularly in recognizing data as an asset and integrating it into corporate financial reporting. Nonetheless, the direct application of China's regulatory framework to these markets requires caution, as differing market structures, investor behavior, and digital infrastructure may lead to varying results. Future research should explore whether China's regulatory framework could be adapted or needs to be adjusted to suit other jurisdictions with more established regulatory regimes.

Policymakers should be mindful of the potential regulatory gaps that could arise as new technologies and business models emerge, ensuring that data asset information disclosure evolves in tandem with these innovations. Therefore, while China's approach provides a valuable starting point, the development of global disclosure standards must be grounded in broader cross-country research to ensure these standards are both applicable and effective across different regulatory environments.

7.2. Limitation

This study has several limitations that provide directions for future research. First, regarding the measurement of data asset information disclosure (DA), the keyword set

expansion relies on the Word2Vec model. While Word2Vec effectively captures semantic similarities through cosine similarity, it may lead to semantic distortion by conflating terms with superficial overlaps but distinct contextual relevance to data assets (e.g., confusing "digital marketing" as a business strategy with "digital infrastructure" as a core data asset). Additionally, the model's performance is highly dependent on the training corpus—even though we integrated multi-level policy texts to enhance validity, the corpus may lack industry-specific terminology (e.g., in fintech or manufacturing), leading to an incomplete keyword set and posing model generalization risks across different sectors.

Second, regarding methodological comparison, while this study elaborates on DML's theoretical advantages over OLS (e.g., handling nonlinear relationships, capturing high-dimensional confounding), we failed to conduct empirical tests on model performance (e.g., out-of-sample predictive accuracy) due to data and computational constraints. This limits our ability to empirically validate whether DML's theoretical merits translate into tangible measurement improvements in the specific context of this research.

Third, regarding mechanism validation, constrained by the focus of the DML framework on core causal effect estimation, this study did not adopt traditional causal mediation models for structural verification.

Finally, the study's market scope is singular, as it only explores Chinese A-share listed firms. This lacks cross-market validation (e.g., mature markets like the U.S. or other emerging markets), limiting the generalizability of findings to global contexts.

Building on the limitations identified, several avenues for future research emerge. First, there is a need to develop more precise measurement methods for information disclosure by integrating expert knowledge with large language models. Additionally, systematic comparisons between machine learning and traditional econometric methods in disclosure research warrant further exploration. Finally, extending the machine learning framework to formal mediation mechanism testing represents a promising path forward. Furthermore, future studies could expand the analysis to include developed economies, comparing the impact of data asset disclosure across different regulatory environments to assess how market maturity and institutional factors shape the effectiveness of disclosure practices in enhancing market stability.

Author Contributions: Conceptualization, J.Z. and Y.G.; methodology, Z.Z. and Y.G.; software, H.W.; validation, J.Z., Z.Z. and Y.G.; formal analysis, J.Z.; investigation, J.Z. and H.W.; resources, Z.Z.; data curation, H.W. and Y.G.; writing—original draft preparation, J.Z., Z.Z., H.W., Y.G., Y.Z. and F.L.; writing—review and editing, J.Z., Z.Z., H.W., Y.G., Y.Z. and F.L.; visualization, H.W.; supervision, Y.Z. and F.L.; project administration, Y.Z.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Guangdong Provincial Department of Education (2024GX1K650).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Construction Process of Data Asset Information Disclosure

To ensure the scientific rigor and comprehensiveness of the corporate data assetization measurement index, this study adopts a text analysis-based approach to construct the variable. Detailed as follows:

Step 1: Determination of Seed Words Based on Conceptual Definition

Against the backdrop of the rapid development of digital technologies such as the Internet of Things (IoT), cloud computing, big data, and artificial intelligence (AI), the term "data" has evolved to be synonymous with "digitized data"—information stored, transmitted, and processed in digital form. The concept of data assets has gradually evolved from "information resources" and "information assets", and its existence relies on carriers such as data middle platforms, network information systems, and financial sharing platforms.

Combining the definition of data assets from the Research Report on Data Asset Recognition and Accounting Measurement by the Policy and Economic Research Institute of China Academy of Information and Communications Technology, which defines data assets as "digitally expressed information resources formed in the process of production, operation and transaction activities of market entities, with clear ownership or control, measurable economic value, and able to bring direct or indirect economic benefits to the holder through reasonable utilization or transaction" (Hu et al., 2022).

Based on this authoritative conceptual framework, four core terms closely corresponding to the essential attributes of data assets—"data", "digital", "information", and "network"—are selected as initial seed words, as they directly reflect the digital form, informational nature, ownership boundary, and technical carrier characteristics emphasized in the report.

Step 2: Construction of a Regulatory Text Corpus

To enhance the objectivity and comprehensiveness of the word set expansion, this study collected legal and regulatory texts related to data assets from both national and local levels in China. These texts cover key domains including data security, fundamental data systems, data ownership, data transactions, and accounting treatment of corporate data resources (He et al., 2024). The legal and regulatory texts included in the corpus are specifically as follows: Cybersecurity Law of the People's Republic of China (2016), Cybersecurity Law of the People's Republic of China (2021), Personal Information Protection Law of the People's Republic of China (2021), 14th Five-Year Plan for National Informatization (2021), Several Views on Fundamental Data Systems (2022), Opinions on Establishing Fundamental Data Systems to Better Leverage the Role of Data Factors (2022), Interim Provisions on Accounting Treatment of Corporate Data Resources (Exposure Draft) (2022), Beijing Digital Economy Promotion Regulations (2022), Shanghai Data Regulations (2021), Tianjin Regulations on Promoting the Development and Application of Big Data (2018), Chongqing Data Regulations (2022), Heilongjiang Regulations on Promoting the Development and Application of Big Data (2022), Jilin Regulations on Promoting the Development and Application of Big Data (2020), Liaoning Big Data Development Regulations (2022), Hebei Digital Economy Promotion Regulations (2022), Shaanxi Big Data Regulations (2022), Shandong Regulations on Promoting the Development of Big Data (2021), Shanxi Regulations on Promoting the Development and Application of Big Data (2020), Anhui Big Data Development Regulations (2021), Jiangsu Digital Economy Promotion Regulations (2022), Sichuan Data Regulations (2022), Guizhou Regulations on Promoting the Development and Application of Big Data (2016), Guizhou Big Data Security Assurance Regulations (2019), Guizhou Regulations on Government Data Sharing and Opening (2020), Guangxi Zhuang Autonomous Region Big Data Development Regulations (2022), Zhejiang Public Data Regulations (2022), Jiangxi Data Application Regulations (Draft) (2022), Guangdong Digital Economy Promotion Regulations (2021), Shenzhen Special Economic Zone Data Regulations (2021), Shenzhen Special Economic Zone Digital Economy Industry Promotion Regulations (2022), Fujian Big Data Development Regulations (2021), and Hainan Big

Data Development and Application Regulations (2019). All these texts were imported into Python (version 3.13) and merged into a unified regulatory corpus.

This approach addresses the limitation of prior studies (Hu et al., 2022), which only included two national-level laws on cybersecurity and data security, resulting in a narrow and incomplete corpus. By incorporating multi-level and multi-domain regulatory texts, this study improves the validity of the corpus and ensures that the expanded word set fully reflects the institutional context of data assetization for A-share listed firms.

Step 3: Expansion of Similar Words Using Word2Vec Neural Network Model

To expand the seed words into a more comprehensive and contextually relevant word set, this study employed the Skip-Gram architecture within the Word2Vec neural network model—a method widely used for capturing semantic relationships in text analysis of A-share corporate reports (Hu et al., 2022). The model was trained on the regulatory corpus to predict context words based on target words, converting segmented words in the corpus into high-dimensional word vectors. For instance, seed words like "data governance" and "data privacy" were expanded to terms such as "data protection", "information security", and "data compliance", which are closely related in the context of data asset management. Cosine similarity was then calculated between the word vectors of the seed words and other words in the corpus. Words with high similarity to the seed words were retained to form an expanded set of similar words, ensuring that the expanded set aligns with the semantic context of data assetization, rather than relying on subjective manual selection.

Step 4: Screening and Validation of Similar Words

To ensure the accuracy and relevance of the expanded word set, two key screening steps were implemented:

1. Exclusion of negative terms: Words containing negative modifiers (e.g., "non-", "without", "no", such as "non-digital" or "data-free") were manually excluded, as they contradict the positive connotation of "data assetization."

2. Manual validation: The remaining similar words were cross-validated by querying authoritative platforms (e.g., Baidu, Sina) to verify their actual semantic relevance to data assets. This step eliminates ambiguous or irrelevant terms, ensuring the final word set accurately reflects the core concept of data assetization for A-share firms.

Step 5: Word Frequency Statistics and Index Construction for A-Share Listed Firms

Annual reports of Chinese A-share listed firms (sample period: 2010–2023) were downloaded from the official websites of the Shenzhen Stock Exchange, Shanghai Stock Exchange, and National Equities Exchange and Quotations. Following the standard practice of A-share empirical studies, we excluded financial and insurance firms, ST/*ST firms, and firms with missing report data to ensure sample validity.

The valid annual reports were converted into TXT format, renamed for traceability, and processed using the Jieba word segmentation tool to extract word frequencies of the expanded similar word set. The total word frequency was used as the initial measure of corporate data assetization (denoted as DA). Given the typical right-skewed distribution of data asset-related word frequencies among A-share firms—where a small number of technology-intensive enterprises exhibit extremely high frequencies—we first winsorized the initial *DA* values at the 1st and 99th percentiles to mitigate the impact of extreme values. Subsequently, the winsorized total word frequency was divided by 100 for normalization. This normalization step compresses the indicator into a manageable range while retaining the relative differences in disclosure intensity across firms. The final normalized value is used as the core *DA* index, with higher values indicating more comprehensive and in-depth data asset information disclosure.

Step 6: Classification of Data Assets by Usage Scenarios

With reference to the methodology of the China Academy of Information and Communications Technology (2020), this study classifies data asset keywords into two sub-types based on their usage within enterprises: operational data assets and transactional data assets. This classification follows the "holding purpose" principle outlined in the Interim Provisions on Accounting Treatment of Corporate Data Resources. Separate word frequency statistics were calculated for each subtype. The specific classification of keywords is shown in Table A1:

Table A1. Classification of Operational and Transactional Data Asset Keywords.

Data Asset Type	Keyword Category	Specific Keywords
Operational Data Assets (ODA) (Low marketization, for internal operations)	Digital-related	Digital infrastructure, digital facilities, digital factory, digital equipment, digital economy, digital technology innovation, digital governance, digital management, digital operation and maintenance, digital design, digital technical means, digital technology, digital resources, digital elements, digital backbone enterprises, digital application, digital ecology, digital capability, digital literacy, digital industry, digital space
	Data-related	Data infrastructure, data laboratory, data collection, data aggregation, data development, data development and utilization, data governance, data storage, data-driven, data processing, data calculation, data utilization, data innovation, data standardization, data resource management, data strategy, data resource integration, data monitoring, data economy, data research, data protection, data interaction, data optimization, data talent training, data encryption, data desensitization, data cleaning, data networking, data operation, data supply, data integration, data connection, data migration to cloud, data management, data acquisition, data analysis, centralized data management, data assets, data technology, data talents, key data technologies, data experts, data algorithms, data information technology, data business, data network, data talent team, data computing power, data engineering, data visualization, data resource advantages, data review, data query, data research institute, data industrial park, data verification, data industry layout, data scale, data application, data simulation, data interface, data security level, data collection, data technology industry, data confidentiality, data industry chain, data project, data artificial intelligence, data empowerment, data production, data usage, data development
	Information-related	Information infrastructure, information facilities, information terminals, information equipment, information disclosure, information acquisition, information collection, information management, information connection, information R&D, information acquisition, information and communication, information resources, information technology, information data, information network, information carrier, information operation, information manufacturing, information protection, information integration, information identification, information monitoring, information recording
	Network-related	Network infrastructure, network information system, network data center, network facilities, network construction, network storage, network governance, network access, network data, network technology, network information technology, network information, network digitalization, network business, network industry chain, network performance, network establishment, network optimization, network function, network application, network expansion, network layout, network operation
Transactional Data Assets (DDA) (High marketization, for external circulation)	Digital-related	Digital platform, digital trade, digital authentication, digital consumption, digital products, digital currency, digital product security
	Data-related	Data platform, data transaction platform, data opening platform, data service platform, data transaction, data opening, data circulation, data resource sharing, data sharing, data cooperation, data collaboration, data delivery, data application services, data hosting, data consumption, data property rights, data usage rights, data security, data information security, data evaluation, data service providers, data disclosure, data supervision, data interconnection, data exchange, data market, data provision

Table A1. *Cont.*

Data Asset Type	Keyword Category	Specific Keywords
	Information-related	Information platform, information sharing, information services, information consumption, information resource sharing, information interconnection, information service system, information exchange, information service level, information service platform
	Network-related	Network transactions, network sales, network data security, network information security, network security, network risks, network provision, network services, network interconnection, network integration, network service providers, cyberspace sovereignty

Notes: The construction method of this dictionary is consistent with the measurement method of the data asset information disclosure index: both rely on text analysis of A-share annual reports and regulatory corpus, and strictly distinguish the semantic connotations of keywords for different asset types. Specifically, ODA keywords (e.g., "data-driven decision-making", "digital operation and maintenance") are defined to reflect internal operational support, while DDA keywords (e.g., "data transaction platform", "digital trade") are defined to reflect external market circulation—aligning with the following asset type definitions: Operational Data Assets (ODA): These have low marketization and primarily serve internal business processes. ODA refers to data resources collected, organized, and analyzed during corporate organization, production, and operation to support business operations and management decisions—ultimately enhancing profitability. According to accounting standards, ODA is generally recognized as an intangible asset. Transactional Data Assets (DDA): These have high marketization and involve external circulation. DDA refers to data resources collected, analyzed, and managed internally by enterprises, which are further developed into tradable data services or products for external exchange. Such assets are typically recognized as inventory in accounting practice due to their purpose of being sold.

This classification facilitates subsequent analysis of the heterogeneous mechanisms of data assetization among A-share firms.

While our initial analysis relied on an automated text mining approach, we recognized the potential biases inherent in this method. To address these concerns, we implemented additional verification steps. Specifically, we manually validated a representative subset of the data and compared the automated results with expert judgments. This process involved expert coders reviewing and categorizing the same data, ensuring that the classifications were consistent with the automated system's output. We also conducted a benchmark comparison with established methods in the field to verify that our results aligned with those obtained through traditional manual approaches. The results of the expert coding were found to be highly consistent with the automated analysis, further validating the reliability of our indicators.

Appendix B. Heterogeneity Analysis

Table A2. Heterogeneity analysis.

	(1)	(2)	(3)	(4)
	FDUVOL	***FDUVOL***	***FNCSKEW***	***FNCSKEW***
Panel A: Heterogeneity Test of Corporate Governance Levels				
	High	Low	High	Low
DA	−0.0318 ***	−0.0500 ***	−0.0472 ***	−0.0632 ***
	(0.011)	(0.015)	(0.016)	(0.023)
Constant	−0.0006	0.0000	−0.0007	−0.0003
	(0.004)	(0.004)	(0.006)	(0.006)
Control Variable First-Order Term	YES	YES	YES	YES
Control Variable Second-Order Term	YES	YES	YES	YES
Time Fixed Effect	YES	YES	YES	YES
Industry Fixed Effect	YES	YES	YES	YES
N	14286	15871	14286	15871
Panel B: Heterogeneity Test for Information Disclosure Quality				
	High	Low	High	Low
DA	−0.0401 ***	−0.0106	−0.0544 ***	−0.0633
	(0.009)	(0.039)	(0.014)	(0.063)
Constant	0.0000	−0.0059	0.0001	0.0015
	(0.003)	(0.015)	(0.004)	(0.025)

Table A2. *Cont.*

	(1)	(2)	(3)	(4)
	FDUVOL	***FDUVOL***	***FNCSKEW***	***FNCSKEW***
Control Variable First-Order Term	YES	YES	YES	YES
Control Variable Second-Order Term	YES	YES	YES	YES
Time Fixed Effect	YES	YES	YES	YES
Industry Fixed Effect	YES	YES	YES	YES
N	29,004	1,153	29,004	1,153
Panel C: Heterogeneity Test of Industry Competition Intensity				
	High	Low	High	Low
DA	−0.0605 ***	−0.0180	−0.0894 ***	−0.0217
	(0.014)	(0.011)	(0.022)	(0.016)
Constant	0.0001	0.0001	−0.0002	0.0009
	(0.004)	(0.004)	(0.006)	(0.006)
Control Variable First-Order Term	YES	YES	YES	YES
Control Variable Second-Order Term	YES	YES	YES	YES
Time Fixed Effect	YES	YES	YES	YES
Industry Fixed Effect	YES	YES	YES	YES
N	15,079	15,078	15,079	15,078
Panel D: Heterogeneity Test of Digital Economy Development Levels				
	High	Low	High	Low
DA	−0.0275 ***	−0.0654 ***	−0.0451 ***	−0.0750 ***
	(0.011)	(0.014)	(0.016)	(0.022)
Constant	−0.0001	0.0003	0.0002	0.0005
	(0.004)	(0.004)	(0.006)	(0.006)
Control Variable First-Order Term	YES	YES	YES	YES
Control Variable Second-Order Term	YES	YES	YES	YES
Time Fixed Effect	YES	YES	YES	YES
Industry Fixed Effect	YES	YES	YES	YES
N	15,153	15,004	15,153	15,004

Notes: standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

References

Bae, J., Yang, X., & Kim, M. I. (2021). ESG and stock price crash risk: Role of financial constraints. *Asia-Pacific Journal of Financial Studies*, *50*(5), 556–581. [CrossRef]

Bakhiet, S. A. (2024). Does audit quality moderate the relationship between financial statements readability and stock price crash risk? *Journal of Financial Reporting and Accounting*. [CrossRef]

Benkraiem, R., Galariotis, E., Guizani, A., & Lakhal, F. (2022). Product market competition and stock price crash risk: Exploring the role of managerial ownership. *Research in International Business and Finance*, *63*, 101786. [CrossRef]

Bleck, A., & Liu, X. (2007). Market transparency and the accounting regime. *Journal of Accounting Research*, *45*(2), 229–256. [CrossRef]

Bodory, H., Huber, M., & Lafférs, L. (2022). Evaluating (weighted) dynamic treatment effects by double machine learning. *The Econometrics Journal*, *25*(3), 628–648. [CrossRef]

China Academy of Information and Communications Technology, Policy and Economic Research Institute. (2020). *Data Assetization: Research Report on Data Asset Recognition and Accounting Measurement*. CAICT.

Collins, V., & Lanz, J. (2019). Managing data as an asset. *The CPA Journal*, *89*(6), 22–27.

Epstein, L. G., & Schneider, M. (2008). Ambiguity, information quality, and asset pricing. *The Journal of Finance*, *63*(1), 197–228. [CrossRef]

Friske, W., Hoelscher, S. A., & Nikolov, A. N. (2023). The impact of voluntary sustainability reporting on firm value: Insights from signaling theory. *Journal of the Academy of Marketing Science*, *51*(2), 372–392. [CrossRef]

Fu, J., Wu, X., Liu, Y., & Chen, R. (2021). Firm-specific investor sentiment and stock price crash risk. *Finance Research Letters*, *38*, 101442. [CrossRef]

Ghasemaghaei, M., & Calic, G. (2020). Assessing the impact of big data on firm innovation performance: Big data is not always better data. *Journal of Business Research*, *108*, 147–162. [CrossRef]

Gong, X. L., & Liu, J. (2023). Institutional investor information network, analyst forecasting and stock price crash risk. *Research in International Business and Finance*, *65*, 101942. [CrossRef]

Gong, Y., Sun, R., Cai, Y., & Li, W. (2025). Executive traits and stock price crash risk: Evidence from China's Growth Enterprise Market. *Corporate Ownership & Control*, *22*(3), 47–61.

He, Y., Chen, L., & Du, Y. (2024). Does Data Assetization Alleviate Financing Constraints of SRDI SMEs. *China Industrial Economy*, *8*, 154.

Hu, C., Li, Y., & Zheng, X. (2022). Data assets, information uses, and operational efficiency. *Applied Economics*, *54*(60), 6887–6900. [CrossRef]

Jackson, A. B., & Luu, S. (2023). Accounting for digital assets. *Australian Accounting Review*, *33*(3), 302–312. [CrossRef]

Kim, J. B., & Zhang, L. (2014). Financial reporting opacity and expected crash risk: Evidence from implied volatility smirks. *Contemporary Accounting Research*, *31*(3), 851–875. [CrossRef]

Kim, J. B., & Zhang, L. (2016). Accounting conservatism and stock price crash risk: Firm-level evidence. *Contemporary Accounting Research*, *33*(1), 412–441. [CrossRef]

Lewbel, A. (1997). Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. *Econometrica: Journal of the Econometric Society*, *65*, 1201–1213.

Li, Y., Luo, C., Dong, L., & Gui, M. (2022). Data asset disclosure and nonprofessional investor judgment: Evidence from questionnaire experiments. *Mobile Information Systems*, *2022*(1), 8116063. [CrossRef]

Lin, B., & Wu, N. (2023). Climate risk disclosure and stock price crash risk: The case of China. *International Review of Economics & Finance*, *83*, 21–34.

Ling, S., Jin, S., Wang, H., Zhang, Z., & Feng, Y. (2024). Transportation infrastructure upgrading and green development efficiency: Empirical analysis with double machine learning method. *Journal of Environmental Management*, *358*, 120922. [CrossRef]

Luo, S. (2025). The impact of data asset information disclosure on bank credit. *Finance Research Letters*, *85*, 108291. [CrossRef]

Murata, R., & Hamori, S. (2021). ESG disclosures and stock price crash risk. *Journal of Risk and Financial Management*, *14*(2), 70. [CrossRef]

Qian, K., Pan, C., & Liang, X. (2025). Does data asset disclosure affect bank lending acquisition? *International Review of Financial Analysis*, *97*, 103834. [CrossRef]

Rasheed, M. S., Kouser, S., & Ling, Z. (2025). Corporate governance and stock price crash risk: Insights from an emerging market. *Asia-Pacific Financial Markets*, *32*(2), 691–709. [CrossRef]

Shi, H., Xia, Y., Li, Y., Hua, Y., & Fu, Y. (2025). Data Asset Information Disclosure Divergence and Green Innovation: Causal Evidence from Double-Debiased Machine Learning. *Finance Research Letters*, *84*, 107757. [CrossRef]

Thai, H. M., Dang, K. N., Nor, N. M., Nguyen, H. T., & Nguyen, K. V. (2025). Corporate tax avoidance and stock price crash risk: The moderating effects of corporate governance. *International Journal of Emerging Markets*, *20*(1), 1–25. [CrossRef]

Wang, L., Cao, Z., & Cao, W. (2024). Do controlling shareholders' equity pledges exacerbate the stock price crash risk?—A study based on margin trading and securities lending transactions in China's securities market. *Heliyon*, *10*(18), e37473. [CrossRef]

Xie, L., Liu, G., & Liu, B. (2023). Patent pledge policy and stock price crash risk: Evidence from China. *Research in International Business and Finance*, *65*, 101967. [CrossRef]

Xu, F., Wang, K., Wang, S., & Jing, K. (2025). Data is power: The impact of data asset management on expected stock returns. *Asia-Pacific Journal of Accounting & Economics*. [CrossRef]

Xu, T., Shi, H., Shi, Y., & You, J. (2023). From data to data asset: Conceptual evolution and strategic imperatives in the digital economy era. *Asia Pacific Journal of Innovation and Entrepreneurship*, *18*(1), 2–20. [CrossRef]

Yao, X., Wu, D., Li, Z., & Xu, H. (2024). On the prediction of stock price crash risk using textual sentiment of management statement. *China Finance Review International*, *14*(2), 310–331. [CrossRef]

Yin, H., Yin, X., & Wen, F. (2025). Artificial intelligence and climate risk: A double machine learning approach. *International Review of Financial Analysis*, *103*, 104169. [CrossRef]

Zhang, M., Xie, L., & Xu, H. (2016). Corporate philanthropy and stock price crash risk: Evidence from China. *Journal of Business Ethics*, *139*(3), 595–617. [CrossRef]

Zheng, X., Zhang, X., Yang, G., & Wang, L. (2024). Product market competition, corporate governance and stock price crash risk: Evidence from China. *Applied Economics Letters*, *31*(12), 1150–1158. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.