Academic Journal of Management and Social Sciences
ISSN: 2958-4396 | Vol. 14, No. 2, 2026

Data Asset Information Disclosure and Stock Mispricing

Ziwen Huang *

Nanjing Normal University, Nanjing Jiangsu, 210000, China
*** Corresponding author:** Ziwen Huang

Abstract: This study focuses on Chinese A-share listed companies from 2011 to 2024, exploring the impact of data asset information disclosure on stock mispricing. Through the analysis of annual report texts, it was found that information disclosure significantly reduced mispricing and improved the efficiency of capital market pricing. This effect operates through four channels: promoting digital transformation, attracting analysts' attention, enhancing environmental, social and governance ratings, and curbing earnings management. This effect is more pronounced in non-state-owned enterprises and high-tech companies. The research results provide empirical guidance for information disclosure regulation and data governance in the digital economy.

Keywords: Data Assets; Stock Mispricing; Capital Market Efficiency; Digital Economy.

1. Introduction

Capital market pricing efficiency constitutes the cornerstone of financial stability and the core of resource allocation optimization. In the digital economy era, data has emerged as a pivotal production factor[1] and a critical source of enterprise value creation and competitive advantage[2]. However, the current accounting standards system lags severely in the recognition, measurement, and disclosure of data assets, with substantial key value information remaining concealed off-balance-sheet, thereby exacerbating information asymmetry.

Information asymmetry represents one of the key factors inducing stock mispricing[3]. Mispricing not only distorts corporate investment and financing decisions but also impairs the overall resource allocation efficiency of capital markets and accumulates systemic risk. Although research generally posits that enhanced information disclosure quality helps curb mispricing[4], existing literature primarily focuses on traditional financial information, with insufficient in-depth investigation into how data assets—a novel, complex, and value-driving type of idiosyncratic information—affect pricing efficiency.

This paper employs Chinese A-share listed companies from 2011–2024 as a sample to empirically examine the impact of data asset information disclosure on stock mispricing. Potential marginal contributions are twofold: First, extending the research perspective to the frontier domain of data assets to directly investigate their relationship with pricing efficiency; second, beyond single-mechanism analysis, constructing a systematic mechanism framework of "information generation-dissemination-internalization-verification" to reveal micro-level transmission channels. The research conclusions provide theoretical and empirical foundations for regulators to refine disclosure standards, for enterprises to optimize data governance, and for investors to enhance information interpretation capabilities.

2. Literature Review

(1) Data Assets and Economic Consequences

Data, as a key production factor in the digital economy era, has become a frontier academic issue regarding its capitalization and the economic consequences of its information disclosure. Existing research primarily centers on conceptual definitions, value assessment, and its broad-ranging impacts.

From an asset attribute perspective, data assets are defined as datasets with ownership rights, value, measurability, and readability. From an accounting viewpoint, emphasis is placed on their nature as non-monetary assets owned or controlled by enterprises that can generate economic benefits. However, data resource value assessment remains challenging, primarily employing the cost approach, income approach, and market approach[5]. At the macro level, scholars attempt to estimate national data capital stock[6]; at the micro level, textual analysis has become the mainstream method for measuring corporate data asset information disclosure levels[7].

Data asset information disclosure generates significant impacts across multiple dimensions. At the corporate level, data assets and their disclosure can enhance enterprise value through mechanisms including attracting technical human capital, alleviating financing constraints, optimizing supply chain configuration[8], and improving total factor productivity[9]. In financing, data assets help reduce corporate debt financing costs and cost of equity capital[10] while strengthening innovation capabilities[11]. At the capital market level, research focus gradually shifts toward pricing efficiency and stability. Scholars find that data asset information disclosure can reduce stock price synchronicity and enhance stock price informativeness[12]. Simultaneously, data assets help reduce stock price volatility and crash risk, bolstering market stability[13]. These studies provide an important foundation for this paper, yet systematic examination of how data asset information disclosure affects stock mispricing and its multidimensional transmission mechanisms remains inadequate.

(2) Measurement and Determinants of Stock Mispricing

Stock mispricing—the phenomenon where market prices deviate from intrinsic value—constitutes a core issue in asset pricing research. Existing literature explores this primarily from measurement methodology and driving factors perspectives.

Measurement approaches divide into indirect and direct methods. The indirect method commonly uses discretionary accruals as a proxy[14], though limited in indicating mispricing direction. Direct methods are more prevalent: first, relative valuation methods, such as comparing firm book-to-market ratios with industry means[15]; second, intrinsic value model approaches. The latter yields several mainstream models: (1) industry multiples-based valuation[16], which relies on industry averages and may contain industry-common noise; (2) residual income model[17], which utilizes book value and future earnings forecasts to estimate intrinsic value, with Frankel and Lee (1998)[18] first introducing analyst forecasts that may suffer optimistic bias[19]; (3) book-to-market ratio decomposition model[20], which decompose individual firms' market-to-book ratios into firm-level mispricing, industry-level mispricing, and long-term growth components, effectively isolating pricing biases from different sources. This model enjoys widespread application and recognition in domestic research[21] and serves as the core measurement approach adopted herein.

Regarding determinants, research primarily adopts information asymmetry and investor bounded rationality theoretical lenses. From the information asymmetry perspective, the information gap between corporate insiders and outside investors represents mispricing's root cause[22]. Consequently, improving information disclosure quality and strengthening internal/external governance are deemed critical corrective measures. Studies demonstrate that higher accounting information quality[3] and stronger internal controls[23] help reduce mispricing. Recently, corporate digital transformation has also been proven to suppress pricing biases by improving internal and external information environments[24]. From the investor bounded rationality perspective, cognitive biases, sentiment, and heterogeneous beliefs constitute behavioral roots causing prices to deviate from fundamentals[25]. Divergence of opinions under short-sale constraints induces overvaluation[26], while psychological biases such as investor overconfidence[27] and money illusion[28] also generate pricing bubbles. Additionally, certain corporate behaviors like complex textual disclosures[29] or earnings smoothing operations[30] may aggravate investor comprehension divergence, leading to mispricing.

3. Research Hypotheses

Based on information asymmetry theory, signaling theory, and behavioral finance theory, this paper constructs a theoretical analysis framework to investigate the impact of data asset information disclosure on stock mispricing and its transmission mechanisms. Specific hypotheses are formulated as follows.

H1: Data asset information disclosure exerts a significant inhibitory effect on stock mispricing.

The Efficient Market Hypothesis posits that asset prices should fully reflect all available information. However, as a novel and complex intangible asset, data assets' value assessment involves high internal specificity, creating severe information gaps between firms and the market. Proactive disclosure of data asset information enables firms to signal their digital technology capabilities, business models, and future growth potential to the market, directly alleviating information asymmetry. Simultaneously, clear and standardized information disclosure helps reduce investors' ambiguity aversion and cognitive biases, guiding them toward more rational valuation judgments. Therefore, high-quality data asset information disclosure is expected to enrich the market's idiosyncratic information set, correcting valuation biases stemming from information scarcity or misinterpretation, thereby suppressing stock mispricing.

H2: Data asset information disclosure inhibits stock mispricing by deepening corporate digital transformation.

The value release of data assets heavily relies on firms' digital infrastructure and governance capabilities. Information disclosure itself creates pressure for firms to improve data governance systems, thereby driving digital transformation deepening. Substantive digital transformation can optimize internal information production processes, enhance operational transparency, and strengthen corporate governance, fundamentally improving data asset information quality and credibility. When high-quality data asset information generated from solid digital transformation is disclosed, its signal-to-noise ratio is lower, enabling more effective transmission of firm-specific information to the market, compressing valuation ambiguity space, and consequently improving pricing efficiency.

H3: Data asset information disclosure inhibits stock mispricing by attracting analyst coverage.

The professional and complex nature of data asset information requires interpretation and dissemination by specialized information intermediaries. Detailed data asset disclosure provides analysts with unique incremental information, attracting their attention and in-depth research. By publishing research reports and adjusting valuations, analysts can reduce information processing costs for ordinary investors and accelerate the diffusion and penetration of idiosyncratic information in the market. Enhanced analyst coverage signifies improved information dissemination efficiency, thereby effectively mitigating market pricing biases caused by insufficient information interpretation or inadequate dissemination.

H4: Data asset information disclosure inhibits stock mispricing by improving corporate ESG ratings.

Information regarding data governance, privacy protection, and digital empowerment of society and environment within data asset disclosures constitutes an important dimension for evaluating firms' sustainable development capabilities. Adequate disclosure of such information helps improve firms' performance in ESG rating systems. Higher ESG ratings convey strong reputation signals of long-term value orientation, sound risk management, and good social responsibility fulfillment to the market, enhancing investors' long-term confidence and reducing firms' financing risk premiums. This guides valuations to revert toward intrinsic value reflecting sustainable development potential, suppressing pricing biases.

H5: Data asset information disclosure inhibits stock mispricing by curbing real earnings management.

High-quality data asset information disclosure enhances overall corporate information transparency, forming stronger external monitoring constraints on managerial opportunistic behavior. Management's motivation and ability to manipulate profits through real earnings management are consequently suppressed. Reduced real earnings management signifies lower information distortion in financial reporting and improved accounting information quality. More truthful and reliable financial information provides investors with a firmer pricing foundation, reducing valuation noise generated by information manipulation, thereby helping alleviate stock mispricing.

4. Research Design

(1) Sample Selection and Data Sources

Based on data availability and the developmental timeline of data asset-related concepts, this paper selects Chinese A-share listed companies from 2010–2024 as the research sample. Considering the delayed release of annual reports and their consequent lagged impact on stock prices, the explanatory variable is lagged by one period. Data are sourced from the CSMAR Database.

The sample is processed as follows: excluding financial listed companies; eliminating observations with missing key variables; removing firms that underwent ST, *ST, or delisting special treatment during listing; excluding insolvent firms with asset-liability ratios greater than 1; deleting observations with less than one year of listing age. Concurrently, to mitigate the influence of extreme values, all continuous variables are Winsorized at the 1% and 99% percentiles. Following these screenings, the final sample comprises 4,086 listed companies, yielding 27,850 firm-year observations.

(2) Variable Definitions

a. Dependent Variable: Stock Mispricing

The dependent variable is the level of stock mispricing. Following Li Shanmin et al. (2020)[21], we measure mispricing by decomposing the market-to-book ratio (M/B), where M represents market value and B represents book value, introducing V to denote intrinsic value:

$$\frac{M}{B}=\frac{M}{V}\times\frac{V}{B} \tag{1}$$

Taking logarithms of both sides of Equation (1), let $\mathrm{m}=\ln\mathrm{M}$, $\mathrm{v}=\ln\mathrm{V}$, $\mathrm{b}=\ln\mathrm{B}$:

$$\mathrm{m}-\mathrm{b}=(\mathrm{m}-\mathrm{v})+(\mathrm{v}-\mathrm{b}) \tag{2}$$

Given that markets are not perfectly efficient, stock market prices deviate due to information asymmetry, investor behavioral biases, and other factors. The term m-v captures the deviation between a firm's market value and its intrinsic value, measuring stock mispricing. The term v-b captures the deviation between intrinsic value and book value, reflecting the firm's growth potential.

Subsequently, for firm i in industry j at time t, introducing firm financial information θ and its estimated coefficient α, Equation (2) is further decomposed into (3):

$$\begin{aligned} m_{i,t}-b_{i,t} &= [m_{i,t}-v(\theta_{i,t};\alpha_{j,t})] \\ &+[v(\theta_{i,t};\alpha_{j,t})-v(\theta_{i,t};\alpha_j)]+[v(\theta_{i,t};\alpha_j)-b_{i,t}] \end{aligned} \tag{3}$$

Where v is expressed as a linear function of firm financial information θ and coefficient α; $v(\theta_{i,t};\alpha_{j,t})$ represents firm's contemporaneous value estimated using industry's time-coefficient $\alpha_{j,t}$ *, with* $m_{i,t}-v(\theta_{i,t};\alpha_{j,t})$ reflecting the deviation between stock price and contemporaneous industry estimates, representing cross-sectional mispricing (firm-level mispricing). $v(\theta_{i,t};\alpha_j)$ denotes firm's long-term value estimated using industry's long-term coefficient α_j*, with* $v(\theta_{i,t};\alpha_{j,t})-v(\theta_{i,t};\alpha_j)$ reflecting the deviation between contemporaneous and long-term industry estimates, representing longitudinal mispricing (industry-level mispricing). $v(\theta_{i,t};\alpha_j)-b_{i,t}$ reflects the deviation between long-term estimated value and book value, measuring growth potential.

Model (4) is used to estimate $v(\theta_{i,t};\alpha_{j,t})$ and $v(\theta_{i,t};\alpha_j)$:

$$\begin{aligned} m_{i,t} &= \alpha_{0\,j,t}+\alpha_{1\,j,t}b_{i,t}+\alpha_{2\,j,t}\ln(NI)^+_{i,t} \\ &+\alpha_{3\,j,t}I_{(<0)}\ln(NI)^+_{i,t}+\alpha_{4\,j,t}LEV_{i,t}+\varepsilon_{i,t} \end{aligned} \tag{4}$$

Where $m_{i,t}$ is firm's market value at year-end t; $b_{i,t}$ is book value; $\ln(NI)^+_{i,t}$ is the natural logarithm of net profit absolute value; $I_{(<0)}$ is an indicator function equal to 1 if net profit is negative and 0 otherwise; $LEV_{i,t}$ is the asset-liability ratio; $\varepsilon_{i,t}$ is the regression residual.

Regressing using firms in industry j at year t yields estimated coefficients ($\alpha_{0\,j,t},\alpha_{1\,j,t},\alpha_{2\,j,t},\alpha_{3\,j,t}$), enabling calculation of contemporaneous estimated value $v(\theta_{i,t};\alpha_{j,t})$. For long-term industry estimates, coefficients are averaged annually: $\bar{\alpha}_j=\frac{1}{T_j}\sum\hat{\alpha}_{j,t}$, yielding long-term coefficients ($\bar{\alpha}_{0\,j},\bar{\alpha}_{1\,j},\bar{\alpha}_{2\,j},\bar{\alpha}_{3\,j},\bar{\alpha}_{4\,j}$) and long-term estimate $v(\theta_{i,t};\alpha_{j,t})$.

This paper focuses on stock mispricing, making firm-level mispricing ($m_{i,t}-v(\theta_{i,t};\alpha_{j,t})$) the core research object, hereafter uniformly denoted as $Misprice_{i,t}$. Using $v(\theta_{i,t};\hat{\alpha}_{j,t})$ from Equation (4), we calculate $Misprice_{i,t}=m_{i,t}-v(\theta_{i,t};\hat{\alpha}_{j,t})$.

A positive $Misprice_{i,t}$ indicates overvaluation, while negative indicates undervaluation. To verify whether data asset disclosure reduces mispricing magnitude, we use the absolute value $|Mispricei,t|$ as the mispricing measure, where larger values indicate greater mispricing.

b. Core Explanatory Variable: Data Asset Information Disclosure

Following Wei Yanlin et al. (2022)[31], we measure data asset information disclosure using annual report textual analysis with a "seed word set + Word2Vec similarity expansion" approach.

The measurement proceeds in three steps: First, based on the definition of data assets in the 2019 *White Paper on Data Asset Management Practice* released by the China Academy of Information and Communications Technology, we select "data resources" and "data assets" as seed words and apply deep learning technology to obtain similar word sets. Words with similarity >0.5 are retained to form a data asset information disclosure dictionary. Finally, using this dictionary, we calculate the frequency of relevant terms in annual reports, summing occurrences and dividing by total annual report word count multiplied by 100 to obtain the disclosure index, where higher values indicate greater disclosure levels.

$$DAD_{i,t}=\frac{\sum Dictionarywords_{i,t,n}}{TotalWords_{i,t}}\times 100 \tag{5}$$

c. Control Variables

Following You Jiaxing and Wu Jing (2012)[32] and Xu Shoufu et al. (2023)[33], control variables are selected from several dimensions. Fundamental characteristics: firm size (Size), return on assets (ROA), leverage (LEV), revenue growth (Grow), and listing age (Age). Governance characteristics: state ownership (SOE), board size (Boardsize), proportion of independent directors (Indir), and CEO duality (DUAL). Ownership structure: proportion of tradable shares (Outshare), ownership concentration (Stock), and ownership balance (Balance).

(3) Model Specification

To test whether data asset information disclosure affects stock mispricing, we construct the following fixed-effects model:

$$\left|Misprice_{i,t}\right| = \beta_0 + \beta_1 DAD_{i,t-1} + \beta_2 Controls_{i,t} + Industry + Year + \varepsilon_{i,t} \quad (6)$$

Table 1. Variable Selection and Definitions

Variable Name	Variable Code	Variable Definition
Stock Mispricing	|Misprice|	Estimated using the above model
Data Asset Information Disclosure Level	DA	Ratio of total frequency of "data asset" keywords in annual reports to total word count
Firm Size	Size	Natural logarithm of total assets
Return on Assets	ROA	Net profit/Total assets
Asset-Liability Ratio	LEV	Total liabilities/Total assets
Revenue Growth Rate	Grow	(Current year revenue - Previous year revenue)/Previous year revenue
Listing Age	Age	ln(Sample current year - Listing year + 1)
Ownership Nature	SOE	Equals 1 for state-owned enterprises, 0 otherwise
Board Size	Boardsize	Number of board directors
Proportion of Independent Directors	Indir	Independent directors/Board size
CEO Duality	DUAL	Equals 1 if CEO chairs the board, 0 otherwise
Proportion of Tradable Shares	Outshare	Tradable shares/Total shares
Ownership Concentration	Stock	Shareholding percentage of largest shareholder
Ownership Balance	Balance	Ratio of second-to-fifth largest shareholders' holdings to largest shareholder's holdings

Where $\left|Misprice_{i,t}\right|$ is firm's mispricing level at year-end *t*, $DAD_{i,t-1}$ is the disclosure level at year-end t-1, $Controls_{i,t}$ are control variables, *Industry* represents industry fixed effects, *Year* denotes year fixed effects, and $\varepsilon_{i,t}$ is the regression residual.

5. Empirical Results and Analysis

(1) Data Asset Information Disclosure and Stock Mispricing

Table 2. Regression Results of Data Asset Disclosure and Stock Mispricing

	(1)	(2)
	|Misprice|	|Misprice|
DAD	-0.705***	-0.766***
	(-3.26)	(-3.64)
Controls	No	Yes
Industry	Yes	Yes
Year	Yes	Yes
N	27850	27850
Adj_R²	0.031	0.057

*Notes: All regressions use firm-level clustered robust standard errors; t-statistics in parentheses are firm-level clustered; *, **, *** denote significance at 1%, 5%, 10% levels.

Table 2 reports baseline regression results, with all specifications controlling for industry and year fixed effects. Column (1) excludes control variables. Results preliminarily indicate that increased data asset disclosure levels help reduce stock mispricing, supporting the core hypothesis. Column (2) includes control variables. After controlling for firm size, profitability, leverage, growth, listing age, ownership nature, and corporate governance characteristics, the inhibitory effect remains robust.

By disclosing data assets as key idiosyncratic information, firms can signal their digital capabilities and growth potential to the market, alleviating information asymmetry and enabling stock prices to more fully reflect intrinsic value, thereby suppressing overall pricing biases.

(2) Mechanism Testing of Data Asset Disclosure Impact

While previous sections analyzed data assets' impact on mispricing, the underlying mechanisms remain unexplored. Building on our theoretical framework, this study examines four dimensions: information generation, dissemination, internalization, and verification. Following Jiang Ting (2022)[34] discussion on mechanism analysis, we focus on data assets' effects on key mechanism variables, constructing the following model:

$$Mediator_{i,t} = \beta_0 + \beta_1 DAD_{i,t-1} + \beta_2 Controls_{i,t} + Industry + Year + \varepsilon_{i,t} \quad (7)$$

Where $\mathrm{Mediator}_{i,t}$ represents mechanism variables, with other variables consistent with the baseline model.

a. Information Generation—Digital Transformation Index

The Digital Transformation Index (Digital) comprehensively measures firms' digital strategy implementation, technology adoption, and organizational change. Table 3 Column (1) shows that data asset disclosure significantly and positively affects digital transformation, indicating that forward-looking strategic information release drives deeper digital transformation in strategic planning, technology adoption, and organizational restructuring. Data asset value cannot exist in isolation from systematic digital infrastructure and governance capabilities. Firms' digital transformation level determines whether raw, disordered data resources can be converted into measurable, verifiable, trustworthy high-quality information, providing an effective pricing basis for capital markets.

b. Information Dissemination—Analyst Coverage

Analyst coverage (Analyst) measures external information intermediaries' research attention. Following existing literature, we use the annual number of analyst teams publishing research reports on the firm (log-transformed) as a proxy. Table 3 Column (2) shows that data asset disclosure effectively attracts more analyst tracking and research, significantly enhancing capital market information supply. As important intangible assets, data assets are crucial for investors to assess future growth potential and performance. Analysts focus more on firms with higher data asset capitalization. Meanwhile, analysts possess significant advantages in extracting company-specific information and disseminating it; increased coverage helps reduce stock mispricing.

c. Information Internalization—ESG Performance

ESG ratings comprehensively assess firms' sustainable development performance across environmental, social, and governance dimensions. Table 3 Column (3) indicates that good ESG performance makes data assets' implicit value explicit and consensus-based, bridging firm data resources and capital markets. ESG rating agencies integrate fragmented information through systematic indicator frameworks, validating and converting it into comparable scores. ESG performance conveys multiple positive signals beyond short-term financial data, effectively alleviating investor concerns about complexity and uncertainty, guiding valuations closer to long-term intrinsic value.

d. Information Verification—Earnings Management

For real earnings management (REM), following Dechow (1998) [35], we use a composite measure of abnormal operating cash flow (abffo), abnormal production costs (abpcost), and abnormal discretionary expenses (abdcost). Table 3 Column (4) shows that data asset disclosure significantly suppresses earnings management activities. As disclosure levels increase, data readability and availability improve, enabling external auditors and regulators to access key information more conveniently, imposing more effective constraints on earnings management, weakening managerial manipulation motives, enhancing information reliability, and reducing valuation noise.

Table 3. Mechanism Testing Results

	(1)	(2)	(3)	(4)
	Digital	**Analyst**	**ESG**	**REM**
DAD	243.368***	2.919***	4.248***	-0.624***
	(23.24)	(3.94)	(5.06)	(-3.96)
Constant	8.135***	-7.289***	-2.348***	-0.013
	(2.95)	(-33.25)	(-9.54)	(-0.26)
Controls	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
N	27850	19794	24475	27558
Adj_R²	0.362	0.369	0.184	0.238

*Notes: t-statistics in parentheses are firm-level clustered; *, **, *** denote significance at 1%, 5%, 10% levels.

(3) Robustness Tests

a. Instrumental Variable Approach

To address endogeneity concerns, we employ instrumental variable estimation. Following Yuan Zeming et al. (2024)[8], we use regional mobile phone penetration rates (IV) as an instrument satisfying relevance and exogeneity conditions. Table 4 Column (1) shows the instrument significantly positively affects data assets. Column (2) confirms the significant inhibitory effect on mispricing, consistent with baseline results. Validity tests include the LM statistic (10.416), rejecting underidentification, and Wald F statistic (10.560), rejecting weak instrument concerns.

Table 4. Instrumental Variable Regression Results

	(1)	(2)
	First stage	**Second stage**
IV	0.0000205***	
	(3.250)	
DAD		-12.1522*
		(-1.825)
Controls	Yes	Yes
KP LM Statistic		10.416
KP Wald F Statistic		10.560
Industry	Yes	Yes
Year	Yes	Yes
N	27846	27846
Adj_R²	0.3725	-0.1764

*Notes: t-statistics in parentheses are firm-level clustered; *, **, *** denote significance at 1%, 5%, 10% levels.

b. Alternative Variable Measurement

First, we change the mispricing measure. To exclude potential estimation bias from specific methods, we use the deviation between firm book-to-market ratio and industry median:

$$\mathrm{Misprice}_{i,t} = \ln\frac{\mathrm{MB}_{i,t}}{\mathrm{MB}_{j,t}} \tag{8}$$

Where $\mathrm{MB}_{i,t}$ is firm's book-to-market ratio and $MB_{j,t}$ is the industry median. Results remain significant. Second, we alter the disclosure measure, using the natural logarithm of (total keyword frequency + 1) to address distributional bias. The inhibitory effect remains robust under alternative measures.

Table 5. Robustness Test Results

	(1)	(2)
	Misprice2	**DAD2**
DAD	-0.644***	-0.008**
	(-2.79)	(-2.14)
Controls	Yes	Yes
Industry	Yes	Yes
Year	Yes	Yes
N	27487	27850
Adj_R²	0.128	0.057

*Notes: t-statistics in parentheses are firm-level clustered; *, **, *** denote significance at 1%, 5%, 10% levels.

(4) Heterogeneity Analysis

a. Ownership

Ownership nature shapes corporate disclosure motives, behavioral patterns, and market reactions. We partition the sample into state-owned (SOE) and non-SOE groups. Table 6 Columns (1)-(2) show the inhibitory effect exists primarily in non-SOE, with no significant effect in SOE.

This difference stems from systematic variations in operational objectives, governance environments, and market constraints. Non-SOE face fiercer market competition and stricter external monitoring, making their financial information and decisions more market-oriented. In this context, high-quality data asset disclosure serves as a positive strategic signal that significantly reduces information asymmetry and corrects pricing biases. Conversely, SOE already possess more standardized mandatory disclosure foundations, with stock prices heavily influenced by non-market factors like industrial policies. Their diversified social objectives may dilute the marginal effect of enhanced disclosure on pricing efficiency.

b. Industry Technology Characteristics

To test whether effects vary by technology intensity, following Peng Hongxing and Mao Xinshu (2017)[36], we classify firms with codes C25–C29, C31–C32, C34–C41, I63–I65, and M73 as high-tech industries. Table 6 Columns (3)-(4) show significant industry heterogeneity: disclosure effectively suppresses mispricing in high-tech industries but shows no effect in non-high-tech sectors.

High-tech industries typically center on technological innovation and data-driven business models, where data assets are key competitive elements. Their disclosures are more substantive, proprietary, and verifiable, conveying high-confidence signals about core competencies and growth potential that investors effectively integrate into valuation models. In contrast, data assets may not be strategic resources in traditional non-tech industries, making disclosures formalistic and lacking quantifiable substance, thus failing to serve as effective information intermediaries.

Table 6. Heterogeneity Analysis Results

	(1)	(2)	(3)	(4)
	SOE	**Non-SOE**	**High-tech**	**Non-high-tech**
DAD	-0.660	-0.821***	-0.894***	0.056
	(-1.45)	(-3.47)	(-3.89)	(0.09)
Constant	-0.212*	-0.321***	-0.403***	-0.062
	(-1.78)	(-2.75)	(-3.74)	(-0.51)
Controls	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
N	10912	16937	15706	12144
Adj_R²	0.063	0.062	0.056	0.070

*Notes: t-statistics in parentheses are firm-level clustered; *, **, *** denote significance at 1%, 5%, 10% levels.

6. Conclusion and Implications

Using Chinese A-share listed companies from 2011–2024, this paper empirically examines the impact of data asset information disclosure on stock mispricing. Findings reveal that data asset disclosure significantly suppresses mispricing, alleviating information asymmetry to enable stock prices to more accurately reflect intrinsic value and improve capital market pricing efficiency. The mechanism operates through four pathways: First, driving digital transformation. Disclosure creates pressure to improve data governance, deepen digital strategy, optimize internal processes, and enhance information quality. Second, attracting analyst coverage. The professional complexity of data assets requires interpretation by information intermediaries; detailed disclosure increases analyst tracking, accelerating information diffusion and reducing investor processing costs. Third, improving ESG ratings. Data governance and privacy protection are key ESG dimensions; good ESG performance conveys sustainable development signals, enhancing long-term confidence and reducing valuation divergence. Fourth, curbing real earnings management. High-quality disclosure improves transparency, strengthening external monitoring constraints, weakening manipulation motives, and enhancing accounting reliability. The inhibitory effect is more pronounced under stronger external monitoring or where data assets hold greater strategic value.

For government: Construct a systematic data asset disclosure and evaluation framework. Accelerate legislative and regulatory improvements; establish clear assessment, transaction, and protection mechanisms; incentivize compliant disclosure through fiscal and tax benefits; strengthen data infrastructure; promote industry-academia-research collaboration; implement differentiated guidance prioritizing high-tech industries, non-SOE, and digitally underdeveloped regions.

For enterprises: Integrate data assets into core strategy; clarify management structures and responsibilities; establish lifecycle governance systems; actively apply big data and AI to enhance analytical capabilities; complete data security and privacy protection mechanisms to prevent misuse risks.

For investors: Enhance professional interpretation and pricing capabilities. Institutional investors should deeply mine data assets' economic substance to identify valuation biases; individual investors should monitor disclosure quality changes, incorporating them into value investment frameworks; both can express data governance concerns through shareholder meetings to drive quality improvements.

References

[1] Cai Jiming, Liu Yuan, Gao Hong, et al. The Approach of Data Factor Participating in Value Creation: A General Equilibrium Analysis Based on the General Theory of Value [J]. Journal of Management World, 2022;38:108–21.

[2] Li Haijian, Zhao Li. Data Becomes a Factor of Production: Characteristics, Mechanisms, and the Evolution of Value Form [J]. Shanghai Journal of Economics, 2021:48–59.

[3] Xu Shoufu, Xu Longbing. Information Disclosure Quality and Misvaluation in Capital Market [J]. Accounting Research, 2015:40–7, 96.

[4] Jin L, Myers S. R2 around the world: New theory and new tests. [J]. Journal of Financial Economics, 2006;79:257–92.

[5] Lu Zheng, Zhou Ting, Wang Li, et al. Data assets and firm development: empirical evidence from Chinese listed companies [J]. Industrial Economics Research, 2023:128–42.

[6] Xu Xiang, Zhao Mofei. Data Capital and Economic Growth Path [J]. Economic Research Journal, 2020;55:38–54.

[7] He Ying, Chen Lili, Du Yaguang. Does Data Assetization Alleviate Financing Constraints of SRDI SMEs [J].China Industrial Economics, 2024:154–73.

[8] Yuan Zeming, Song Yuqian, Yu Xiang. Data Asset Sharing, Diversified Supply Chain Configuration and Enterprise Value [J]. Statistics & Decision, 2024;40:172–7.

[9] Yao Huina, Zhang Jinchang. A Study on the Influence of Corporate Data Assets on Total Factor Productivity [J]. Economic Survey, 2024;41:107–19.

[10] Niu Biao, Yu Xiang. Has the Data Asset Gained Favor among Investors? - From the Perspective of Equity Capital Cost [J]. Securities Market Herald, 2024:68–79.

[11] Li Jian, Dong Xiaofan, Zhang Jinlin, et al. The Impact of Data Assets on Enterprise Innovation Investment [J]. Foreign Economics & Management, 2023;45:18–33.

[12] Li Shigang, Shao Hongbin, Fang Fang, et al. Firm Data Asset Disclosure and Capital Market Pricing Efficiency [J].China Industrial Economics, 2025:138–55.

[13] Niu Biao, Yu Xiang, Yuan Zeming. Enterprise Data Assetization and Capital Market Stability [J]. Journal of Modern Finance, 2025;30:53–63, 77.

[14] Polk C, Sapienza P. The stock market and corporate investment: A test of catering theory [J]. Review of Financial Studies, 2009;22:187–217.

[15] Lin JB, Pantzalis C, Park JC. Corporate hedging policy and equity mispricing [J]. Financial Review; 45:803–24.

[16] Berger PG, Ofek E. Diversification's effect on firm value [J]. Journal of Financial Economics, 1995;37:39–65.

[17] Ohlson JA. Earnings, book values, and dividends in equity valuation* [J]. Contemporary Accounting Research, 1995; 11: 661-87.

[18] Frankel R, Lee CMC. Accounting valuation, market expectation, and cross-sectional stock returns [J]. Journal of Accounting and Economics, 1998;25:283–319.

[19] Rao Pingui, Yue Heng. Residual Income Model and Stock Future Returns [J]. Accounting Research, 2012:52–8, 97.

[20] Rhodes-Kropf M, Robinson DT, Viswanathan S. Valuation waves and merger activity: The empirical evidence [J]. Journal of Financial Economics, 2004:0–43.

[21] Li Shanmin, Huang Zhihong, Guo Jingjing. Capital Market Pricing and Corporate Acquisitions: Evidence from Chinese A-share Firms [J]. Economic Research Journal, 2020;55:41–57.

[22] Myers SC, Majluf NS. Corporate financing and investment decisions when firms have information that investors do not have [J]. Journal of Financial Economics, 1984;13:187–221.

[23] Wang Shengnian, Wang Songhe. Internal Governance and Asset Mispricing: A Perspective of Internal Control System Construction [J]. Modern Finance and Economics-Journal of Tianjin University of Finance and Economics, 2018;38:95–112.

[24] Li Zhenzhen, Wang Aidong, Li Haijian. Can Corporate Digital Transformation Curb Stock Mispricing? [J]. Contemporary Finance & Economics, 2023:133–43.

[25] Barberis N, Shleifer A, Vishny R. A model of investor sentiment1 [J]. Journal of Financial Economics, 1998;49:307–43.

[26] Miller EM. Risk, uncertainty, and divergence of opinion [J]. The Journal of Finance, 1977;32:1151–68.

[27] Wu Weixing, Wang Yongxiang, Liang Hengyi. Overconfidence Limited Participation and Asset Price Bubbles [J]. Economic Research Journal, 2006:115–27.

[28] Chen Guojin, Zhang Yijun, Wang Jing. Resale Option Inflation Illusion and the Chinese Stock Market Bubbles [J]. Economic Research Journal, 2009;44:106–17.

[29] He Kang, Song Bingjie, Liu Wei. Complexity of Annual Report and Asset Mispricing--Empirical Research Based on Text Analysis [J]. Collected Essays on Finance and Economics, 2020: 64–73.

[30] Zhang Jing, Wang Shengnian. Earning Smoothing, Investor Heterogeneous Beliefs and Asset Mispricing: An Analysis based on 2000-2012 Data of Shanghai and Shenzhen Stock Markets [J]. Commercial Research, 2017:53–9, 187.

[31] Wei Yanlin, Zhang Junrui, Wang Fangjun, et al. Research of data assets disclosure and analyst forecasts: Evidence from text analysis [J]. Journal of Industrial Engineering / Engineering Management, 2022;36:130–41.

[32] You Jiaxing, Wu Jing. Spiral of Silence: Media Sentiment and the Asset Mispricing [J]. Economic Research Journal, 2012; 47: 141-52.

[33] Xu Shoufu, Zheng Yingfei, Zhang Jiachen. Interaction Online, Strategic Response of Listed Company and Stock Mispricing [J]. Business and Management Journal, 2023;45:189–208.

[34] Jiang Ting. Mediating Effects and Moderating Effects in Causal Inference [J]. China Industrial Economics, 2022:100–20.

[35] Dechow PM, Hutton AP, Kim JH, et al. Detecting earnings management: A new approach[J]. Journal of Accounting Research, 2012.

[36] Peng Hongxing, Mao Xinshu. Government Subsidies for Innovation, Company Executives Background and R&D Investment-Evidence from the High-tech Industry [J]. Finance & Trade Economics, 2017;38:147–61.