Is Universal Disclosure Equivalent to No Disclosure? Data Asset Disclosure, Investors' Limited Attention and Capital Market Pricing Efficiency

Linjing Peng, Queens Yukun Yang[†]

Abstract: This paper utilizes data from Chinese A-share listed companies and, based on the limited attention theory, re-examines the impact of data asset disclosure on capital market pricing efficiency. The study finds that the economic consequences of data asset disclosure are not static but exhibit a significant marginal decreasing feature as the overall market disclosure level rises. Specifically, when the overall market disclosure level is low, corporate disclosure of data asset information can significantly reduce stock price synchronization and enhance capital market pricing efficiency by increasing the supply of idiosyncratic information, enhancing information diffusion, and accelerating information integration. However, with the recent widespread disclosure behavior, there is no evidence that the above mechanisms remain effective. Further analysis indicates that disclosure behavior can maintain its positive effect only when the overall market disclosure level of data assets is low, and the firm's disclosed idiosyncratic information is relatively limited, the information dissemination environment is favorable, or institutional investor attention is not excessively diffused. This paper expands the boundaries of the limited attention theory, extending the research perspective to the dimension of the dispersion of investor attention by homogenized information, and reveals the value dilution problem faced by data asset information disclosure in the context of universalization.

Keywords: Data asset disclosure, Limited attention theory, Capital market pricing efficiency, Stock price synchronicity, Value dilution

1 Introduction

In recent years, as data has been formally established as a key factor of production, the management, development, and value realization of data resources by enterprises have become pivotal topics in the era of the digital economy (Jones and Tonetti, 2020). The "Interim Provisions on Accounting Treatment of Enterprise Data Resources[①]" (hereafter referred to as the "Interim Provisions"), implemented in China in 2024, provides normative guidance for the accounting treatment of corporate data resources (Huang et al., 2025). However, constrained by the recognition criteria under accounting standards, data assets disclosed by enterprises in financial statements are primarily measured at historical cost, which fails to fully reflect the added value

[†] Information of authors: Linjing Peng is currently a student majoring in Accounting at the School of Management and Economics, Beijing Institute of Technology, with an expected graduation date of July 2027. E-mail: Linda13677066050@163.com. Queens Yukun Yang is currently a student majoring in Information Management and Information Systems at the School of Management and Economics, Beijing Institute of Technology, with an expected graduation date of July 2027. E-mail: yangyukun2005@163.com. Yukun Yang is the corresponding author of this paper. (Information as of February 7, 2026)
The latest version of the paper can be obtained by visiting Queens Yukun Yang's personal webpage: https://yangyukun2005.github.io/publications/.
① https://www.gov.cn/zhengce/zhengceku/202308/content_6899395.htm

and future potential of data resources (Wang and Yang, 2024). Therefore, the voluntary disclosure of data asset-related information in the text of annual reports can provide external investors with significant incremental information (Bochkay et al., 2023; Li et al., 2025). The "Interim Provisions" also encourage enterprises to voluntarily disclose information regarding application scenarios and usage effectiveness for data resources not recognized as intangible assets or inventories, based on their actual circumstances.

Against this backdrop, Li et al. (2025) utilized machine learning methods to analyze the annual reports of Chinese A-share listed companies from 2007 to 2022, measuring their data asset information disclosure levels (as well as stock price synchronicity). Their research indicates that corporate data asset information disclosure can reduce stock price synchronicity and improve capital market pricing efficiency. However, while researching by using the data publicly provided by Li et al. (2025) on the official website of *China Industrial Economics*,[②] we found that the relationship between data asset information disclosure and capital market pricing efficiency exhibits significant time-varying characteristics. When the test is conducted using only samples from recent years (e.g., 2020–2022), the impact of data asset information disclosure on capital market pricing efficiency is no longer statistically significant. This finding challenges the prevailing perception of the effects of data asset information disclosure in existing research, suggesting a need to re-examine the economic consequences of such disclosures.

Stock price synchronicity is an essential proxy for measuring capital market pricing efficiency, reflecting the extent to which firm-specific information is integrated into stock prices (Chan and Hameed, 2006). According to the "information efficiency hypothesis," higher stock price synchronicity indicates that prices primarily reflect market or industry-wide information rather than the firm's own operational conditions, representing lower capital market pricing efficiency (Morck et al., 2000). Existing research suggests that textual disclosures—such as social responsibility reports (Wang et al., 2014), internal control reports (Zhao and Liu, 2024), and Management Discussion and Analysis (Song et al., 2024)—contribute to enhancing capital market pricing efficiency.

However, unlike traditional disclosures centered on compliance or accountability, data asset disclosure directly manifests a firm's capability to integrate and apply data factors in the digital economy. It serves as a vital signal of digital competence, technical prowess, and innovation potential (Yuan et al., 2022). Nevertheless, current literature predominantly focuses on the independent effect of individual firm disclosure, overlooking the role of the overall market disclosure level. In reality, acquiring and processing information is not costless (Blankespoor et al., 2020). Based on the limited attention theory, investors possess finite cognitive resources and processing capacities, making it impossible for them to attend to and process all available information simultaneously (Hirshleifer et al., 2011; Hirshleifer and Teoh, 2003; Peng and Xiong, 2006). Consequently, when a specific type of information is widely disclosed across the market, investors' attention resources may be dispersed, diminishing the marginal value of any single firm's disclosure. Furthermore, as more firms disclose, information homogenization intensifies, potentially drowning out firm-specific signals in overall market noise and hindering the efficient incorporation of information into stock prices.

With the vigorous development of the digital economy, profound changes are taking place in

---

② *China Industrial Economics* is a top economics journal in Chinese, the webpage of their paper is: https://ciejournal.ajcass.com/Magazine/show/?id=119653

the market environment and information disclosure practices, making it increasingly common for listed companies to disclose data assets (Liu et al., 2026). Li et al. (2025), Zhang and Zhang (2025), and Wei et al. (2022) adopted similar methods based on machine learning models to analyze financial reports and measure the data asset information disclosure levels of Chinese A-share listed companies. Furthermore, both Li et al. (2025) and Zhang and Zhang (2025) cited the paper by Wei et al. (2022) as a reference. The annual average measurement results by Zhang and Zhang from 2007 to 2022 are shown in Figure 1; Li et al. (2025) used two indicators to measure corporate data asset disclosure levels, and their annual averages from 2007 to 2022 are shown in Figure 2. [③] Their studies all indicate that corporate data asset information disclosure levels have significantly increased over time. This change may alter the informational value and market response of data asset disclosure: when all listed companies disclose data asset information with similar content, investors may reduce their attention to individual company disclosures or perceive data asset information as an industry norm rather than a firm-specific trait, thereby weakening its value as a differentiating signal. This phenomenon can be summarized as "universal disclosure equivalent to no disclosure."

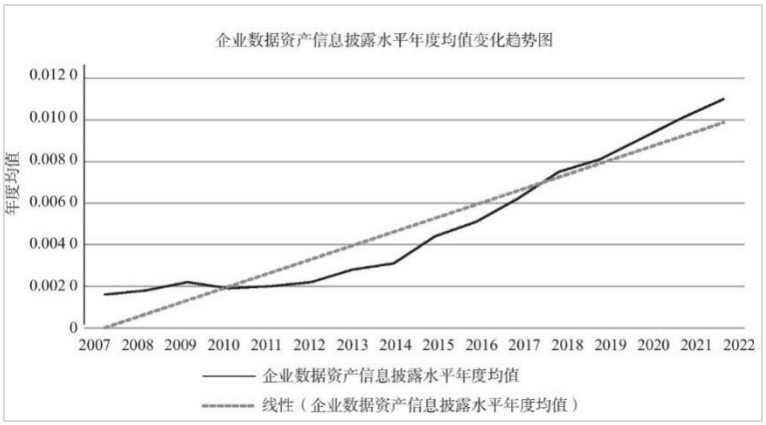

Figure 1. The annual average of Zhang and Zhang's measurement results on data asset disclosure levels
Note: This figure was taken directly from Zhang and Zhang's paper without any modifications.

[③] Zhang and Zhang (2025) first designated "data resources" and "data assets" as seed words. They then utilized the Word2Vec neural network model to identify a set of related terms, retaining those with a cosine similarity greater than 0.5 to construct a comprehensive dictionary. Finally, based on the text of annual reports from listed companies, they calculated the index of corporate data asset disclosure by counting the frequency of each dictionary term. This was achieved through a weighted summation—where weights represent the similarity of each term to the seed words—which was then normalized by dividing by the total word frequency of the report. The methodology and the two specific indicators developed by Li et al. (2025) will be introduced in detail in the subsequent sections.

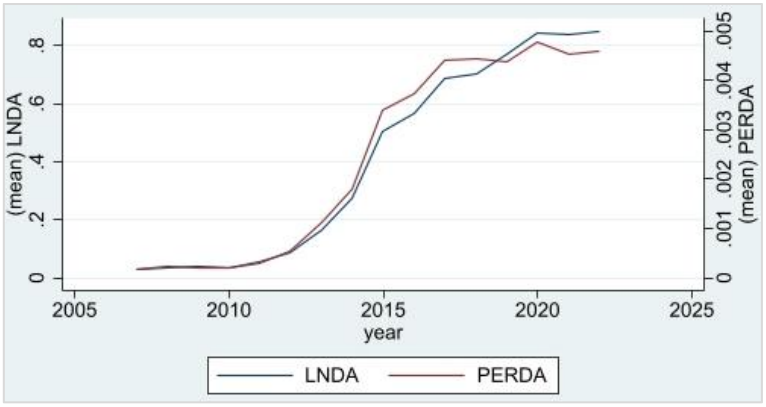

Figure 2. The annual average of Li et al.'s measurement results on data asset disclosure levels

Note: This figure was drawn by us based on data published by Li et al.

Therefore, to test this proposition, this paper also utilizes data from A-share listed companies from 2007 to 2022 and follows the methodology of Li et al. (2025) to construct an indicator for the overall market level of data asset information disclosure, examining the "moderating effect" of overall disclosure quality on the effectiveness of individual firm disclosure. Empirical results show that as the overall market disclosure level continues to rise, the marginal impact of individual firm data asset disclosure on capital market pricing efficiency significantly decreases; furthermore, in recent years when the overall disclosure level has been high, its independent effect is no longer statistically significant. Mechanism tests indicate that when the overall disclosure level is low, corporate data asset disclosure can reduce stock price synchronicity through channels such as increasing information supply, enhancing information diffusion, and accelerating information integration; however, when the overall disclosure level is high, these transmission mechanisms are significantly weakened. Heterogeneity analysis further reveals that data asset disclosure can significantly reduce stock price synchronicity only in scenarios where the overall disclosure level is low, information homogenization is low, the information dissemination environment is favorable, and institutional investors' attention is not dispersed. Overall, from the perspectives of limited attention constraints and dynamic information environments, this paper characterizes the context-dependency and evolutionary traits of the economic consequences of data asset disclosure. It extends the limited attention theory to the perspective of the "dilution effect of overall market disclosure density on information value" and provides new empirical evidence for understanding the economic consequences of information disclosure.

We believe that the main contributions of this paper are primarily reflected in the following aspects:

(1) At the theoretical level, this paper expands the boundaries of the limited attention theory and (to our best knowledge) for the first time extends the research perspective to the dimension of "investor attention dispersion caused by homogenized information." Existing research focuses more on the problem of investors' attention allocation among heterogeneous information, emphasizing how attention constraints affect investors' reactions to single information events. This paper introduces the limited attention theory into the dynamic perspective of information environment evolution, pointing out that when a certain type of information shifts from scarcity to ubiquity in the market, homogenized information may disperse investors' attention, thereby weakening the marginal informational value generated by individual firm disclosures.

(2) At the theoretical level, this study contributes to the limited attention literature by

examining how homogenized information environments shape the dispersion of investor attention. By constructing a market-level data asset disclosure index and incorporating it into a framework of threshold regression and grouped regression, this paper captures the changes in the effectiveness of corporate disclosure under different information environments. This approach allows the study to distinguish the "economic consequences of the disclosure behavior itself" from the "marginal effect decay caused by changes in the information environment," providing a new identification path for explaining conflicting conclusions in existing literature and offering an extensible empirical paradigm for future research to analyze the context-dependency of other types of information disclosure.

(3) At the level of conclusions, this paper provides a series of challenging empirical evidence regarding the economic consequences of data asset information disclosure. It finds that data asset disclosure significantly enhances capital market pricing efficiency only during stages when the overall disclosure level is low and information is relatively scarce; as listed companies widely adopt such disclosures, its marginal pricing role weakens or even disappears. This result indicates that information disclosure is not a case of "the more, the better," and its market function has an "endogenous boundary." Consequently, this paper reveals the potential "value dilution" issue in information disclosure, providing new empirical evidence and insights for understanding the relationship between corporate disclosure and capital market pricing efficiency.

2 Literature Review and Research Hypothesis

2.1 Investors' Limited Attention

The limited attention theory posits that human cognitive resources and information-processing capacities are subject to inherent constraints, making it impossible to simultaneously attend to and process all available information (Egeth and Kahneman, 1975). In capital markets, this constraint forces investors to selectively focus on massive amounts of information, thereby influencing their decision-making behavior and market pricing efficiency (Odean, 1999).

Inspired by psychological evidence that "attention is a scarce cognitive resource," Peng and Xiong (2006) pioneered an attention allocation model for investor learning and examined its impact on asset price dynamics. They discovered that limited attention leads to "category-learning behavior," where investors tend to focus more on market- and industry-level information while neglecting firm-specific information. Hirshleifer et al. (2011) proposed a theoretical model of investor limited attention to explain why the market underreacts to aggregate earnings but overreacts to earnings components. Their results indicate that investors' neglect of earnings information leads to post-earnings-announcement drift, while neglect of earnings components results in accrual and cash flow anomalies. Lu (2022) presented a theoretical model of financial reporting disclosure, demonstrating that when investors' information-processing capacity is limited, providing more granular and detailed information may instead lead to information overload, making it difficult for them to effectively extract key signals and thus reducing decision quality. In contrast, although summarized information loses some potential detail, its concise structure and lower cognitive load allow for higher information extraction efficiency under conditions of low attention. Liu et al. (2023, 2014) constructed a theoretical model illustrating that IPO underpricing is a strategic tool

used by issuing companies to attract investors with limited attention—by promising to allocate underpriced shares to institutional investors participating in roadshows, they compensate for the opportunity cost of their attention. They found that this attention-driven underpricing is positively and asymmetrically correlated with media attention, being significant only in IPOs with upward price revisions, and is influenced by the proportion of shares retained and the extent of expansion.

A substantial body of empirical evidence supports the aforementioned theories. Hirshleifer et al. (2009) found that investors' reactions to earnings announcements weaken when the market is saturated with a large volume of competing information. Aboody et al. (2010) observed that stocks with superior past performance exhibit significant positive returns before earnings announcements, followed by a strong reversal; this occurs because stocks with substantial gains tend to attract investor attention and capital, particularly prior to their earnings releases. DellaVigna and Pollet (2009) compared investor reactions to financial reports released on Fridays with those on other weekdays, finding that reactions to Friday announcements are significantly weaker. Ferracuti and Lind (2025) discovered that when many firms release earnings announcements simultaneously, investors rationally shift their limited attention from firm-specific information to macroeconomic information, as clustered earnings announcements more effectively predict macroeconomic conditions. Kwan et al. (2026) found that institutional investors strategically redirect their attention from individual stock news to macro news during periods of high market volatility; funds (especially hedge funds) that effectively perform this attention reallocation achieve higher returns. Wang et al. (2019) explored how the management of listed companies strategically and selectively discloses information by exploiting investors' attention constraints. Based on the limited attention hypothesis, Jiang et al. (2023) constructed a corporate affiliation network using analyst co-coverage and identified significant and persistent momentum spillover effects among affiliated companies in the Chinese A-share market. Fang et al. (2018) found that the presence of competing information disperses investor attention away from analyst research reports. As the number of analyst reports released on the same day increases (i.e., higher information competition), the immediate investor response to analyst rating reports becomes weaker.

Although existing studies have examined the competition and allocation of investor attention among different signals under the coexistence of multi-source heterogeneous information, and revealed how competitive information shocks weaken investor reactions to specific corporate events or disclosures, they have not yet focused on whether homogenized information itself—when a large number of firms in the market simultaneously disclose highly similar content—weakens the marginal informational value of an individual firm's disclosure by dispersing investor attention.

2.2 Factors affecting the pricing efficiency of capital markets

(1) Information Disclosure and Capital Market Pricing Efficiency

Capital market pricing efficiency is significantly influenced by the quality and quantity of corporate information disclosure. According to the "information efficiency hypothesis," information asymmetry between firms and external parties is considered a critical cause of low resource allocation efficiency in capital markets (Morck et al., 2000; Yi et al., 2019). Existing research primarily explores the mechanisms through which information disclosure affects pricing efficiency from two dimensions: "disclosure level" and "disclosure characteristics" (Li et al., 2025). Regarding the disclosure level, a higher level of disclosure provides investors with more firm-specific information, thereby reducing their reliance on market- and industry-level macro

information. This allows firm-specific information to be integrated into stock prices through trading, thus enhancing pricing efficiency (Li et al., 2025). A wealth of empirical studies has confirmed this view from various perspectives, including corporate social responsibility disclosure (Dhaliwal et al., 2011; Wang et al., 2014), key audit matters (Goh et al., 2024; Wang and Li, 2019), and social media disclosures (Bartov et al., 2018; Blankespoor et al., 2014; Guo et al., 2025; Hu and Wang, 2015).

However, the key to whether the disclosure level affects market information efficiency lies in whether it increases the content of firm-specific information (Li et al., 2025). If a listed company's disclosure provides more market or industry information rather than idiosyncratic information, it may actually reduce capital market pricing efficiency (Wang et al., 2014). Therefore, improvements in disclosure quality—such as enhanced transparency of firm-specific information—can also improve pricing efficiency (Jin and Myers, 2006). With the development of big data mining and textual analysis technologies, an increasing number of studies have focused on the incremental information provided by the textual characteristics of disclosures (Brown and Tucker, 2011). Research indicates that capital market pricing efficiency tends to rise when textual similarity is low (Song et al., 2024; Zhao and Liu, 2024), readability is high (Bai et al., 2019), and the emotional tone is more authentic (Wu et al., 2020). Nevertheless, it is important to remain vigilant, as management may manipulate disclosures for private gain, exacerbating information asymmetry through untruthful positive reporting and thereby reducing pricing efficiency (Yu et al., 2013).

Existing literature suggests that corporate data asset disclosure, as a significant form of non-financial information, can significantly reduce stock price synchronicity and improve pricing efficiency through three channels: increasing information supply, enhancing information diffusion, and accelerating information integration (Li et al., 2025). At the same time, however, disclosures regarding data assets are becoming increasingly common. When the overall market disclosure level rises, the marginal value of an individual firm's disclosure may decline due to investors' limited attention constraints (Peng and Xiong, 2006), leading to the phenomenon of "everyone disclosing means no one disclosing."

(2) Information Environment and Capital Market Pricing Efficiency

Capital market pricing efficiency is also profoundly influenced by the external information environment (Song et al., 2024). As an emerging capital market, the Chinese securities market is characterized by a high proportion of individual investors (Chen et al., 2024; He et al., 2025; Ma and Yu, 2023). Due to the lack of reliable information sources and information mining capabilities, individual investors often exhibit significant irrational investment behaviors and need to rely on the assistance of other market information intermediaries, such as analysts and the news media (Huang and Guo, 2014). Existing literature presents divergent views on the role of analysts: some scholars argue that analyst coverage is an important channel for mining firm-specific information (Zhu et al., 2007), which can enhance capital market pricing efficiency by reducing information asymmetry (Piotroski and Roulstone, 2004). This effect is more pronounced when the information content of analyst reports is higher and the analysts' influence is greater (Yi et al., 2019). Conversely, other scholars contend that analysts act merely as intermediaries for information transmission (Hu and Wang, 2015) rather than improving pricing efficiency by mining firm-specific information; thus, analysts' behavior in transmitting market- or industry-level information may actually reduce market information efficiency (Huang and Guo, 2014). With the

improvement of communication infrastructure and technological progress, the role of news reporting in the information transmission process has become increasingly prominent. Investors can obtain more corporate information through channels such as financial newspapers, television, and the internet, and use this information for trading, thereby improving capital market pricing efficiency (Huang and Guo, 2014). Compared to individual investors, institutional investors possess superior capabilities in information searching, interpretation, processing, and large-scale trading, and can promote the integration of information into stock prices through trading mechanisms (Hou and Ye, 2008). Furthermore, institutional investors can exercise a monitoring role, enhancing pricing efficiency by improving information transparency (Wang et al., 2009). [④]

Existing literature has relatively comprehensively validated the impact of individual firm information disclosure on capital market pricing efficiency, yet it overlooks the contextual factor of overall market information disclosure density. In an environment with high overall information density, will the effectiveness of an individual firm's disclosure be diluted?

2.3 Economic Consequences of Data Asset Disclosure

Existing research demonstrates that data asset information disclosure exerts multi-dimensional positive impacts on corporate value creation and capital market efficiency. At the micro-firm level, data asset disclosure significantly alleviates financing constraints. He et al. (2024) found that data assetization effectively mitigates financing difficulties for "Specialized, Refined, Differential, and Innovational" (SRDI) SMEs through signaling and collateral effects. Simultaneously, data asset disclosure caters to investor preferences; Niu and Yu (2024) discovered that such disclosure significantly reduces the cost of equity capital by mitigating information asymmetry and enhancing corporate transparency. At the capital market level, Li et al. (2025), based on data from A-share listed companies from 2007 to 2022, systematically demonstrated that data asset disclosure significantly reduces stock price synchronicity and enhances pricing efficiency through three mechanisms: increasing information supply, enhancing information diffusion, and accelerating information integration. Furthermore, Wei et al. (2022) showed that data asset disclosure can increase analyst coverage and reduce analyst forecast bias, thereby strengthening market information efficiency. Yuan et al. (2022) found that data asset disclosure enhances enterprise value, an effect more pronounced in firms with higher institutional shareholding.

However, the economic consequences of information disclosure are not static but highly dependent on the context in which the disclosure occurs. Yuan et al. (2025) indicated that institutional investors can directly and effectively monitor the behavior of listed companies' management; however, when their attention is dispersed, management is more likely to engage in tone manipulation in annual reports to strategically mislead investors. As the degree of data asset disclosure among listed companies has risen in recent years (Li et al., 2025; Zhang and Zhang, 2025), the effectiveness of such disclosure in reducing stock price synchronicity is likely to diminish. Using publicly available data by Li et al., we re-examined the consequences of data asset disclosure, found that when the regression is conducted using only recent data (e.g., 2020–2022), the impact of data asset disclosure on capital market pricing efficiency is attenuated, whereas it remains statistically significant when using the full sample (2007–2022). This

---

[④] However, the behavior of institutional investors is not always rational, and their herding behavior can also reduce capital market pricing efficiency (Xu et al., 2013).

time-varying characteristic suggests that the economic consequences of data asset disclosure may dynamically evolve with the overall market disclosure environment. Based on the limited attention theory, when the overall market disclosure level increases, investors' attention resources become dispersed, potentially leading to a decline in the marginal value of an individual firm's disclosure.

2.4 The Proposal of Hypotheses

Based on the aforementioned theories, we contend that the impact of data asset information disclosure on capital market pricing efficiency is not static but is likely moderated by the overall market information disclosure environment. This perspective stems from the inherent logic of the interaction between the limited attention theory and the information environment. When the overall level of data asset disclosure in the market is low, the data asset information disclosed by an individual firm possesses high information scarcity and signaling value. Investors can concentrate their limited attention resources on these disclosures, effectively identifying firm-specific information and incorporating it into stock prices through trading activities, thereby reducing stock price synchronicity and enhancing capital market pricing efficiency. However, with the in-depth development of the digital economy, an increasing number of enterprises have begun to disclose information related to data assets. When the overall market disclosure level is high, investors' attention resources are dispersed by a vast amount of similar information. At this point, a firm's idiosyncratic information may be drowned out by homogenized information within market noise. In such a scenario, individual data asset disclosures become more difficult for investors to notice, and their impact on stock price synchronicity consequently weakens. From an empirical standpoint, based on publicly available data by Li et al., we observed that when testing with the full sample period (2007–2022), the role of data asset disclosure in reducing stock price synchronicity is indeed significant. However, when using only the recent sample (e.g., 2020–2022) characterized by a high overall market disclosure level, this relationship loses its statistical significance. This phenomenon cannot be fully explained by existing theories but aligns closely with the intuition that " universal disclosure equivalent to no disclosure."

Based on these arguments, we propose the following hypotheses:

H1: When the overall market level of data asset information disclosure is low, an increase in the disclosure level of data assets by listed companies is conducive to reducing their stock price synchronicity and enhancing capital market pricing efficiency.

H2: When the overall market level of data asset information disclosure is high, the effect of increasing the disclosure level of data assets by listed companies on reducing stock price synchronicity and enhancing capital market pricing efficiency will be attenuated.

3 Empirical Design

3.1 Data and Sample

Due to the implementation of new accounting standards in 2007, which led to changes in the calculation of certain annual financial indicators, this paper selected all A-share listed companies from 2007 to 2022 as the initial sample to ensure the comparability and consistency of accounting data. The samples were screened and processed as follows: (1) Excluding financial and insurance companies; (2) Excluding companies that were subject to special treatment such as ST, *ST, PT, or

delisted during the listing period; (3) Excluding samples with an asset-liability ratio greater than one (i.e., insolvency); (4) Excluding samples with missing values. Furthermore, to mitigate the potential interference of extreme values on the estimation results, all continuous variables were winsorized at the 1% and 99% quantiles. After these screening steps, a final total of 30,600 firm-year observations was obtained. Among them, corporate data asset information disclosure data were extracted by processing the text of annual reports; news media-related data were sourced from the Chinese Research Data Services (CNRDS) platform; and the remaining financial and corporate governance variables were primarily obtained from the China Stock Market & Accounting Research (CSMAR) database. [5]

3.2 Econometric models setting

To examine the impact of data asset information disclosure on stock price synchronicity, we first construct the following basic model.

$$\boldsymbol{SYN_{i,t+1} = \alpha_0 + \alpha_1 DA_{i,t} + \gamma Controls_{i,t} + \sum Firm + \sum Year + \sum Area + \varepsilon_{i,t}} \quad (1)$$

Where i represents firm, t represents year, the dependent variable SYN represents stock price synchronicity, the core explanatory variable DA is the level of corporate data asset information disclosure, $Controls$ represents a set of control variables, and $\varepsilon_{i,t}$ represents the random error term. To enhance estimation reliability, the model further controls for firm, year, and province fixed effects ($\sum Firm$, $\sum Year$, and $\sum Area$), and clusters the standard errors at the firm level.

Next, referring to the work of Hansen (1999), we construct the following single-threshold panel regression model (2). We first conduct a threshold effect test based on Model (2). Upon confirming the existence of a significant threshold effect, we split the 16-year sample (2007–2022) into two groups. Each subsample is then regressed using Model (1). Additionally, following the methodology of Jiang (2022), we introduce a grouping dummy variable and its interaction term with the core explanatory variable to determine whether the regression coefficients of the two groups differ significantly.

$$\boldsymbol{SYN_{i,t+1} = \alpha_0 + \alpha_{11} DA_{i,t} \cdot I(\overline{DA_t} \le \lambda) + \alpha_{12} DA_{i,t} \cdot I(\overline{DA_t} > \lambda) + \gamma Controls_{i,t} + \sum Firm + \sum Year + \sum Area + \varepsilon_{i,t}} \quad (2)$$

Where the threshold variable $\overline{DA_t}$ represents the cross-sectional mean of the core explanatory variable—corporate data asset information disclosure level—in year t, which measures the overall market-level data asset disclosure intensity; λ denotes a specific threshold value; $I(\cdot)$ is an indicator function that equals 1 if the condition within the parentheses holds true, and 0 otherwise; the interpretations of other variables remain unchanged.

Based on the theoretical analysis and hypotheses above, we expect that the regression coefficient α_1 for the low group of overall market-level data asset information disclosure will be significantly negative, whereas the coefficient for the high group will be statistically insignificant.

---

[5] To ensure the comparability of the empirical results with existing studies, we directly utilized the data publicly provided by Li et al. on the website of China Industrial Economics. This paragraph is essentially a restatement of their data collection and processing procedures.

3.3 Definition of Variables

(1) Dependent Variable: Capital Market Pricing Efficiency

We employ stock price synchronicity as a proxy for capital market pricing efficiency. A higher value indicates that stock prices reflect more market- or industry-level common information rather than firm-specific information, implying lower pricing efficiency in the capital market.

Following the approach of Durnev et al. (2003), we first perform the following regression on weekly stock returns. This procedure essentially corresponds to the first stage of the Fama–MacBeth regression (Fama and MacBeth, 1973); however, our objective is not to estimate factor exposures but rather to obtain the coefficient of determination R^2.

$$\boldsymbol{R_{i,t} = \alpha_i + \beta_{i,1} R_{m,t} + \beta_{i,2} R_{I,t} + \varepsilon_{i,t}} \quad (3)$$

Where $R_{i,t}$ denotes the weekly return of stock $i$in week t; $R_{m,t}$ represents the weekly return of the market index in week t; $R_{I,t}$ denotes the weekly return of industry I in week t; and $\varepsilon_{i,t}$ is the random error term. All returns are computed based on backward-adjusted prices using the total market capitalization weighted average method. The returns for individual stocks, the market, and industries all incorporate cash dividend reinvestment. Industry classification follows the China Securities Regulatory Commission 2012 standard.

The R^2 from the above regression captures the extent to which stock i's return co-moves with market and industry returns—that is, the proportion of price variation explained by market- and industry-level information. To render the R^2 measure more normally distributed, we further apply the following logit transformation to obtain the final stock price synchronicity indicator SYN:

$$\boldsymbol{SYN = ln(\frac{R^2}{1 - R^2})} \quad (4)$$

(2) Explanatory Variable and Threshold Variable: Data Asset Information Disclosure Level and Its Annual Means

Following the "seed word + Word2Vec similar word expansion" method proposed by Yuan et al. (2022), this paper defines data assets based on the *White Paper on Data Asset Management Practices (Version 5.0)*. We select "data assets" and "data resources" as seed words, utilize the Word2Vec model to expand into a set of semantically similar keywords, and calculate the total frequency of the top ten keywords with the highest similarity within annual reports. Given that this frequency follows a right-skewed distribution, we construct two proxy variables: first, the natural logarithm of the total word frequency plus one, denoted as *LNDA*; second, the proportion of relevant keyword frequency relative to the total word count of the annual report, denoted as *PERDA*. Finally, the annual means of these two variables are calculated to serve as threshold variables.

(3) Control Variables

Drawing on the research of Wang and Li (2019), Song et al. (2024), and Zhao and Liu (2024), this study incorporates several control variables: firm size (*Size*), listing age (*Age*), nature of ownership (*SOE*), Big Four auditor (*Big4*), leverage (*LEV*), operating revenue growth rate (*Grow*), return on assets (*ROA*), market-to-book ratio (*MB*), stock turnover (*Turnover*), separation of ownership and control (*Separation*), audit opinion (*Opinion*), ownership concentration (*TOP1*), CEO duality (*Dual*), management shareholding ratio (*Manhold*), board size (*Boardsize*),

proportion of independent directors (*Indir*), and the regional marketization index (*Market*). To ensure the reliability of the regression results, this paper further controls for firm, year, and province fixed effects, with standard errors clustered at the firm level. Detailed variable definitions and calculation methods are presented in Table 1.

Table 1. Main variables' definition in this article

Variables' name	Variables' Symbol	Definition
Capital market pricing efficiency	SYN	Stock price synchronicity indicator for the next period, calculated based on market and industry returns using Equations (3) and (4).
Level of corporate data asset disclosure	LNDA	The first indicator of data asset disclosure, measured as the natural logarithm of (total frequency of keywords related to "data assets" + 1).
	PERDA	The second indicator of data asset disclosure, measured as the ratio of the total frequency of keywords related to "data assets" to the total word count in the annual report.
Firm size	Size	Natural logarithm of total assets.
Listing years	Age	ln(Current year of the sample - Listing year + 1).
Nature of ownership	SOE	Dummy variable: 1 if the firm is a state-owned enterprise, 0 otherwise.
Big 4 audit	Big4	Dummy variable: 1 if the financial report is audited by a "Big 4" accounting firm, 0 otherwise.
Leverage	LEV	Total liabilities / Total assets.
Operating revenue growth rate	Grow	(Main business revenue of current year - Main business revenue of previous year) / Main business revenue of previous year.
Return on assets	ROA	Net profit / Total assets.
Market-to-book ratio	MB	Market value of the firm / Book value.
Stock turnover rate	Turnover	Natural logarithm of the average daily turnover rate during the year, calculated using tradable shares.
Separation of two rights	Seperation	The difference between control rights and ownership rights.
Audit opinion	Opinion	Dummy variable: 1 if the audit opinion of the financial report is a standard unqualified opinion, 0 otherwise.
Ownership concentration	TOP1	Shareholding percentage of the largest shareholder.
CEO duality	Dual	Dummy variable: 1 if the chairman and general manager (CEO) are the same person, 0 otherwise.
Management shareholding ratio	Manhold	Ratio of shares held by directors, supervisors, and senior executives to total shares.
Board size	Boardsize	Total number of directors on the board.

Proportion of independent directors	Indir	Number of independent directors / Total number of board members.
Marketization index	Market	The "Marketization Index Score" of the province where the company is registered.

3.4 Descriptive Statistics of Variables

The descriptive statistics for the main variables of this study are presented in Table 2. The results show that the mean of the dependent variable, *SYN*, is approximately -0.4473, with a standard deviation of 1.0517, and minimum and maximum values of -3.9233 and 1.8669, respectively. This indicates significant variation in stock price synchronicity across different firms. Regarding the primary explanatory variables, the first proxy, *LNDA*, has a mean of 0.5274 and a median of 0, with minimum and maximum values of 0 and 3.9703. These figures suggest that, across the entire sample, the level of data asset disclosure among listed companies remains relatively low. The second proxy, *PERDA*, has a mean of approximately 0.003, indicating that data asset-related information accounts for an average of only 0.3% of the total word count in annual reports. This further confirms the overall insufficiency of data asset disclosure among listed firms. Nevertheless, as previously illustrated in Figure 1 and Figure 2, there is a clear upward trend in the extent of data asset information disclosure over time.

Table 2. Descriptive statistics of the main variables in this paper

Variables	Obs	Mean	SD	Min	P25	Median	P75	Max
SYN	30600	-0.4473	1.0517	-3.9233	-1.0478	-0.3499	0.2739	1.8669
LNDA	30600	0.5274	0.9038	0.0000	0.0000	0.0000	0.6931	3.9703
PERDA	30600	0.0031	0.0065	0.0000	0.0000	0.0000	0.0028	0.0253
Size	30600	22.1703	1.2623	19.8881	21.2648	21.9854	22.8961	26.0988
Age	30600	2.1406	0.7705	0.6931	1.6094	2.3026	2.7726	3.3322
SOE	30600	0.3837	0.4863	0.0000	0.0000	0.0000	1.0000	1.0000
Big4	30600	0.0563	0.2306	0.0000	0.0000	0.0000	0.0000	1.0000
LEV	30600	0.4281	0.2008	0.0586	0.2684	0.4223	0.5789	0.8821
Grow	30600	0.3573	1.0143	-0.7108	-0.0394	0.1190	0.3850	7.3003
ROA	30600	0.0535	0.0633	-0.2033	0.0274	0.0514	0.0830	0.2413
MB	30600	3.9626	2.7768	1.2707	2.3187	3.1898	4.5681	19.2529
Turnover	30600	6.1864	0.7879	4.1038	5.6631	6.2248	6.7600	7.8472
Seperation	30600	4.7701	7.3862	0.0000	0.0000	0.0000	7.6639	28.1632
Opition	30600	0.9765	0.1516	0.0000	1.0000	1.0000	1.0000	1.0000
TOP1	30600	34.7191	14.6280	9.1818	23.2367	32.5915	44.6389	74.2950
Dual	30600	0.2735	0.4457	0.0000	0.0000	0.0000	1.0000	1.0000
Manhold	30600	12.7933	19.1266	0.0000	0.0006	0.3540	23.4489	67.5000
Boardsize	30600	8.5916	1.6846	5.0000	7.0000	9.0000	9.0000	15.0000
Indir	30600	37.3984	5.2352	33.3300	33.3300	33.3300	42.8600	57.1400
Market	30600	9.6106	1.7950	4.1410	8.6680	9.8130	10.8270	12.8640

4 Empirical results

4.1 Basic regression results

We first conducted threshold tests on the two threshold variables. As shown in Table 3, the threshold value for variable $\overline{LNDA}_t$ is 0.8402, and that for $\overline{PERDA}_t$ is 0.0046; both are statistically significant at the 1% level. Based on these threshold values, the 16-year sample spanning 2007–2022 was divided into two groups according to the market-level data asset information disclosure intensity, with the grouping results also presented in Table 3. When $\overline{LNDA}_t$ is used as the threshold variable, the years 2020–2022 are classified as the "high market-level disclosure" group, while the remaining years constitute the "low market-level disclosure" group. When $\overline{PERDA}_t$ serves as the threshold variable, 2020 and 2022 are categorized as the high-disclosure group, with all other years forming the low-disclosure group.

Subsequently, regression analyses were performed separately on the two subsamples, with results reported in Table 4. The findings reveal that, regardless of whether LNDA or PERDA is employed as the explanatory variable, only the regression coefficients for the low market-level disclosure group are significantly negative, whereas those for the high market-level disclosure group lack statistical significance. Moreover, the coefficient differences between the two groups are statistically significant. These results indicate that in the early period—when data asset disclosure had not yet become widespread among listed firms—enhancing the extent of data asset disclosure indeed contributed to reducing stock price synchronicity. However, in later periods when data asset disclosure had evolved into a prevalent practice across the market, no evidence suggests that further increasing disclosure intensity continues to lower stock price synchronicity.

Table 3. Threshold test results and grouping results based on threshold values

Threshold Variable	Threshold Value	F statistic (P-value in parentheses)	Results Grouped by Year			
			$\overline{LNDA}_t \leq 0.8402$	$\overline{LNDA}_t > 0.8402$	$\overline{PERDA}_t \leq 0.0046$	$\overline{PERDA}_t > 0.0046$
$\overline{LNDA}_t$	0.8402	77.57*** (0.0050)	2007-2019	2020-2022		
$\overline{PERDA}_t$	0.0046	69.48*** (0.0025)			2007-2019、2021	2020、2022

Note: Unless otherwise stated, this table and other tables follow the following rules: ***, ** and * represent significance at the 1%, 5% and 10% levels, respectively.

Table 4. Grouped regression results based on threshold test

Panel A. Using $LNDA$ as the independent variable		
	$\overline{LNDA}_t \leq 0.8402$	$\overline{LNDA}_t > 0.8402$
	(1)	(2)
$LNDA$	-0.021*	-0.022
	(-1.72)	(-0.81)
Control variables	Yes	Yes
Three fixed effects	Yes	Yes
P-value of between-group	0.002	

coefficient difference test		
Obs.	21,017	8,669
Adj. R^2	0.511	0.686
Panel B. Using $PERDA$ as the independent variable		
	$\overline{PERDA}_t \leq 0.0046$	$\overline{PERDA}_t > 0.0046$
	(1)	(2)
$PERDA$	-2.723*	-0.706
	(-1.74)	(-0.13)
Control variables	Yes	Yes
Three fixed effects	Yes	Yes
P-value of between-group coefficient difference test	0.001	
Obs.	23,669	4,938
Adj. R^2	0.499	0.737

Note: In Panel A, column (1) presents the regression results for the subsample with low market-level data asset information disclosure intensity using LNDA as the explanatory variable, column (2) presents results for the high market-level disclosure subsample using LNDA. The P-value for the between-group coefficient difference test indicates whether the coefficients in columns (1) and (2) differ significantly. In Panel B, column (1) presents regression results for the low market-level disclosure subsample using PERDA as the explanatory variable, and column (2) presents regression results for the high market-level disclosure subsample using PERDA as the explanatory variable. The P-value for the between-group coefficient difference test indicates whether the coefficients in columns (1) and (2) differ significantly. Due to the inclusion of firm, year, and province fixed effects, the sample size may be slightly reduced. Unless otherwise specified, this table and other regression result tables follow the convention that ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively; t-statistics are reported in parentheses, and standard errors are clustered at the firm level.

4.2 Robustness Check

(1) Bidirectional Clustering Adjustment

To address potential cross-sectional correlation at the firm level and autocorrelation in the time series inherent in panel data, this study implements bidirectional clustering adjustments at both the firm and year levels based on the baseline regression. The results are presented in Table 5. As shown, the core findings remain robust after bidirectional clustering adjustment. Specifically, in the subsample with low market-average disclosure levels, the regression coefficients of both core explanatory variables are significantly negative; whereas in the high market disclosure level group, the regression coefficients lack statistical significance. Meanwhile, the coefficient difference between the two groups remains statistically significant.

Table 5. Bidirectional clustering adjusted regression results

Panel A. Using $LNDA$ as the independent variable		
	$\overline{LNDA}_t \leq 0.8402$	$\overline{LNDA}_t > 0.8402$
	(1)	(2)
$LNDA$	-0.0213*	-0.022
	(-2.09)	(-1.02)

Control variables	Yes	Yes
Three fixed effects	Yes	Yes
P-value of between-group coefficient difference test	0.041	
Obs.	21017	8669
Adj. R^2	0.4251	0.4984
Panel B. Using $PERDA$ as the independent variable		
	$\overline{PERDA}_t \leq 0.0046$	$\overline{PERDA}_t > 0.0046$
	(1)	(2)
$PERDA$	-2.723* (-2.44)	-0.706 (-0.13)
Control variables	Yes	Yes
Three fixed effects	Yes	Yes
P-value of between-group coefficient difference test	0.018	
Obs.	23,669	4,938
Adj. R^2	0.4189	0.4633

(2) Excluding non-high-tech enterprises

Compared to ordinary firms, high-tech enterprises possess richer data assets and are more capable of generating information that can be disclosed in annual reports. Accordingly, this study conducts additional robustness checks focusing exclusively on high-tech firm samples. Following the classification approach of Peng and Mao (Peng and Mao, 2017), we define high-tech firms as those belonging to 15 major categories within manufacturing, 3 major categories within information transmission, software, and information technology services, and the scientific research and technical services sector. Separate regressions were performed on this subsample, with results presented in Table 6. The findings indicate that during periods of low market-average disclosure levels, the coefficients of both LNDA and PERDA remain significantly negative; however, as the overall market disclosure level rises, this statistical significance disappears. This result suggests that even among high-tech enterprises—where data assets are presumably more substantive and disclosure more prevalent—the effectiveness of data asset information disclosure in enhancing capital market pricing efficiency is still constrained by the industry's average disclosure intensity.

Table 6. Regression results without non-high-tech enterprises

Panel A. Using $LNDA$ as the independent variable		
	$\overline{LNDA}_t \leq 0.8402$	$\overline{LNDA}_t > 0.8402$
	(1)	(2)
$LNDA$	-0.025** (-2.37)	-0.012 (-0.92)
Control variables	Yes	Yes
Three fixed effects	Yes	Yes
P-value of between-group coefficient difference test	0.002	

Obs.	26674	18071
Adj. R^2	0.4884	0.4642
Panel B. Using $PERDA$ as the independent variable		
	$\overline{PERDA}_t \leq 0.0046$	$\overline{PERDA}_t > 0.0046$
	(1)	(2)
$PERDA$	-3.559** (-2.40)	-1.558 (-0.86)
Control variables	Yes	Yes
Three fixed effects	Yes	Yes
P-value of between-group coefficient difference test	0.001	
Obs.	26674	18071
Adj. R^2	0.4884	0.4642

4.3 Endogeneity Treatment

(1) Accounting for Time Trend Variations

To eliminate potential confounding effects arising from inherent regional or industry-specific trends—stemming from asynchronous development paces in the digital economy—this study augments the baseline model with interactive fixed effects for province-by-year (Area×Year) and industry-by-year (Ind×Year), followed by re-estimation. The regression results demonstrate that the core findings remain robust after controlling for omitted variables that may evolve over time and across spatial dimensions.

Table 7. Regression results after accounting for time Trend variations

Panel A. Using $LNDA$ as the independent variable				
	$\overline{LNDA}_t \leq 0.8402$		$\overline{LNDA}_t > 0.8402$	
	(1)	(2)	(3)	(4)
$LNDA$	-0.025** (-2.00)	-0.043*** (-3.21)	-0.013 (-0.48)	-0.019 (-0.74)
Control variables	Yes	Yes	Yes	Yes
Three fixed effects	Yes	Yes	Yes	Yes
Provinces×years fixed effect	Yes	No	Yes	No
Industries×years fixed effect	No	Yes	No	Yes
P-value of between-group coefficient difference test	0.002	0.657	-	-
Obs.	21017	21017	8669	8669
Adj. R^2	0.428	0.441	0.503	0.529
Panel B. Using $PERDA$ as the independent variable				
	$\overline{PERDA}_t \leq 0.0046$		$\overline{PERDA}_t > 0.0046$	
	(1)	(2)	(3)	(4)
$PERDA$	-3.150*	-5.280***	-2.476	-3.471

	(-1.88)	(-2.93)	(-0.61)	(-0.89)
Control variables	Yes	Yes	Yes	Yes
Three fixed effects	Yes	Yes	Yes	Yes
Provinces×years fixed effect	Yes	No	Yes	No
Industries×years fixed effect	No	Yes	No	Yes
P-value of between-group coefficient difference test	0.001	0.076	-	-
Obs.	21017	21017	8669	8669
Adj. R^2	0.428	0.441	0.503	0.529

(2) Entropy Balancing Test

To alleviate the selection bias arising from systematic differences in observable characteristics between the treatment and control groups, this study employs the entropy balancing method to reweight the sample, ensuring that the two groups are balanced across all control variables. After eliminating the differences in observable covariates using entropy balancing weights, the results of the subsample regressions remain consistent with the baseline findings. Specifically, after constructing a more comparable analytical sample, the dynamic pattern—where the average disclosure level is "effective in the low-level group but ineffective in the high-level group"—remains robust.

Table 8. Regression results with entropy balancing test

Panel A. Using $LNDA$ as the independent variable		
	$\overline{LNDA}_t \leq 0.8402$	$\overline{LNDA}_t > 0.8402$
	(1)	(2)
$LNDA$	-0.025*	-0.027
	(-1.89)	(-0.99)
Control variables	Yes	Yes
Three fixed effects	Yes	Yes
P-value of between-group coefficient difference test	<0.001	
Obs.	21017	8669
Adj. R^2	0.442	0.512
Panel B. Using $PERDA$ as the independent variable		
	$\overline{PERDA}_t \leq 0.0046$	$\overline{PERDA}_t > 0.0046$
	(1)	(2)
$PERDA$	-3.027*	-3.773
	(-1.73)	(-0.91)
Control variables	Yes	Yes
Three fixed effects	Yes	Yes
P-value of between-group coefficient difference test	<0.001	

Obs.	21017	8669
Adj. R^2	0.442	0.512

5 Mechanism Analysis

According to Li et al. (2025), the disclosure of information related to data assets helps external market participants more clearly grasp the development dynamics of a firm's data resources. Due to its idiosyncratic characteristics, such information can effectively capture investor attention, thereby expanding the supply of firm-specific information. Meanwhile, as a critical indicator of a firm’s digital technological prowess and future growth potential, data asset information is highly likely to attract active attention and coverage from information intermediaries such as the news media. This, in turn, generates a significant information diffusion effect and facilitates the flow of idiosyncratic information within the market. Furthermore, leveraging their extensive information channels and professional analytical capabilities, institutional investors can delve deep into such information and accelerate its integration into stock prices. Consequently, data asset disclosure by listed companies can reduce stock price synchronicity and enhance capital market pricing efficiency through three mechanisms: increasing information supply, strengthening information diffusion, and accelerating information integration. However, the regression results mentioned above indicate that the effect of data asset disclosure on reducing stock price synchronicity is moderated by the overall market-wide disclosure level. Does this impact similarly manifest across these three specific mechanisms? This section will conduct a re-examination of these three mechanisms.

5.1 Improving Information Supply

Existing research suggests that data asset disclosure can enhance capital market pricing efficiency by increasing the supply of firm-specific information (Li et al., 2025; Niu and Yu, 2024; Wei et al., 2022). Given that traditional financial statements often fail to fully reflect the value of data assets, data asset disclosures within the textual portions of annual reports provide investors with crucial incremental information (Liu et al., 2025). This allows for a more comprehensive understanding of a firm’s digital technology applications, data governance capabilities, and future growth potential, thereby mitigating information asymmetry. Simultaneously, the process of developing and disclosing data assets helps optimize internal information flows, improve accounting information quality, and reduce information screening costs for investors. As vital information intermediaries, analysts can provide in-depth interpretations of disclosed data asset information and transmit this to the market through research reports, reducing their reliance on macro and industry-level information while improving the specificity and accuracy of their forecasts. Consequently, data asset disclosure can significantly increase analyst coverage and reduce forecast bias, thereby enhancing the content of firm-specific information embedded in stock prices. However, analysts are likewise constrained by limited attention. When the overall market-wide level of data asset disclosure continues to rise and the content becomes increasingly homogenized, analysts' cognitive resources may be dispersed by a vast amount of similar information, making it difficult to conduct deep dives into individual firms. In such scenarios, the marginal effect of an individual firm increasing its data asset disclosure to boost analyst coverage

and improve forecast quality may diminish.

Therefore, we re-examine the mechanism through which data asset disclosure reduces stock price synchronicity by improving information supply. Specifically, following the methodology of Wei et al. (2022), we measure analyst coverage (*Follow*) using the number of analyst teams tracking a firm within a year, and measure analyst forecast bias (*Bias*) using the ratio of the standard deviation of all analysts' most recent earnings-per-share forecasts to the stock price at the beginning of the period. These two indicators are used to gauge the level and quality of information disclosure. Based on the grouping results from the previous threshold test, regressions are conducted for both groups, with results presented in Table 9. The results indicate that when the overall market-wide level of data asset disclosure is low, an increase in a firm's data asset disclosure indeed helps to enhance analyst coverage and reduce forecast bias. However, when the overall market-wide disclosure level is high, there is no evidence to suggest that increasing data asset disclosure still yields these benefits.

Table 9. Test results of improving the information supply mechanism

Panel A. Using $LNDA$ as the independent variable				
	$\overline{LNDA}_t \leq 0.8402$		$\overline{LNDA}_t > 0.8402$	
	(1)	(2)	(3)	(4)
	$Follow$	$Bias$	$Follow$	$Bias$
$LNDA$	0.033**	-0.0005***	-0.011	0.0003
	(2.31)	(-3.12)	(-0.64)	(0.58)
Control variables	Yes	Yes	Yes	Yes
Three fixed effects	Yes	Yes	Yes	Yes
P-value of between-group coefficient difference test	0.010	0.655	-	-
Obs.	21,017	12,555	8,669	3,722
Adj. R^2	0.748	0.383	0.900	0.635
Panel B. Using $PERDA$ as the independent variable				
	$\overline{PERDA}_t \leq 0.0046$		$\overline{PERDA}_t > 0.0046$	
	(1)	(2)	(3)	(4)
	$Follow$	$Bias$	$Follow$	$Bias$
$PERDA$	4.177**	-0.059***	0.671	0.143
	(2.32)	(-3.23)	(0.19)	(1.41)
Control variables	Yes	Yes	Yes	Yes
Three fixed effects	Yes	Yes	Yes	Yes
P-value of between-group coefficient difference test	0.107	0.946	-	-
Obs.	23,669	13,715	4,938	1,884
Adj. R^2	0.698	0.235	0.780	0.360

Note: In Panel A, Column (1) presents the regression results for the group with a low market-wide level of data asset disclosure, using Follow as the dependent variable and LNDA as the explanatory variable; Column (2) presents the results for the low-level group using Bias as the dependent variable and LNDA as the explanatory variable; Column (3) presents the results for the high-level group using Follow as the dependent variable and LNDA as the explanatory variable; and Column (4) presents the results for the high-level group using Bias as the

dependent variable and LNDA as the explanatory variable. In Column (1), the p-value of the test for coefficient differences between groups indicates whether there is a significant difference between the coefficients of Columns (1) and (3); in Column (2), the p-value indicates whether there is a significant difference between the coefficients of Columns (2) and (4). In Panel B, Column (1) presents the regression results for the group with a low market-wide level of data asset disclosure, using Follow as the dependent variable and PERDA as the explanatory variable; Column (2) presents the results for the low-level group using Bias as the dependent variable and PERDA as the explanatory variable; Column (3) presents the results for the high-level group using Follow as the dependent variable and PERDA as the explanatory variable; and Column (4) presents the results for the high-level group using Bias as the dependent variable and PERDA as the explanatory variable. In Column (1), the p-value of the test for coefficient differences between groups indicates whether there is a significant difference between the coefficients of Columns (1) and (3); in Column (2), the p-value indicates whether there is a significant difference between the coefficients of Columns (2) and (4).

5.2 Strengthening Information Diffusion

Data asset disclosure may also reduce stock price synchronicity by strengthening information diffusion. Existing research suggests that, compared to traditional financial information, data asset information possesses greater novelty and topicality, making it more likely to capture the attention of analysts and the media. Such information is often viewed by the market as a critical signal of a firm's digital capabilities and future growth potential, and is thus more likely to be widely disseminated in capital markets through channels such as research reports and news coverage (Li et al., 2025). During this process, firm-specific information originally confined to annual reports diffuses to a broader group of investors, accelerating its integration into asset prices. This mechanism is particularly crucial in the Chinese capital market, which is dominated by individual investors. Individual investors rely more heavily on media reports and secondary information for decision-making; thus, the media's reprocessing of data asset information helps lower the cost of information comprehension. This enables more investors to identify differences in firms' data resource allocation and business models, thereby reducing stock price reliance on market and industry-level information and increasing the weight of firm-specific information in prices. However, the effectiveness of the information diffusion mechanism is also conditioned by the overall information environment. When the market-wide level of data asset disclosure is low, such information possesses high scarcity and distinctiveness, making it easier to attract sustained attention and achieve effective diffusion. As the number of disclosing firms increases and information content tends toward homogeneity, the marginal attention from the media and investors may decline, and firm-specific disclosures are more likely to be perceived as general industry characteristics rather than idiosyncratic traits. Under the constraint of limited attention, the improvement effect of information diffusion on stock price synchronicity may subsequently diminish.

Therefore, we re-examine the mechanism through which data asset disclosure reduces stock price synchronicity by strengthening information diffusion. Specifically, we use the number of online financial news reports (*News_cont1*) and the total number of financial newspaper and online news reports (*News_cont2*) from the Chinese Research Data Services (CNRDS) platform to measure the media's impact on the diffusion of firms' data asset information. Based on the grouping results from the previous threshold test, regressions are conducted for both groups, and the results are presented in Table 10. The results indicate that when the market-wide level of data

asset disclosure is low, increasing the level of data asset disclosure indeed helps attract news media attention. However, when the market-wide disclosure level is high, there is no evidence to suggest that increasing data asset disclosure still attracts news media attention.

Table 10. Test results of strengthening information diffusion mechanism

Panel A. Using $LNDA$ as the independent variable				
	$\overline{LNDA}_t \leq 0.8402$		$\overline{LNDA}_t > 0.8402$	
	(1)	(2)	(3)	(4)
	$News\_cont1$	$News\_cont2$	$News\_cont1$	$News\_cont2$
$LNDA$	11.953***	15.315**	3.019	4.019
	(2.60)	(2.56)	(0.79)	(0.83)
Control variables	Yes	Yes	Yes	Yes
Three fixed effects	Yes	Yes	Yes	Yes
P-value of between-group coefficient difference test	0.001	0.003	-	-
Obs.	20,987	8,663	20,825	8,006
Adj. R^2	0.779	0.932	0.816	0.950
Panel B. Using $PERDA$ as the independent variable				
	$\overline{PERDA}_t \leq 0.0046$		$\overline{PERDA}_t > 0.0046$	
	(1)	(2)	(3)	(4)
	$News\_cont1$	$News\_cont2$	$News\_cont1$	$News\_cont2$
$PERDA$	1,404.331**	1,680.394**	700.075	832.296
	(2.41)	(2.20)	(1.05)	(0.95)
Control variables	Yes	Yes	Yes	Yes
Three fixed effects	Yes	Yes	Yes	Yes
P-value of between-group coefficient difference test	0.013	0.028	-	-
Obs.	23,635	4,934	23,299	4,416
Adj. R^2	0.773	0.948	0.807	0.959

Note: In Panel A, Column (1) presents the regression results for the group with a low market-wide level of data asset disclosure, using News_cont1 as the dependent variable and LNDA as the explanatory variable; Column (2) presents the results for the low-level group using News_cont2 as the dependent variable and LNDA as the explanatory variable; Column (3) presents the results for the high-level group using News_cont1 as the dependent variable and LNDA as the explanatory variable; and Column (4) presents the results for the high-level group using News_cont2 as the dependent variable and LNDA as the explanatory variable. In Column (1), the p-value of the test for coefficient differences between groups indicates whether there is a significant difference between the coefficients of Columns (1) and (3); in Column (2), the p-value indicates whether there is a significant difference between the coefficients of Columns (2) and (4). In Panel B, Column (1) presents the regression results for the group with a low market-wide level of data asset disclosure, using News_cont1 as the dependent variable and PERDA as the explanatory variable; Column (2) presents the results for the low-level group using News_cont2 as the dependent variable and PERDA as the explanatory variable; Column (3) presents the results for the high-level group using News_cont1 as the dependent variable and PERDA as the explanatory variable; and Column (4) presents the results for the high-level group using News_cont2 as the dependent variable and PERDA as the explanatory variable. In Column (1), the p-value of the test for coefficient differences between groups indicates

whether there is a significant difference between the coefficients of Columns (1) and (3); in Column (2), the p-value indicates whether there is a significant difference between the coefficients of Columns (2) and (4).

5.3 Accelerating Information Integration

In addition to improving information supply and strengthening information diffusion, data asset disclosure may also enhance capital market pricing efficiency by accelerating the process of information integration. Information integration emphasizes the speed and sufficiency with which firm-specific information is reflected in stock prices following its disclosure. When disclosures are more comprehensive and clearly structured, investors can more quickly identify the implications of the information and adjust their trading behavior accordingly, thereby shortening the time it takes for information to be incorporated into prices (Li et al., 2025). Compared to information implicitly embedded in operating results, data asset information reveals characteristics of a firm's digital investment, data governance capabilities, and future strategic direction in a more direct and forward-looking manner. This helps investors form expectations in advance and reflect them in trades (McCAHERY et al., 2016). From a market operation perspective, systematic data asset disclosure helps reduce uncertainty and the proportion of noise trading, ensuring that investor decisions are based more on fundamentals rather than market resonance, thus improving price discovery efficiency and reducing stock price synchronicity. However, this mechanism is also constrained by the information environment. When the market-wide level of data asset disclosure is low, the relevant information possesses strong incremental characteristics, making it easier for investors to view it as valuable new information and rapidly incorporate it into pricing. As disclosure behavior becomes widespread and information content tends toward homogeneity, its marginal information content may decline, and investors' reactions to such disclosures tend to become more rational, thereby weakening the enhancement effect on the speed of information integration.

Therefore, we re-examine the mechanism through which data asset disclosure reduces stock price synchronicity by accelerating information integration. Specifically, we use the institutional investor shareholding ratio (*Inshold*) as a proxy for the effect of institutional investors accelerating information integration. Based on the grouping results from the previous threshold test, regressions are conducted for both groups, and the results are presented in Table 11. The results indicate that when the market-wide level of data asset disclosure is low, increasing a firm's data asset disclosure indeed helps enhance attention from institutional investors and encourages them to increase their shareholding. However, when the market-wide disclosure level is high, there is no evidence to suggest that increasing data asset disclosure still attracts institutional investor attention or prompts them to increase their holdings.

Table 11. Test results of accelerating information integration

Panel A. Using $LNDA$ as the independent variable		
	$\overline{LNDA}_t \leq 0.8402$	$\overline{LNDA}_t > 0.8402$
	(1)	(2)
	$Inshold$	$Inshold$
$LNDA$	1.532	3.660
	(1.00)	(1.58)

Control variables	Yes	Yes
Three fixed effects	Yes	Yes
P-value of between-group coefficient difference test	0.321	
Obs.	21,017	8,669
Adj. R^2	0.547	0.789
Panel B. Using $PERDA$ as the independent variable		
	$\overline{PERDA}_t \leq 0.0046$	$\overline{PERDA}_t > 0.0046$
	(1)	(2)
	$Inshold$	$Inshold$
$PERDA$	309.034	728.750
	(1.64)	(1.31)
Control variables	Yes	Yes
Three fixed effects	Yes	Yes
P-value of between-group coefficient difference test	0.893	
Obs.	23,669	4,938
Adj. R^2	0.519	0.711

Note: In Panel A, Column (1) presents the regression results for the group with a low market-wide level of data asset disclosure, using Inshold as the dependent variable and LNDA as the explanatory variable; Column (2) presents the results for the high-level group using Inshold as the dependent variable and LNDA as the explanatory variable. The p-value of the test for coefficient differences between groups indicates whether there is a significant difference between the coefficients of Columns (1) and (2). In Panel B, Column (1) presents the regression results for the group with a low market-wide level of data asset disclosure, using Inshold as the dependent variable and PERDA as the explanatory variable; Column (2) presents the results for the high-level group using Inshold as the dependent variable and PERDA as the explanatory variable. The p-value of the test for coefficient differences between groups indicates whether there is a significant difference between the coefficients of Columns (1) and (2).

6 Heterogeneity Analysis

The preceding regression and mechanism test results indicate that the impact of corporate data asset disclosure on capital market pricing efficiency is not universally consistent across all contexts; rather, it is contingent upon the market-wide level of data asset disclosure. When the overall market disclosure level is low, data asset information disclosed by an individual firm can indeed reduce stock price synchronicity—thereby enhancing pricing efficiency—through channels such as improving information supply, strengthening information diffusion, and accelerating information integration. Conversely, when the market-wide disclosure level is high, the marginal information value of an individual firm's disclosure diminishes significantly, weakening its impact on capital market pricing efficiency. However, notable differences exist across firms regarding their information environments, information production methods, and the characteristics of information users. Even within the same stage of overall market disclosure, the economic consequences of data asset disclosure may exhibit significant heterogeneity. Therefore, building

upon the aforementioned "high vs. low market-wide disclosure" groupings, this paper conducts a heterogeneity analysis based on the three functional paths of information supply, diffusion, and integration. Specifically, we select textual information similarity, regional informatization level, and institutional investor distraction as moderating variables.

The textual information similarity index measures the degree of variation in a firm's disclosure content across different years, facilitating an analysis of the differentiated impacts of data asset disclosure when the supply of firm-specific information varies. The degree of regional digital development determines the speed and breadth of information transmission, demonstrating the diffusion effect of data asset disclosure content across different geographical regions from a dissemination perspective. Finally, the institutional investor distraction index quantifies the concentration of professional investors' attention; it reflects whether data asset-related information can be fully captured by institutional investors during the information integration process and eventually be reflected in stock prices through market trading behavior. Through these three dimensions of heterogeneity analysis, we can effectively discern the differentiated performance of data asset disclosure's impact on pricing efficiency under various scenarios, further validating and enriching our previous conclusions.

6.1 Textual Information Similarity

The Management Discussion and Analysis (MD&A) section is one of the most critical and information-rich components of corporate annual reports (Wang et al., 2018). Its content, which includes management’s insights into significant events and outlook on future prospects, effectively meets the needs of information users for relevant, reliable, and forward-looking information (Clarkson et al., 1999). Textual similarity is a key characteristic of annual reports; specifically, the cosine similarity of MD&A texts between adjacent years can indirectly reflect the level of firm-specific information content in the current period (Brown and Tucker, 2011). Existing research suggests that data asset disclosure can enhance capital market pricing efficiency by supplementing firm-specific information and alleviating information asymmetry (Li et al., 2025). However, according to the preceding analysis, when firms universally disclose data assets, it becomes more difficult for an individual firm’s disclosure to capture investor attention. In this context, even if a firm’s MD&A text shows high similarity to the previous period, the function of data asset disclosure in supplementing idiosyncratic information may be weakened. Therefore, we expect that when the overall level of data asset disclosure is low and a firm’s MD&A similarity is high, data asset disclosure will play a significant role in helping investors mitigate information asymmetry and enhance pricing efficiency. Conversely, when the overall market-wide disclosure level is high, this effect will no longer be significant regardless of the degree of MD&A similarity.

To test the heterogeneous effect of MD&A textual similarity on data asset disclosure under different overall disclosure levels, this study follows the methodology of Wang et al. (2018). We use the textual similarity of the MD&A section in annual reports as a measure of firm-specific information disclosure, dividing the sample into high-similarity and low-similarity groups based on industry-year medians. Regression analyses are then conducted under different levels of overall market disclosure, with results presented in Table 12. The results show that the regression coefficients are statistically significant only when the overall data asset disclosure level is low and the firm’s MD&A similarity is high. This suggests that when the market-wide disclosure level is low, data asset disclosure indeed provides more idiosyncratic information, thereby more effectively reducing stock price synchronicity and improving pricing efficiency. However, as the

overall market disclosure level increases, this effect diminishes accordingly.

Table 12. Heterogeneity analysis results of textual information similarity

Panel A. Using $LNDA$ as the independent variable				
	$\overline{LNDA}_t \leq 0.8402$		$\overline{LNDA}_t > 0.8402$	
	(1)	(2)	(3)	(4)
	High	Low	High	Low
$LNDA$	-0.036*	-0.050	-0.046	-0.046
	(-1.84)	(-1.34)	(-1.31)	(-0.34)
Control variables	Yes	Yes	Yes	Yes
Three fixed effects	Yes	Yes	Yes	Yes
Obs.	8866	3320	5904	494
Adj. R^2	0.391	0.374	0.460	0.594
Panel B. Using $PERDA$ as the independent variable				
	$\overline{PERDA}_t \leq 0.0046$		$\overline{PERDA}_t > 0.0046$	
	(1)	(2)	(3)	(4)
	High	Low	High	Low
$PERDA$	-4.521*	-1.411	2.288	2.365
	(-1.81)	(-0.58)	(0.33)	(0.34)
Control variables	Yes	Yes	Yes	Yes
Three fixed effects	Yes	Yes	Yes	Yes
Obs.	10,078	9,963	3,047	3,185
Adj. R^2	0.568	0.558	0.741	0.730

6.2 Regional Informatization Level

The regional informatization level is a critical external environmental factor influencing information dissemination efficiency and investors' information acquisition capabilities. The ability of investors to promptly obtain and process stock market-related information significantly impacts their investment decisions. From an information channel perspective, the proliferation of information technologies, such as computers and the internet, not only lowers transaction costs but also substantially reduces barriers to obtaining firm-specific information, thereby enhancing market participation and decision quality (Goldstein et al., 2023; Guo and Liang, 2014). Existing research suggests that data asset disclosure can attract the attention of information intermediaries like the news media, diffusing idiosyncratic information to a broader group of investors and enhancing capital market pricing efficiency. The regional informatization level, as an indicator of information transmission infrastructure, directly determines whether information disseminated by the media can efficiently reach target investors (Li et al., 2025). However, according to the preceding analysis, when firms universally disclose data assets, investor attention becomes scattered. Even in regions with high informatization levels, the information disclosed by an individual firm may become more difficult to effectively identify and absorb. Therefore, we expect that when the overall level of data asset disclosure is low and the regional informatization level is high, data asset disclosure can significantly reduce stock price synchronicity through efficient information diffusion channels. Conversely, when the overall disclosure level is high, this effect

will no longer be significant regardless of the regional informatization level.

To test the heterogeneous effect of the regional informatization level on data asset disclosure under different overall disclosure levels, this study uses the number of computers per 100 people in the province where the firm is located as a proxy for the regional informatization level. The sample firms are divided into high and low informatization groups based on the province-year median, and regression analyses are conducted under different levels of overall market disclosure. [6] The results, presented in Table 13, show that the regression coefficients are statistically significant only when the overall data asset disclosure level is low and the regional informatization level is high. These findings suggest that when the market-wide disclosure level is low, data asset disclosure indeed significantly reduces stock price synchronicity and improves pricing efficiency through efficient information diffusion; however, as the overall market disclosure level increases, this effect diminishes.

Table 13. Heterogeneity analysis results of regional informatization level

Panel A. Using $LNDA$ as the independent variable				
	$\overline{LNDA}_t \leq 0.8402$		$\overline{LNDA}_t > 0.8402$	
	(1)	(2)	(3)	(4)
	High	Low	High	Low
$LNDA$	-0.043**	-0.007	-0.003	0.025
	(-2.27)	(-0.37)	(-0.05)	(0.53)
Control variables	Yes	Yes	Yes	Yes
Three fixed effects	Yes	Yes	Yes	Yes
Obs.	10,078	9,963	3,047	3,185
Adj. R^2	0.568	0.558	0.741	0.730
Panel B. Using $PERDA$ as the independent variable				
	$\overline{PERDA}_t \leq 0.0046$		$\overline{PERDA}_t > 0.0046$	
	(1)	(2)	(3)	(4)
	High	Low	High	Low
$PERDA$	-3.173*	-5.567	-1.182	-3.169
	(-1.74)	(-1.27)	(-0.15)	(-0.78)
Control variables	Yes	Yes	Yes	Yes
Three fixed effects	Yes	Yes	Yes	Yes
Obs.	18376	4077	1998	8669
Adj. R^2	0.432	0.369	0.467	0.499

6.3 Institutional Investor Distraction

As primary participants in capital markets, institutional investors play a pivotal role in the process of information integration. Compared to individual investors, institutional investors leverage their professional information mining capabilities and large-scale trading advantages to deeply identify firm-specific information and rapidly reflect it in stock prices, thereby enhancing

[6] When using PERDA as the explanatory variable, for years with a high market-wide level of data asset disclosure (2020 and 2022), the sample size for the group with a low regional informatization level was insufficient to yield regression results. Therefore, we adopted a "compromise method" by including the 2021 sample in the regression.

capital market pricing efficiency (Hou and Ye, 2008). Existing research suggests that data asset disclosure can accelerate the integration of idiosyncratic information into stock prices by attracting institutional investor attention (Li et al., 2025). However, as per the preceding analysis, institutional investors are also subject to cognitive resource constraints (Kwan et al., 2026). When the market-wide level of data asset disclosure is high, the limited attention of institutional investors becomes dispersed (Yuan et al., 2025). In such cases, even if their level of distraction is relatively low, it remains difficult for them to effectively identify the idiosyncratic data asset information of an individual firm. Therefore, we expect that when the overall disclosure level is low and institutional investor distraction is low, data asset disclosure will significantly reduce stock price synchronicity by garnering institutional attention. Conversely, when the overall disclosure level is high, this effect will diminish.

To test the heterogeneous impact of institutional investor distraction under different overall disclosure levels, this study follows the methodology of Yuan et al. (2025) to construct a firm-level indicator of institutional investor distraction. The sample is divided into high-distraction and low-distraction groups based on industry-year medians, and regression analyses are conducted under different levels of overall market disclosure. The results, shown in Table 14, indicate that the regression coefficients are statistically significant only when the market-wide data asset disclosure level is low and institutional investor distraction is low. These findings demonstrate that when the overall market disclosure level is low, data asset disclosure can indeed significantly reduce stock price synchronicity and improve pricing efficiency by attracting institutional investor attention; however, this effect weakens when the market-wide disclosure level is high.

Table 14. Heterogeneity analysis results of institutional investor distraction

Panel A. Using $LNDA$ as the independent variable				
	$\overline{LNDA}_t \leq 0.8402$		$\overline{LNDA}_t > 0.8402$	
	(1)	(2)	(3)	(4)
	High	Low	High	Low
$LNDA$	-0.004	-0.046**	-0.050	-0.026
	(-0.25)	(-2.24)	(-1.22)	(-0.53)
Control variables	Yes	Yes	Yes	Yes
Three fixed effects	Yes	Yes	Yes	Yes
Obs.	10,134	9,970	3,621	3,350
Adj. R^2	0.552	0.537	0.717	0.717
Panel B. Using $PERDA$ as the independent variable				
	$\overline{PERDA}_t \leq 0.0046$		$\overline{PERDA}_t > 0.0046$	
	(1)	(2)	(3)	(4)
	High	Low	High	Low
$PERDA$	-1.386	-5.322*	-6.153	-10.222
	(-0.60)	(-1.96)	(-1.04)	(-1.36)
Control variables	Yes	Yes	Yes	Yes
Three fixed effects	Yes	Yes	Yes	Yes
Obs.	10,134	9,970	3,621	3,350
Adj. R^2	0.552	0.537	0.717	0.717

7 Conclusion

Grounded in limited attention theory and the perspective of a dynamic information environment, this study utilizes data from Chinese A-share listed companies from 2007 to 2022 to re-examine and challenge prevailing assumptions regarding the economic consequences of data asset disclosure. The findings indicate that the enhancing effect of data asset disclosure on capital market pricing efficiency is not static; rather, it diminishes as the overall level of market disclosure rises. Specifically, during periods of low market-wide disclosure, data asset disclosures effectively reduce stock price synchronicity by improving the quality of information supply, strengthening media dissemination effects, and accelerating the integration of information by institutional investors. However, as data asset disclosure becomes prevalent among listed companies, homogenized information scatters the limited attention of investors, rendering these transmission mechanisms ineffective and resulting in a phenomenon where universal disclosure renders individual disclosures ineffective. Further heterogeneity analysis reveals that disclosure maintains its positive effects only under conditions where the overall market disclosure level is low, and when the firm discloses relatively little idiosyncratic information, the information dissemination environment is superior, or institutional investors' attention is not over-dispersed. These findings not only extend the theoretical boundaries of limited attention within the information environment and offer a reference for broader disclosure research, but also provide critical policy insights. When promoting the accounting recognition and disclosure of data resources, regulatory bodies should look beyond mere quantitative growth and focus on establishing differentiated, substantive disclosure standards to curb "hype-driven" homogenized reporting. Concurrently, firms should move away from boilerplate disclosure templates, focusing instead on identifying and presenting data asset information of unique value to regain investor attention through substantive differentiation in a crowded information market, thereby facilitating the effective transmission of data value and enhancing capital market efficiency.

References

Aboody, D., Lehavy, R., Trueman, B., 2010. Limited attention and the earnings announcement returns of past stock market winners. Rev Account Stud 15, 317–344. https://doi.org/10.1007/s11142-009-9104-9

Bai, X., Dong, Y., Hu, N., 2019. Financial report readability and stock return synchronicity. Applied Economics 51, 346–363. https://doi.org/10.1080/00036846.2018.1495824

Bartov, E., Faurel, L., Mohanram, P.S., 2018. Can Twitter Help Predict Firm-Level Earnings and Stock Returns? The Accounting Review 93, 25–57. https://doi.org/10.2308/accr-51865

Blankespoor, E., deHaan, E., Marinovic, I., 2020. Disclosure processing costs, investors' information choice, and equity market outcomes: A review. Journal of Accounting and Economics 70, 101344. https://doi.org/10.1016/j.jacceco.2020.101344

Blankespoor, E., Miller, G.S., White, H.D., 2014. The Role of Dissemination in Market Liquidity: Evidence from Firms' Use of Twitter™. The Accounting Review 89, 79–112. https://doi.org/10.2308/accr-50576

Bochkay, K., Brown, S.V., Leone, A.J., Tucker, J.W., 2023. Textual Analysis in Accounting:

What's Next? Contemporary Accounting Research 40, 765–805. https://doi.org/10.1111/1911-3846.12825

Brown, S.V., Tucker, J.W., 2011. Large-Sample Evidence on Firms' Year-over-Year MD&A Modifications. J of Accounting Research 49, 309–346. https://doi.org/10.1111/j.1475-679X.2010.00396.x

Chan, K., Hameed, A., 2006. Stock price synchronicity and analyst coverage in emerging markets. Journal of Financial Economics 80, 115–147. https://doi.org/10.1016/j.jfineco.2005.03.010

Chen Y., Yang Z., Wen X., 2024. Expectation Guidance, Economic Resilience and Macroeconomic Governance. Management World 40, 66–88. https://doi.org/10.19744/j.cnki.11-1235/f.2024.0124

Clarkson, P.M., Kao, J.L., Richardson, G.D., 1999. Evidence That Management Discussion and Analysis (MD&A) is a Part of a Firm's Overall Disclosure Package. Contemporary Accounting Research 16, 111–134. https://doi.org/10.1111/j.1911-3846.1999.tb00576.x

Dellavigna, S., Pollet, J.M., 2009. Investor Inattention and Friday Earnings Announcements. The Journal of Finance 64, 709–749. https://doi.org/10.1111/j.1540-6261.2009.01447.x

Dhaliwal, D.S., Li, O.Z., Tsang, A., Yang, Y.G., 2011. Voluntary Nonfinancial Disclosure and the Cost of Equity Capital: The Initiation of Corporate Social Responsibility Reporting. The Accounting Review 86, 59–100. https://doi.org/10.2308/accr.00000005

Durnev, A., Morck, R., Yeung, B., Zarowin, P., 2003. Does Greater Firm-Specific Return Variation Mean More or Less Informed Stock Pricing? Journal of Accounting Research 41, 797–836. https://doi.org/10.1046/j.1475-679X.2003.00124.x

Egeth, H., Kahneman, D., 1975. Attention and Effort. The American Journal of Psychology 88, 339. https://doi.org/10.2307/1421603

Fama, E.F., MacBeth, J.D., 1973. Risk, Return, and Equilibrium: Empirical Tests. Journal of Political Economy 81, 607–636. https://doi.org/10.1086/260061

Fang J., Wu Q., Fu Q., 2018. Limited Attention, Competitive Information and Market Reaction to Analyst Rating Reports. Journal of Financial Research 193–206.

Ferracuti, E., Lind, G., 2025. Macroeconomic Information Acquisition Around Earnings Clusters. J of Accounting Research 1475–679X.70020. https://doi.org/10.1111/1475-679X.70020

Goh, B.W., Lee, J., Li, D., Wang, M., 2024. Informativeness of Key Audit Matters: Evidence from China. AUDITING: A Journal of Practice & Theory 43, 139–164. https://doi.org/10.2308/AJPT-2020-099

Goldstein, I., Yang, S., Zuo, L., 2023. The Real Effects of Modern Information Technologies: Evidence from the EDGAR Implementation. Journal of Accounting Research 61, 1699–1733. https://doi.org/10.1111/1475-679X.12496

Guo, F., Lyu, B., Lyu, X., Zheng, J., 2025. Social Media Networks and Stock Price Synchronicity: Evidence from a Chinese Stock Forum. Abacus 61, 419–461. https://doi.org/10.1111/abac.12341

Guo S., Liang P., 2014. Information Channels and Household Stock Market Participation: Empirical Evidence Based on 2011 China Household Finance Survey. Economic Research Journal 49, 116–131.

Hansen, B.E., 1999. Threshold effects in non-dynamic panels: Estimation, testing, and inference. Journal of Econometrics 93, 345–368. https://doi.org/10.1016/S0304-4076(99)00025-1

He J., Qu B., Guo J., 2025. Fund Manager Media Coverage and Individual Investor Behavior: Evidence from Micro Big Data of Fund Subscription and Redemption. Journal of Financial Research 188–206.

He Y., Chen L., Du Y., 2024. Can Data Assetization Alleviate Financing Constraints of "Specialized, Sophisticated, Distinctive and Innovative" Small and Medium-sized Enterprises? China Industrial Economics 154–173. https://doi.org/10.19581/j.cnki.ciejournal.2024.08.008

Hirshleifer, D., Lim, S.S., Teoh, S.H., 2011. Limited Investor Attention and Stock Market Misreactions to Accounting Information. Rev Asset Pric Stud 1, 35–73. https://doi.org/10.1093/rapstu/rar002

Hirshleifer, D., Lim, S.S., Teoh, S.H., 2009. Driven to Distraction: Extraneous Events and Underreaction to Earnings News. The Journal of Finance 64, 2289–2325. https://doi.org/10.1111/j.1540-6261.2009.01501.x

Hirshleifer, D., Teoh, S.H., 2003. Limited attention, information disclosure, and financial reporting. J. Account. Econ. 36, 337–386. https://doi.org/10.1016/j.jacceco.2003.10.002

Hou Y., Ye D., 2008. Institutional Investors, Insider Trading and Market Efficiency: Empirical Evidence from China's Capital Market. Journal of Financial Research 131–145.

Hu J., Wang Z., 2015. Weibo, Idiosyncratic Information Disclosure and Stock Price Synchronicity. Journal of Financial Research 190–206.

Huang J., Guo Z., 2014. News Media Coverage and Capital Market Pricing Efficiency: Analysis Based on Stock Price Synchronicity. Management World 121–130. https://doi.org/10.19744/j.cnki.11-1235/f.2014.05.010

Huang R., Tao C., Han R., 2025. Accounting Treatment of Data Assets for Financial Service-Oriented Data Platform Enterprises: A Case Study Based on Shuku Technology. Management World 41, 181–199. https://doi.org/10.19744/j.cnki.11-1235/f.2025.0119

Jiang D., Peng X., Zhu H., 2023. Stock Momentum Spillover Effects of Related Companies in A-share Market: Evidence from Analysts' Common Coverage. Systems Engineering—Theory & Practice 43, 1891–1909.

Jiang T., 2022. Mediating and Moderating Effects in Empirical Research on Causal Inference. China Industrial Economics 100–120. https://doi.org/10.19581/j.cnki.ciejournal.2022.05.005

Jin, L., Myers, S.C., 2006. *R*2 around the world: New theory and new tests. Journal of Financial Economics 79, 257–292. https://doi.org/10.1016/j.jfineco.2004.11.003

Jones, C.I., Tonetti, C., 2020. Nonrivalry and the Economics of Data. American Economic Review 110, 2819–2858. https://doi.org/10.1257/aer.20191330

Kwan, A., Liu, Y., Matthies, B., 2026. Institutional Investor Attention. The Journal of Finance jofi.70009. https://doi.org/10.1111/jofi.70009

Li S., Shao H., Fang F., Lu F., 2025. Corporate Data Asset Information Disclosure and Capital Market Pricing Efficiency. China Industrial Economics 138–155. https://doi.org/10.19581/j.cnki.ciejournal.2025.07.010

Liu, D., Li, Mingzhu, Li, Mingsheng, Shi, J., 2025. Market reaction to announcement of accounting treatment of data assets: evidence from China. Journal of Accounting Literature. https://doi.org/10.1108/JAL-10-2024-0290

Liu, J.-Y., Jia, F.-X., Zhang, Y.-N., Liu, Z.-Y. (Ralph), Tse, Y.K. (Mike), 2026. Does data asset

disclosure mitigate stock mispricing? A signalling perspective. Enterprise Information Systems 2607361. https://doi.org/10.1080/17517575.2025.2607361

Liu, L.X., Lu, R., Sherman, A.E., Zhang, Y., 2023. IPO underpricing and limited attention: Theory and evidence. Journal of Banking & Finance 154, 106932. https://doi.org/10.1016/j.jbankfin.2023.106932

Liu, L.X., Sherman, A.E., Zhang, Y., 2014. The Long-Run Role of the Media: Evidence from Initial Public Offerings. Management Science 60, 1945–1964. https://doi.org/10.1287/mnsc.2013.1851

Lu, J., 2022. Limited Attention: Implications for Financial Reporting. Journal of Accounting Research 60, 1991–2027. https://doi.org/10.1111/1475-679X.12432

Ma Y., Yu Y., 2023. Many Hands Make Light Work: Social Networks with Imperfect Information Exchange and Market Quality. Systems Engineering—Theory & Practice 43, 3424–3440.

McCAHERY, J.A., Sautner, Z., Starks, L.T., 2016. Behind the Scenes: The Corporate Governance Preferences of Institutional Investors. The Journal of Finance 71, 2905–2932. https://doi.org/10.1111/jofi.12393

Morck, R., Yeung, B., Yu, W., 2000. The information content of stock markets: why do emerging markets have synchronous stock price movements? Journal of Financial Economics, Special Issue on International Corporate Governance 58, 215–260. https://doi.org/10.1016/S0304-405X(00)00071-4

Niu B., Yu X., 2024. Do Data Assets Attract Investor Preference? Evidence from the Perspective of Equity Capital Cost. Securities Market Herald 68–79.

Odean, T., 1999. Do Investors Trade Too Much? American Economic Review 89, 1279–1298. https://doi.org/10.1257/aer.89.5.1279

Peng H., Mao X., 2017. Government Innovation Subsidies, Executive Background and R&D Investment: Empirical Evidence from China's High-tech Industry. Finance & Trade Economics 38, 147–161.

Peng, L., Xiong, W., 2006. Investor attention, overconfidence and category learning. Journal of Financial Economics 80, 563–602. https://doi.org/10.1016/j.jfineco.2005.05.003

Piotroski, J.D., Roulstone, D.T., 2004. The Influence of Analysts, Institutional Investors, and Insiders on the Incorporation of Market, Industry, and Firm-Specific Information into Stock Prices. The Accounting Review 79, 1119–1151. https://doi.org/10.2308/accr.2004.79.4.1119

Song X., Chen Y., Lu D., Cheng J., 2024. Information Environment, Incremental Information Disclosure of Listed Companies and Capital Market Pricing Efficiency: Evidence from MD&A Text Similarity. Nankai Business Review 27, 30–39.

Wang M., Li D., 2019. New Audit Reports and Stock Price Synchronicity. Accounting Research 86–92.

Wang X., Gao X., He J., 2018. Annual Report Risk Information Disclosure and Audit Fees: Evidence from Text Cosine Similarity. Auditing Research 98–104.

Wang Y., Liu H., Wu L., 2009. Information Transparency, Institutional Investors and Stock Price Synchronicity. Journal of Financial Research 162–174.

Wang Y., Peng Z., Gao J., 2019. Investor Attention and Management Performance Forecast Timing: Active Monitoring or Excessive Pressure. Economic Management 41, 139–155. https://doi.org/10.19616/j.cnki.bmj.2019.02.009

Wang Y., Yang D., 2024. Transformation of China's Management Accounting System: From Data Elements to Data Assets. Management World 40, 171–189. https://doi.org/10.19744/j.cnki.11-1235/f.2024.0111

Wang Y., Yu L., An R., 2014. Can Non-financial Information Disclosure Improve the Capital Market Information Environment? Evidence from Social Responsibility Report Disclosure. Journal of Financial Research 178–191.

Wei Y., Zhang R., Wang F., Cheng M., 2022. A Study on the Relationship between Data Asset Information Disclosure and Analyst Earnings Forecasts: Empirical Evidence Based on Text Analysis. Journal of Industrial Engineering and Engineering Management 36, 130–141. https://doi.org/10.13587/j.cnki.jieem.2022.05.011

Wu W., Yue Zhao, Yan J., Wang S., 2020. Does Analyst Textual Tone Affect Stock Price Synchronicity? Mediation Effect Test Based on Stakeholder Behavior. Journal of Management Sciences in China 23, 108–126.

Xu N., Yu S., Yi Z., 2013. Herding Behavior of Institutional Investors and Stock Price Crash Risk. Management World 31–43. https://doi.org/10.19744/j.cnki.11-1235/f.2013.07.004

Yi Z., Yang S., Chen Q., 2019. Can Analysts Reduce Stock Price Synchronicity? Empirical Evidence Based on Textual Analysis of Research Reports. China Industrial Economics 156–173. https://doi.org/10.19581/j.cnki.ciejournal.2019.01.009

Yu, Z., Li, L., Tian, G., Zhang, H., 2013. Aggressive reporting, investor protection and stock price informativeness: Evidence from Chinese firms. Journal of International Accounting, Auditing and Taxation 22, 71–85. https://doi.org/10.1016/j.intaccaudtax.2013.07.004

Yuan Y., Wu Z., Liao J., 2025. Does Institutional Investor Distraction Exacerbate Management Language Inflation? Empirical Evidence Based on Annual Report Textual Tone. Nankai Business Review 28, 123–134.

Yuan Z., Yu X., Li M., 2022. Data Asset Information Disclosure, Heterogeneity of Institutional Investors and Corporate Value. Modern Finance and Economics (Journal of Tianjin University of Finance and Economics) 42, 32–47. https://doi.org/10.19559/j.cnki.12-1387.2022.11.003

Zhang R., Zhang Y., 2025. Data Asset Information Disclosure and Analyst Optimism Bias: "Halo Effect" or "Information Risk"? Journal of Central University of Finance and Economics 74–90. https://doi.org/10.19681/j.cnki.jcufe.2025.11.003

Zhao N., Liu J., 2024. Textual Characteristics of Internal Control Evaluation Reports and Capital Market Pricing Efficiency. Auditing Research 148–160.

Zhu H., He X., Tao L., 2007. Can Securities Analysts Improve Capital Market Efficiency in China? Empirical Evidence Based on Stock Price Synchronicity and Stock Price Information Content. Journal of Financial Research 110–121.