Membrane Technology
ISSN (online): 1873-4049

Data Asset Information Disclosure and Capital Market Efficiency: Empirical Research Based on Big Data Text Mining

Shuhui Wang*

Hainan University, NO.58 Remin Avenue, Meilan District Haikou, Hainan Province, 570228, China

**Corresponding author.*

Abstract:

In the era of big data, the significance of data has skyrocketed. It has transformed into a new type of asset and a significant element of production, which contributes immensely to the sustainable development and competitiveness enhancement of enterprises.This research focuses on the A-share listed companies in China over the period of 2007 to 2022, and uses Word2Vec neural network model and deep learning text mining method to investigate the ramifications and mechanism of data asset information disclosure on the effectiveness of corporate capital market. The findings indicate that: 1) data asset information disclosure noticeably enhances the efficiency of corporate capital market; 2) data asset information disclosure have a positive impact on corporate capital market efficiency by enhancing corporate governance, increasing institutional investors' shareholding and improving stock liquidity; 3) the data asset information disclosure has a more pronounced improvement effect on corporate capital market efficiency in growing and mature firms, and stronger in high-technology, state-owned, media-focused and investor-focused enterprises. The research conclusions are of great significance to improve the behavior of enterprise data assets information disclosure, enhance the capital market operation efficiency, as well as fostering the development of a robust and superior quality capital market.

Keywords: Data asset; Information disclosure; Capital market efficiency; Stock return synchronization; Text miningntroduction.

INTRODUCTION

With the accelerated pace of digital technology innovation and iteration, China listed data as the fifth major factor of production after land, labor, capital and technology in 2020, which became an crucial force to drive economic and social development. The Digital China Development Report (2022) shows that the scale of China's digital economy in 2022 will have exceeded 50 trillion yuan, with the digital economy accounting for 41.5% of GDP, ranking second in the world. This reflects China's rapid development and innovation in the Internet, artificial intelligence and other areas of the digital economy, with a large global influence, and the digital economy has injected a new impetus into China's economic development. Meanwhile, there are already enterprises that have gradually accumulated data assets in the booming data market, but there are no uniform provisions in the enterprise accounting standards on the issue of how to standardize the accounting treatment of data assets. Therefore, in August 2023, the Ministry of Finance issued the Interim Provisions on Accounting Treatment of Enterprise Data Resources, which came into effect on January 1, 2024, which means that the work of getting enterprise data assets onto the balance sheet has officially started. This will help enterprises to better strengthen the whole process management of data assets and delve into data assets' worth. However, from the perspective of specific operation, it is still difficult to effectively evaluate enterprises' level of information disclosure on data assets. At present, enterprises still mainly disclose data asset information through text of annual reports, but due to the large content of annual reports, external investors need to invest a lot of time and energy to obtain data asset information. At the same time, if investors cannot identify the practice of packaging adverse information in the annual report, they will make a wrong judgment on the value of the enterprise and make incorrect investment decisions. However, with the progress of computer technology and the maturity of machine deep learning methods, it is possible to use Word2Vec neural network model [1] to train the text corpus of enterprise annual report, obtain the similar word set of seed words, and construct the dictionary of data asset information disclosure, so as to measure the information content of data assets in companies' annual report text and help investors evaluate the level of companies' data assets information disclosure.

Data assets, as an emerging asset type in the process of economic and social digital transformation, are increasingly becoming an important strategic resource to encourage corporate development. Through the innovative application of data assets, enterprises can develop more intelligent and efficient products and services, and enhance the competitiveness and market position of enterprises. Based on the data presented in Fig. 1, the level of data assets information disclosure of Chinese listed enterprises between 2007 and 2022 is generally on rise. And investors, as rational economic people, will increasingly scrutinize the data assets information disclosed by enterprises. In addition, in this process, the information disclosure behavior of enterprises' data assets will also have a certain impact on capital market activities.

An efficient capital market allocates limited resources to the most efficient enterprises and industries by utilizing the signal mechanism of stock price, and the guiding role depends on the degree to which it reflects the fundamental information of the company that distinguishes from the whole market. The concept of stock return synchronization pertains to the extent of similarity in the trend of price fluctuation between different stocks over a certain period of time, in other words, changes of individual stock price will cause the average change of the market. Therefore, the synchronization of stock return is one of indicators to gauge capital market effectiveness. In contrast to the well-developed capital market, the emerging capital market presents a greater challenge for informed trading, resulting in a diminished presence of firm-specific information within stock prices., which leads to "signal" failure of the stock price, low market efficiency, and resources cannot flow to high-quality enterprises. China capital market is an emerging capital market, and the recurring pattern of stock return synchronization has always been one of the "stumbling blocks" hindering China's economic development. Thus, it has become an important task for China's financial reform to explore the factors that impact effectiveness of capital market and find an effective path to improve its effectiveness, which holds immense importance for maximizing the allocation of resources in capital market, improving capital market operation efficiency and fostering the sound growth of the China capital market.

Based on this, this paper discusses the relationship between enterprise data asset information disclosure and capital market efficiency. The following are the paper's innovations: 1) From the standpoint of data asset information disclosure, the paper embed the framework of data asset disclosure within the study of capital market efficiency, which is helpful to understand how such disclosure can influence capital market efficiency, and to offer a specific point for subsequent help of firm to measure the cost-benefit of data asset information disclosure. 2) The article studies the way in which data assets information disclosure affects the efficiency of enterprise capital market, and clarifies the "black box" of mechanism via which information disclosure of data assets impacts capital market efficiency, further expanding the research horizon of the financial effects of data assets disclosure. 3) form the perspective of enterprise life cycle, industry, property right nature, media attention and investor attention, the article explores heterogeneous effect of data asset disclosure on capital market efficiency. It has positive policy implications for the improvement and promotion of the information disclosure mechanism of enterprise data assets, and also provides fresh thoughts for enhancing capital market efficiency, which helps investors gain a deeper comprehension of the effectiveness of Chinese securities market. 4) In terms of variable measurement, the Word2Vec neural network model is used to gauge the level of data assets information disclosure by constructing a text dictionary and mine the data asset information in the text of enterprise annual reports. Businesses can use this as a helpful guide to assess the degree of information disclosure of their data assets.

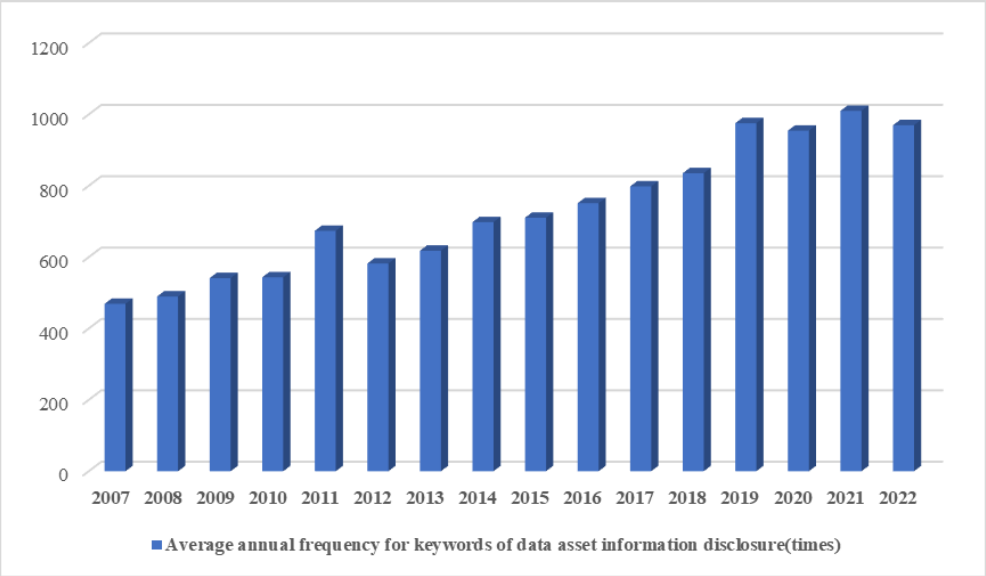

Figure 1. Status of data asset information disclosure of listed companies in China, 2007-2022

LITERATURE REVIEW

Related Research on Data Assets

Richard Peterson first proposed the notion of "data assets" in 1974. At present, the academic research on data assets and data asset information disclosure primarily centers around the subsequent aspects: 1) The definition of data assets. From an accounting point of view, in "White Paper on Data Asset Management Practices (6.0)" recently issued by China in 2023, data assets refer to the

data that an organization owns or controls, recorded electronically or otherwise. These data assets are measurable and can be bought or sold, ultimately generating economic and social advantages, either directly or indirectly. This means that only those data that can generate value for the enterprise and bring in economic benefits are data assets. However, unlike the above definition where data assets are "recorded electronically or otherwise", Veldkamp believes that only data recorded electronically is a data asset[2]. And from a broader perspective, data assets may not be limited to data controlled by the enterprise itself, but may also include other public data and open data, which will also enable enterprises to obtain valuable information and enhance their ability to gain insight into business opportunities[3]. 2) The benefits and costs of data asset information disclosure for enterprises. Firms that engage in information disclosure of data assets can gain access to more strategic resources and improve the level of enterprise innovation [4], and also further improve information transparency, attract high-quality talents and capital, reduce financing constraints [5], increase enterprise productivity [6] and drive the improvement of business models [7] thereby enhancing corporate value. However, data assets are a double-edged sword, which makes enterprises spend a lot of manpower, material and financial resources to collect, process, maintain and transform data, and there is uncertainty about whether these costs can ultimately create value for the enterprises [8].

Related Research on the Efficiency of Capital Market

In recent years, the research literature on capital market efficiency mainly elaborated two different views: 1) According to the perspective of information efficiency, which believes that augmentation of idiosyncratic information regarding a company in the stock price will lead to a decline in stock return synchronization [9] consequently an increase in capital market effectiveness. First of all, from the perspective of external factors of enterprises, green credit policy [10] and other factors will lessen the quantity of information about a company that is specific to stock prices; while factors such as, institutional investors [11], media coverage[12], help the company idiosyncratic information integrate into the stock price, thereby diminishing the stock return synchronization and improving capital market effectiveness. Secondly, in terms of internal factors of the enterprise, family control [13], share repurchase [14], managerial ability [15], quantity and quality of information disclosure [16] and so on will increase the company idiosyncratic information content reflected by stock prices, lower stock return synchronization, as well as increasing capital market efficiency. 2) The irrational behavior view, which argues that stock return synchronization is driven by irrational behaviors such as noise trading, investor sentiment, and bubbles. That is, the stock return synchrony decreases with increasing noise within stock prices [17], and therefore capital market efficiency declines. From the outside of the enterprise, Atmaz and Basak believe that stock price synchronicity will decrease with investors' heterogeneous beliefs[18], which in turn reduces capital market efficiency. From an intra-firm perspective, company site visits [20] have a noise reduction effect, which will reduce the irrational behavior of investors and enhance the synchronization of stock return, thus improving capital market effectiveness.

In summary, the role and impact of data asset information disclosure behavior in the capital market still need to be explored, and there is no further in-depth investigation into the impact and workings of data asset disclosure on capital market effectiveness. Therefore, on the basis of analyzing the current situation and future development of data asset disclosure in China, there is a great deal of room for research to study the influence of data asset disclosure on capital market effectiveness.

THEORETICAL DERIVATION AND HYPOTHESIS PROPOSITION

With the development of digital economy, data assets have been applied to various industries such as finance and industry, which has profoundly changed the value creation process of enterprises. Through the analysis of data assets, enterprises can understand key information such as market trends and competitive situation to provide decision-making support. Enterprises can also mine data assets to discover new business opportunities, and enhance their competitiveness. At the same time, by collecting and analyzing customer data, enterprises can better understand customer needs and preferences, so as to provide personalized products and services and increase customer loyalty. Therefore, the application of data assets in business activities has become an important measurement indicator to evaluate and predict the future growth potential and development prospects of enterprises, which means that data assets contain information about the corporate fundamentals, and can reflect and convey deeper understanding of the firm's fundamentals to external world. And the annual report serves as a crucial tool for investors to acquire and analyze the company's information. Therefore, when the company discloses the information of data assets in the annual report, it will inevitably attract the attention of investors, so that investors can receive more company-level information from the company and lessen information asymmetry. Meanwhile, data asset information disclosure proves that the enterprise has fulfilled its social responsibility, improved its reputation, established a good corporate image, and transmitted a positive signal to external investor, thus winning the trust of investors. These will cause the company's individual stock price to produce certain fluctuations, but the market and industry return will not change, that is, the change of single stock price and the market share price is not synchronized, and the capital market is more efficient.

In light of this, research hypothesis 1 is put out in this paper.

H1: data asset disclosure can enhance capital market efficiency.

Data asset disclosure can improve corporate governance level, thereby improving the capital market efficiency. Data asset disclosure can promote realization of internal and external governance incentive mechanism. On the one hand, data asset information disclosure evaluates and supervises internal governance; on the other hand, data asset disclosure helps information exchange between insiders and external investors, diminishes information asymmetry level, facilitates supervision of management behavior, reduces the moral hazard of management, suppresses managerial opportunism, and saves agency cost of the company, thus promoting the virtuous cycle of internal and external governance and enhancing corporate governance. Further, a higher level of corporate governance will also impose incentives and constraints on management, inhibit their motivation to use corporate resources to whitewash financial statements and seek self-interest, which is conducive to alleviating agency conflicts, improving corporate information transparency, promoting incorporation of corporate idiosyncratic information into stock prices, thus improving capital market efficiency. Moreover, an elevated level of corporate governance reduces cost private information collection and investors obtaining company-specific information cost. The decline in the cost to obtain information in turn has a ripple effect on arbitrage traders, fueling their risk-taking behavior., thus more corporate idiosyncratic information in stock price [9], leading to a more efficient capital market.

In light of this, research hypothesis 2 is put forth in this paper.

H2: Revealing data asset can boost capital market efficiency by improving corporate governance level.

Revealing data asset can increase institutional investors' shareholdings, thereby improving capital market effectiveness. Firstly, data asset information disclosure can enable institutional investors to better understand how enterprises apply data assets to production and operation activities, reduce investment risks, enhance the confidence of institutional investors, and thus attract more institutional investors to hold shares. Secondly, enterprises use data assets to promote technological innovation and product upgrading, which promotes the improvement of enterprise innovation level, brings a large number of external economic benefits into corporate, and increases investment value of business. Institutional investors can more comprehensively understand the application prospect of data assets, so as to find more investment opportunities, and therefore more likely to be favored by institutional investors. Compared with ordinary small and medium-sized investors, institutional investors not only possess the capability to obtain relevant firm status through external channels such as field research, but also can obtain information from within the company, so institutional investors have more diversified channels to obtain information, and there is an information advantages. Moreover, upon attaining a particular threshold of shareholding, institutional investors have outstanding expertise in professional knowledge and decision-making, and stronger information interpretation ability. They can transfer the value information they hold about this enterprise to the capital market through purchase and sale transactions, and the information environment is improved [21], and then the company fundamentals integrated into stock valuation increases, and capital market efficiency experiences a favorable boost.

In light of this, research hypothesis 3 is put out in the paper.

H3: Disclosure of data asset information can bolster capital market efficacy by enhancing shareholding of institutional investors.

Revealing data assets can enhance the fluidity of stocks, thereby augmenting the proficiency of the financial market. First of all, data assets, as an emerging asset type, have increased investors' demand for their private information; data asset information disclosure can convey private information to the market, increase private information supplement [5], and narrow information gap between informed and uninformed stakeholders. Consequently, investors are in a better position to forecast and make judgments on the future profitability as well as stock price risk of the company to reduce investment risks, so more investor participate in the market, and increase stock liquidity. Secondly, as an information intermediary, when enterprises disclose information about data assets, they will attract the attention of analysts, who use their advantages of information search and integration to interpret the information disclosed of data assets in an all-round, accurate and easy-to-understand manner, thus enhancing the transparency of corporate information and stock liquidity [5]. Thirdly, in the digital age, the disclosure of data asset information is a general trend, which is consistent with the national policy guidelines, so as to attract a large number of investors and increase the probability of stock trading, thus bringing about the improvement of stock liquidity. Finally, the information disclosure of data asset conveys positive news to the market, so investors will invest more in it, and the financial situation of enterprises will be improved, then the market will tilt resources to enterprises with good financial situation, thus improving stock liquidity. Stronger stock liquidity increases attention to the stock, lowers transaction costs, improves information

transmission efficiency, and makes stock valuation more conducive to fully manifesting the company's characteristic information, all of which contribute to improve capital market efficiency.

In light of this, research hypothesis 4 is put out in this paper.

H4: Information disclosure of data assets has the potential boost capital market efficiency by improving stock liquidity.

The theoretical framework of H1, H2, H3, H4 is graphically presented in following Fig. 2.

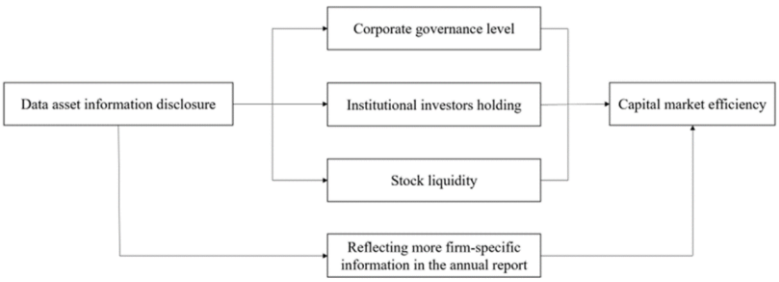

Figure 2. Relationship between variables

RESEARCH DESIGN

Sample Selection and Data Sources

In 2007, China Accounting Standards (CAS) were changed, so this article employs a sample of data encompassing all A-share listed enterprises in China from 2007 until 2022. And the following is how this paper handles with the samples: 1) Not including financial enterprises. 2) Not including ST, *ST, PT enterprises. 3) Excluding companies listed later than 2018. 4) Not including corporate with missing values of relevant data. 5) To diminish partiality in calculating capital market efficiency, stocks with fewer than 30 weeks of annual individual returns data are removed from the sample. Finally, this paper obtained 25,322 year-firm sample observations. The relevant data of all variables in this paper are derived from CSMAR, CNRDS and Sens Data. In order to mitigate the impact of outliers and improve the effectiveness of results, the continuous variables are winsorized at 1% and 99% levels.

Selection and Measurement of Variables

Core explanatory variable

The primary explanatory variable in the study is the unveiling of data asset information (data). Because before 2023, the CSA has not yet issued accounting regulations related to data assets, and data assets have not been included in the balance sheet, so the disclosure of data assets information is still voluntary information disclosure. However, in fact, some enterprises have already disclosed information on data assets in their annual reports, so this paper measures the degree of data assets disclosure via mining text information related to data assets in the annual reports.

The research utilizes text mining techniques to assess the level of data asset information disclosure of enterprises [22]. The measurement procedures are outlined below: First, in practice, some companies have begun to apply term of “data assets”, so they use “data assets” as a seed word. Secondly, the “White Paper 4.0 on Data Asset Management Practices” released by China in 2019 mentioned that data assets are data resources that yield significant financial gains for businesses. Therefore, the term “data resources” is also used as a seed word. Thirdly, according to the above seed words, the Word2Vec neural network model and deep learning technology are applied to obtain the similar word set of seed words [1,23], and only the words with similarity greater than or equal to 0.5 are retained to complete the construction of the dictionary, which makes the dictionary more accurate to measure the level of information disclosure of data assets. Third, mine the word frequency of all words in dictionary of annual report every year, and calculate the level of information disclosure of data assets. The calculation formula is as follows:

$$data_{it} = \frac{\sum dictionarywords_{itn}}{totalwords_{it}} \times 100 \quad (1)$$

Among them, $dictionarywords_{itn}$ is the exact word frequency of the nth word in the dictionary in individual stock i’s annual report for year t; $totalwords_{it}$ represents the total word frequency of individual stock i’s annual report for year t (excluding

English and numbers). The greater the magnitude of *data*, the greater the amount of details revealed regarding enterprise's data assets.

Explained variable

Capital market efficiency (*CME*) is represented by stock price synchronicity (*SYN*). The sub-yearly regression of model (2) is calculated to obtain the goodness of fitting $R_{i,t}^2$ then $R_{i,t}^2$ is logarithmically processed, as in model (3), and *SYN* as the yearly synchronization of stock price of company i is calculated [24]. The larger *SYN* stands for the lower capital market efficiency.

$$r_{i,t}=\beta_0+\beta_1 r_{m,t}+\beta_2 r_{I,t}+\varepsilon_{i,t} \tag{2}$$

$$SYN_{i,t}=ln\left(\frac{R_{i,t}^2}{1-R_{i,t}^2}\right) \tag{3}$$

Among them, $r_{i,t}$ stands for weekly stock yield of firm i accounting for reinvestment of cash dividend in week t; $r_{m,t}$ demotes circulation market value weighted comprehensive weekly market yield of company i in week t factoring cash dividends reinvested; $r_{I,t}$ indicates circulation market value weighted comprehensive weekly market yield of other stocks in week t ,taking into account cash dividend reinvested after excluding stock i in the industry where company i is located. Among them, the industry classification refers to the 2012 industry classification standard of the China Securities Regulatory Commission, with secondary industry classification for manufacturing and primary industry classification for others.

Mechanism variables

The study employs the principal component analysis approach to analyze the nine metrics in three dimensions: shareholders, board of directors and incentive mechanism to construct a comprehensive index, and selects the leading major principal component to evaluate extent of corporate governance (*Govern*) [25]. The greater the magnitude of *Govern*, the higher level of corporate governance. An the specific design is as follows: First, the shareholder level is expressed by four indicators: the ownership proportion of topmost shareholder, proportion of shares held by second through tenth largest shareholders compared to that of topmost shareholder, whether it is a state-owned company or not, and institutional investors' shareholding ratio. Secondly, the board level is measured by three indicators: the ratio of non-executive directors, the board size, and whether the chairman and the general manager are in one. Finally, from the perspective of incentive mechanism, it is expressed by two indicators: the management shareholding ratio and the top three management's remuneration.

In this paper, cumulative amount of shares owned collectively by all institutional investors as a percentage of the outstanding A-shares' listed enterprises denotes institutional investors (*Inst*). The calculation formula is as below:

$$Inst=\frac{cumulative\ amount\ of\ shares\ owned\ by\ all\ institutional\ investors}{Circulating\ A\text{-}shares}\times 100\% \tag{4}$$

The stock liquidity (Liqui) is expressed by the annual average daily turnover rate (TO), and the computation formula is given as below.

$$TO_{it}=\frac{Volume_{it}}{Outshare_{it}}\times\frac{1}{D_{it}} \tag{5}$$

Among them, TO_{it} is the stock i's mean daily turnover ratio for year t; $Volume_{it}$ is the stock i's quantity of transaction for year t; $Outshare_{it}$ stands for stock i's number of outstanding shares for year t; D_{it} denotes total stock i's number of trading days for year t.

Control variables

Table 1 displays the relevant control variables that are selected for model construction based on the literature.

Table 1. Explanation of variables

Variable category	Variable name	Variable symbol	Variable definition and measurement
Core explanatory variable	Data asset information disclosure	data	See (1) for details
Explained variable	Capital market efficiency	CME	See (2)(3) for detailed formulas
Mechanism variables	Level of corporate governance Institutional investor holdings Stock liquidity	Govern Inst Liqui	See detailed process described in the paper See (4) for details See details (5)
Control variables	Company scale	Employee	The quantity of staff members is taken as natural logarithm
	Financial leverage	Lev	The ratio of year end liabilities to year-end assets
	Return on total assets	ROA	Profit after tax divided by assets
	Tobin q	TobinQ	Market capitalization of company divided by year-end assets
	Net cash flow	Cashflow	The ratio of net cash flow from operating activities to assets
	Fixed asset investment ratio	FIXED	The ratio of total fixed assets to assets
	Enterprise growth	Growth	Growing of main business revenue
	Shareholding of top five shareholder	TOP5	Shareholdings of top five shareholders as a percentage of all shares
	Age of enterprise	Age	Time span between release of firm's annual report and the time of company's incorporation is taken as natural logarithm
	Number of Board of Directors	Board	The natural logarithm is used to determine the count of individuals serving on a board
	Integration of two functions	Dual	Whether the chairman and CEO of the enterprise are same person, if so take the value of 1, and vice versa take 0
	Proportion of independent directors	Indep	The proportion of year-end non-executive directors to the overall number of directors
	Year	Year	Control the year fixed effect of the selected samples
	Industry	Ind	Control the industry fixed effect of the selected samples

Model Design

The study develops a regression model (6) to examine hypothesis 1.

$$SYN_{it}=\alpha_0+\alpha_1 data+\sum_k \varphi_k Controls_{kit}+\sum Year_{it}+\sum Ind_{it}+\gamma_{it} \quad (6)$$

Among them, the capital market efficiency (*CME*) is the explained variable, which can be measured by stock price synchrony (*SYN*). Data asset information disclosure (*data*) is the core explanatory variable, *Controls* is the set of control variables, *k* represents the number of above control variables, *Year and Ind* are the year and industry fixed effects, and γ is random error term. To enhance the accuracy of estimation, the research has carried on robust standard error treatment.

In order to further test Hypothesis 2, Hypothesis 3 and Hypothesis 4 of the mechanism path, the study presents the construction of following framework.

$$mediator_{it}=\alpha_0+\alpha_1 data+\sum_k \varphi_k Controls_{kit}+\sum Year_{it}+\sum Ind_{it}+\gamma_{it} \quad (7)$$

Among them, *mediator* is the mechanism variable, including corporate governance level, the shareholding of institutional investors, stock liquidity. The defining of remaining variables align with those in preceding part.

EMPIRICAL ANALYSIS

Descriptive Statistic

Table 2 shows descriptive statistical results. The mean of data asset information disclosure (data) is 1.3051, the standard deviation is 0.2505, and the maximum is 2.2582, which demonstrates that the extent of data asset disclosure varies widely among different firms. The standard deviation of the stock price synchronicity (SYN) is 1.0270, the average is -0.5404, and the minimum is -3.9109, which suggests that a significant disparity in capital market efficiency among companies listed. The average of corporate

governance (Govern) is 0.0331, and has a standard deviation of 0.9882, a minimum of -2.0302, indicating that enterprise's corporate governance level is generally low, and gap of the corporate governance level among corporate is substantial. Institutional investors (Inst) has a average value of is 0.6931, a standard deviation of 0.7233, and a maximum of 6.0869. This shows that ownership amongst institutional investors is unevenly distributed and it is lower. The mean value of stock liquidity (Liqui) is 2.6820, and it has a minimum is 0.2496, and a maximum is 19.7388, which says that there is a significant gap in stock liquidity among companies and most of the companies have poor stock liquidity.

Table 2 Descriptive statistical

	count	mean	sd	min	p50	max
SYN	27384	-0.5404	1.0270	-3.9109	-0.4219	1.4055
data	27384	1.3051	0.2505	0.7488	1.2803	2.2582
Govern	27384	0.0331	0.9882	-2.0302	0.1385	1.9173
Inst	27384	0.6931	0.7233	0.0062	0.5396	6.0869
Liqui	27384	2.6820	2.2551	0.2496	2.0147	19.7388
Employee	27384	7.7479	1.2692	2.3026	7.6748	13.2535
Lev	27384	0.4299	0.1992	0.0508	0.4260	0.8975
ROA	27384	0.0431	0.0624	-0.2298	0.0401	0.2262
TobinQ	27384	2.0160	1.2536	0.8534	1.6189	8.3532
Cashflow	27384	0.0508	0.0689	-0.1716	0.0489	0.2482
FIXED	27384	0.2167	0.1603	0.0021	0.1847	0.6980
Growth	27384	0.1635	0.3699	-0.5740	0.1106	2.5807
TOP5	27384	0.5295	0.1516	0.2011	0.5286	0.8853
Age	27384	2.2027	0.7591	0.0000	2.3026	3.3322
Board	27384	2.1324	0.1976	1.6094	2.1972	2.7081
Indep	27384	0.3747	0.0532	0.3125	0.3333	0.5714
Dual	27384	0.2679	0.4429	0.0000	0.0000	1.0000

Benchmark Regression Analysis

Benchmark regression results are depicted in Table 3. Columns (1), (2), (3) demonstrate that after controlling fixed effects of year and industry, with control variables are gradually incorporated, the coefficients of data asset information disclosure(*data*) are -0.1200, -0.0877 and -0.1417 respectively, and the regression coefficients exhibit a substantial negative at the level of 1%. And for every standard deviation (0.2505) increase in the degree of data asset revelation, the capital market efficiency improves by an average of about 6.57% (0.2505×0.1417/0.5404). This suggests that more an enterprise reveals information about data assets, the more information about enterprise's fundamentals is conveyed to the outside world, leading to firm idiosyncratic information encompassed within stock price will increase, and capital market function more effectively, i.e., the information disclosure of data assets has a significant enhancement impact on capital market efficiency, displaying remarkable positive correlation among the two. Thus, H1 has been verified.

Table 3 Baseline regression results

	(1)	(2)	(3)
	SYN	*SYN*	*SYN*
data	-0.1200***	-0.0877***	-0.1417***
	(-4.7664)	(-3.5122)	(-5.6580)
Employee	0.1338***	0.1083***	0.0981***
	(28.0794)	(21.9114)	(19.0405)
Lev	-0.1923***	-0.1631***	-0.2674***
	(-5.8803)	(-4.9907)	(-8.0841)
ROA	-0.4951***	0.2462**	0.5367***
	(-5.3384)	(2.3436)	(5.0805)
TobinQ		-0.0953***	-0.1019***
		(-18.4717)	(-19.7310)
Cashflow		-0.0167	-0.0147
		(-0.1979)	(-0.1760)
FIXED		0.0463	0.0320
		(1.1541)	(0.8034)

Growth		-0.1556***	-0.1414***
		(-10.4755)	(-9.5279)
TOP5			-0.4070***
			(-10.8665)
Age			0.0945***
			(11.1883)
Board			0.0870***
			(2.8383)
Indep			0.2604**
			(2.4111)
Dual			-0.0118
			(-1.0143)
Constant	-1.3169***	-0.9970***	-1.0741***
	(-24.5277)	(-17.8847)	(-10.3461)
Observations	27,383	27,383	27,383
R-squared	0.3777	0.3901	0.3994

Note: t-values in parentheses; *, ** and *** indicate significant at the 10%, 5%, and 1% levels.

Robustness Test

Adjust the time range of the sample

The sample time range of the previous paper is 2007-2022, but considering that the outbreak of the epidemic in 2019-2022 will have some interference with China capital market efficiency, the paper excludes the samples that are in the interval of 2019-2022 and repeats the previous test for regression. The findings from the regression analysis presented in column (1) of Table 5 indicate that that after excluding the impact of epidemic, the information disclosure of data assets passes the 10% significance test and is negatively correlated, which suggests that outcomes of the study are robust.

Fixed effect model

To further mitigate the influence of the company' own characteristic differences regarding regression outcomes, the study re-runs regression after controlling individual fixed effects. The data presented in Column (2) of Table 5 shows that the regression coefficient of disclosure about data asset information remains still noteworthy negative at 1% level. This suggests that the conclusion of this paper is reliable and the absence of variables that remain constant over time will not alter the outcome.

Replacement of explained variable measurement

The stock return synchronicity (*SYN1*) can be re-measured representing capital market efficiency [26]. The calculation process is as follows:

$$R_{i,w,t}=\beta_0+\beta_1 R_{M,w,t}+\beta_2 R_{M,w-1,t}+\beta_3 R_{I,w,t}+\beta_4 R_{I,w-1,t}+\varepsilon_{i,w,t} \quad (8)$$

$$SYN1_{i,t}=ln\left(\frac{R^2_{i,t}}{1-R^2_{i,t}}\right) \quad (9)$$

Among them, $R_{i,w,t}$ is stock i's weekly individual stock return rate of week w for year t considering reinvestment of cash dividends; $R_{M,w,t}$ stands for outstanding market capitalization-weighted composite weekly market yield of all A-share listed companies of week w for year t; $R_{I,w,t}$ denotes outstanding market capitalization-weighted composite weekly market yield of week w for year t considering cash dividend reinvestment of other stocks that are in industry which stock i belongs to after excluding stock i, and R^2 is calculated. And logarithmic processing of R^2, getting the company i's stock return synchronization (*SYN1*) for year t.

Then, *SYN1* is brought into main regression model for re-testing. The findings displayed in column (3) of Table 4, indicate that coefficient of data asset information disclosure is still substantially inversely linked negatively correlated at 1% threshold, and fundamental not changes. The outcomes are robust to findings that we drawn.

Replace the measurement of core explanatory variable

The above is the ratio of key word frequency of data asset information disclosure to the overall word occurrence frequency of this paper. Since the data asset information disclosure does not conform to the normal distribution, the above equation (1) is logarithmically processed and normalized respectively, to obtain a new way to measure the degree of companies' data asset disclosure *data1* and *data2*. As indicated by regression findings of column (4), (5) in Table 4, that after retesting baseline regression

model with data1 and data2, the level of information disclosure of data assets is markedly negative at 1% threshold. The findings validate that the conclusions drawn in this paper are not affected by the measurement methods of core explanatory variables, and the results of disclosure on data asset significantly boosting capital market efficacy remain robust.

The aforementioned robustness test findings collectively show that, following a battery of robustness tests, the core conclusion "disclosure about data asset information can considerably enhance capital market effectiveness" remains true.

Table 4 Robustness test results

	(1)	(2)	(3)	(4)	(5)
VARIABLES	*SYN*	*SYN*	*SYN1*	*SYN*	*SYN*
data	-0.0507*	-0.2178***	-0.1310***		
	(-1.7963)	(-5.8077)	(-5.8489)		
data1				-0.3697***	
data2				(-6.3058)	-1.3339***
					(-6.3058)
Controls	Yes	Yes	Yes	Yes	Yes
Constant	-0.6021***	-0.8557**	-0.8825***	-0.9416***	-0.9416***
	(-5.2863)	(-4.3122)	(-9.4557)	(-8.4487)	(-7.7737)
Industry	Yes	No	Yes	Yes	Yes
Stock	No	Yes	No	No	No
Year	Yes	Yes	Yes	Yes	Yes
Observations	19,040	27,309	27,383	27,383	27,383
R-squared	0.3130	0.5268	0.4017	0.3996	0.3996

Endogeneity Test

Instrumental variables test

There may be a reciprocal causal relationship between the information disclosure of data assets and capital market efficacy, that is, an rise in capital market efficiency may also result in an increase in the amount of information disclosed about data assets. Therefore, to tackle the problem, the instrumental variable is represented by the industry mean of the data asset information disclosure excluding the enterprise's own data for the same year, as well as applies 2SLS for analysis. In the first stage, the Cragg-Donald Wald F statistic is 206.16, which is much larger than 10, and passes the Weak identification test. Column (1) in Table 5 reveals that the coefficient of instrumental variable to data asset information disclosure is 0.2855, indicating a strongly positive correlation at 1% threshold. Upon examining the findings of column (2) in Table 5, it is evident that in the second stage of regression, after controlling the endogenous problem, the regression coefficient of data asset information disclosure is -0.9076, which displays a significantly negative relationship at 5% threshold. This demonstrates that capital market effectiveness increases with the degree of data asset information disclosure. The basic conclusion of this paper has not changed, and the results are robust.

First-order difference model test

To check if there is also a significant improvement association among the increment of capital market efficiency and increment of data asset disclosure, this article uses the first-order difference model to verify. The specific step is to differentiate the variables on both sides of formula (6). Column (3) of Table 5 displays that coefficient of the first-order differential data asset information disclosure (Δ*data*) is -0.1421 and significant at 1% threshold, which provided additional backing to conclusions of Hypothesis 1.

The above endogenous test results further confirm that the primary regression conclusion, asserting that "the substantial positive correlation between data asset information disclosure and capital market efficiency" is robust.

Table 5 Endogeneity test results

	(1)	(2)	(3)
VARIABLES	data	SYN	ΔSYN
data_iv	0.2855***		
	(10.2784)		
data		-0.9076**	
		(-2.1499)	
Δdata			-0.1421***

			(-2.7015)
Controls	Yes	Yes	Yes
Constant	1.1713*** (23.2704)		-0.0789*** (-7.5040)
Industry Stock Year	Yes No Yes	No Yes Yes	Yes No Yes
Observations	27,289	27,289	23,743
R-squared	0.7804	0.0136	0.3516

FURTHER ANALYSIS

Mechanism Test Analysis

Based on the previous analysis of hypothesis 2,3 and 4, the srticle confirms the mechanism of disclosing data asset to improve capital market efficiency from three aspects: corporate governance level, institutional investor shareholding and stock liquidity. The mechanism path test results are displayed in Table 6. The column (1), (2) (3) of Table 6 are the test results of corporate governance level mechanism, institutional investor shareholding mechanism and stock liquidity mechanism, respectively. First of all, it is verified through the channel of "data asset information disclosure→(improve) corporate governance level→(improve) capital market efficiency". Column (1) reveals that the coefficient of data asset information disclosure is 0.1052 and highly positive at 1% significance threshold. It suggests that corporate data asset information disclosure can perfect the corporate governance model, enhance corporate governance level, augment characteristic informative value of stock price, thus enhance capital market efficacy. Secondly, it is verified through the channel of "data asset information disclosure→(increase) institutional investor shareholding→(improve) capital market efficiency". As can be seen in Column (2), the coefficient for disclosing data asset information is 0.0403 and is found to be at 5% statistically significance at least at the 5% threshold, which proves effectiveness of institutional investors shareholding as an intermediary channel, that is, data asset information disclosure is conducive to enhancing confidence and trust of institutional investors in enterprises and increasing institutional investors' shareholding proportion. Therefore, enterprises can take advantage of economies of scale such as diversified information channels and specialization of knowledge to enhance the of stock prices to accurately mirror the company idiosyncratic information, thereby maximizing capital markets effectiveness. Finally, it is verified through the channel of "data asset information disclosure→(improve) stock liquidity→(improve) capital market efficiency". As demonstrated in column (3) that coefficient of disclosing data asset is 0.3681 and at1% significance, verifying the mechanism of stock liquidity, that is, data asset information disclosure can increase the probability of investors' stock trading, enhance stock liquidity, and then improve efficiency of stock price reflecting firm idiosyncratic information, so as to increase the efficiency of capital market. In summary, H2, H3 and H4 are supported by the evidence and can thus be verified.

Table 6 Mechanism test

	(1)	(2)	(3)
VARIABLES	*Govern*	*Inst*	*Liqui*
data	0.1052***	0.0403**	0.3681***
	(5.5519)	(2.2624)	(6.1843)
Controls	Yes	Yes	Yes
Constant	-4.9180***	-1.1467***	9.2109***
	(-65.0011)	(-13.6250)	(39.7400)
Industry Year	Yes Yes	Yes Yes	Yes Yes
Observations	27,383	27,383	27,383
R-squared	0.6355	0.3553	0.3278

Heterogeneity Test

Enterprise life cycle

As per life cycle hypothesis: First, the main objective of growing enterprises is to rapidly open up market space and enhance competitiveness. At the same time, due to the rapid expansion of the scale, enterprises may seek external financing to support their sustainable development. Second, the profits of enterprises in the mature period remain at a high level, the market share and

organizational structure tends to be stable, and attention is paid to innovation and diversified development. Third, in the decline period, the organizational structure of enterprises is rigid, the market share is gradually declining, the profit is also plummeting, and the financial situation is beginning to become tense. Enterprises may face problems such as shortage of funds, increase in debt and lack of innovation, as well as having difficulty in finding new growth points. This paper divides the enterprise life cycle into three stages: growing phase (=1), mature phase (=2) and declining phase (=3) depending on combination of net operational, net investment and net financing cash flow, both positive and negative [27] as illustrated in Table 7. Table 8 demonstrates the findings of the grouping test, which reveals that coefficient for disclosing data asset is noticeably negative at 1% threshold for enterprises in growth and maturity stages, while it is not significant for enterprises in declining phase. This shows that in contrast with the decline period, the data asset information disclosure of enterprises which are in the growth or maturity stage has a substantial improvement effect on capital market efficacy. First of all, for enterprises in the growth period, this is mainly because enterprises usually disclose data asset information more actively in order to attract investment and expand market share. At the same time, because enterprises in the growth period usually show higher growth potential and uncertainty, which makes investors focus more on specific operation, future growth foreground and potential risks of enterprises, and the demand for information is more urgent. Therefore, when companies disclose data asset information in time, it is able to assist investors better understand operating status and potential value of firm, improve efficiency of stock price to reflect characteristic information, and then enhance capital market efficacy. Secondly, for mature enterprises, this is mainly because their business environment and performance are usually stable, and there are sufficient resources to drive enterprise innovation, which means that the enterprise 's data asset information is more reliable and predictive. Meanwhile, because at this time investors' focus on enterprises has gradually shifted from high-speed growth to stable profitability and sustained growth, and disclosing data asset can meet investors' information needs in these aspects, making stock prices more reflective of the intrinsic value, thus enhancing capital market effectiveness.

Table 7 Division of enterprise life cycle based on cash flow

Net cash flow	Growing phase		Mature phase	Declining phase				
	Introduction phase	Growing phase	Mature phase	Declining phase	Declining phase	Declining phase	shakeout phase	shakeout phase
Operating	−	+	+	−	+	+	−	−
Investment	−	−	−	−	+	+	+	+
Financing	+	+	−	−	+	−	+	−

Table 8 Heterogeneity test results 1

	(1)	(2)	(3)
	Gowth	Maturity	Decline
VARIABLES	*SYN*	*SYN*	*SYN*
data	-0.1875***	-0.1409***	-0.0722
	(-4.9988)	(-3.3399)	(-1.2704)
Controls	Yes	Yes	Yes
Constant	-0.8927***	-1.0882***	-0.9747***
	(-5.9196)	(-6.1484)	(-4.0310)
Industry	Yes	Yes	Yes
Year	Yes	Yes	Yes
Observations	11,539	10,477	5,258
R-squared	0.4163	0.3806	0.4288

Industry of enterprise

In different industries, high-tech businesses typically place a strong emphasis on technology, while non-high-tech enterprises may focus more on existing products or services, have less dependence on innovation, and focus more on traditional business models and operating modes [28]. In 2012, the China Securities Regulatory Commission (CSRC) categorized the industry, and the companies whose company classification codes belong to C25-C29, C31-C32, C34-C41, 163-165 and M73 are defined as high-tech corporate, hence the rest belong to non-high-tech industry companies. If the business operates in the high-tech sector, its *HTec* value is 1; otherwise, it's 0. The findings in column (1) of Table 9 reveals that the coefficient of interaction term linking the industry

and the information disclosure of data assets (*HTec*×*data*) is -0.0914 and statistically at 5% significance threshold, which suggests that within high-tech enterprises, the disclosure of data assets information exerts a more pronounced impact on efficiency of capital market. This is due to the fact that in contrast to non-high-tech companies, high-tech companies have a greater capacity for advancing technology, data collection and analysis, and tend to have advanced technology and data assets, while most non-high-tech firms do not have the ability to own data assets. Therefore, among high-tech firms, data asset disclosure has a stronger role in promoting capital market efficiency.

Nature of business ownership

Compared with privately-owned businesses, the appointment of management in state-owned businesses usually falls under government intervention, also there may be a mismatch between personal skills and positions [29]. Moreover, the business objectives of state-owned enterprises extend beyond maximizing profits, but also bear social responsibilities and national strategic tasks, such as ensuring employment and maintaining social stability. Therefore, state-owned enterprises may abandon high-risk but high-return investment projects and reduce R&D investment under this goal; whereas privately-owned corporate prioritize the improvement of financial efficiency and profitability, and pursue market competition and commercial interests. Therefore, separating businesses into state-owned businesses and privately-owned businesses in terms of nature of property ownership(*SOE*). If it's state-owned enterprise, the value-taking of *SOE* is 1, conversely it's 0. As indicated in column (2) of Table 9, the coefficient for interaction term between property ownership and data assets disclosure (*SOE*×*data*) is -0.1051 passing at 1% significance threshold, which suggests that in state-owned enterprises, data assets information disclosure has a more significant effect on the improvement of capital market efficacy. This is because state-owned enterprises are stricter oversight, transparency and accuracy of data asset information disclosure are higher to meet the government's supervision requirements for state-owned enterprise. This high degree of transparency and accuracy helps to reduce information asymmetry. Also, state-owned businesses enjoy policy support' advantages and resource acquisition, which makes them have more resources and capabilities in data asset information disclosure, which will release more company-specific information and make the capital market more efficient.

Media attention

Companies with high media attention will cause damage to the reputation of directors and companies if they have negative news; if there is positive news, it will promote the spread of this management trend, thereby improving corporate governance. To express media attention, using natural logarithm of company's annual media coverage plus 1. According to the median of media attention, the enterprise surpassing the average is the high media attention group, denoted by a Media value of 1, vice versa, the low media attention group, with a *Media* value of 0. Column (3) of Table 9 reveals that the interaction coefficient of media attention and data asset information disclosure (*Media*×*data*) is -0.1048, exhibiting a remarkable correlation at 1% threshold and demonstrating that in enterprises with high media attention, data asset information disclosure is particularly significant in enhancing capital market efficacy. It is because that, first, companies with high media attention are affected by market pressure to will disclose data asset information more energetically in order to maintain their market image and reputation. Second, companies with high media attention are often subject to more investor attention and market observation. Such supervision can encourage companies to enhance quality of disclosing information, and reduce degree information asymmetry. Third, high media attention helps to improve the visibility and influence of enterprises, thereby increasing the liquidity and trading volume of stocks. In this case, the influence of data asset disclosure is much more significant on stock valuation efficiency. Therefore, in enterprises with high media attention, a greater amount of information regarding fundamentals of enterprises gets assimilated in stock price, resulting in higher capital market efficiency.

Investor attention

The effect of data asset information disclosure on improving efficiency of capital market will also be affected by the external constraint mechanism of investor attention. Relative to enterprises with low investor attention, the stock prices of enterprises with high investor attention tend to be more active and volatile. This is because more individuals invest in the purchase of stocks, resulting in price pressure that can make temporary deviation of stock prices. This paper defines investor attention as natural logarithm of enterprises' annual Baidu index median plus 1. According to the median of investor attention, the company higher than the median is high investor attention group, the *Attention* value is 1, on the contrary, the group with low investor attention, the *Attention* value is 0. Column (4) of Table 9 The interaction coefficient between investor attention and data asset information disclosure (*Attention*×*data*) is -0.0768 at 10% significance threshold, noting that in enterprises with high investor attention, data asset information disclosure further enhances capital market efficicacy. It is because, first, high investor attention means that investors have higher expectations and requirements for these companies. Therefore, in order to cater to investors, companies will improve the transparency of data asset information disclosure. Second, companies with high investor attention usually mean that

their stocks are favored and concerned by more investors in the market. The higher the degree of connection between company's stock price and external environment, the easier it is to be fed back and adjusted by the market. Therefore, when companies disclose data asset information, the company idiosyncratic information becomes assimilated in the stock value faster, rendering the capital market more efficient.

Table 9 Heterogeneity test results 2

	(1)	(2)	(3)	(4)
VARIABLES	*SYN*	*SYN*	*SYN*	*SYN*
data	-0.0883***	-0.1165***	-0.0966***	-0.0926**
	(-2.5831)	(-3.8639)	(-3.0367)	(-2.4229)
HTec	-0.0202			
	(-0.2258)			
HTec×data	-0.0914**			
	(-1.9843)			
SOE		0.2634***		
		(4.9300)		
SOE×data		-0.1051***		
		(-2.5975)		
Media			0.0709	
			(1.3871)	
Media×data			-0.1048***	
			(-2.7037)	
Attention				0.0772
				(1.3508)
Attention×data				-0.0768*
				(-1.8192)
Controls	Yes	Yes	Yes	Yes
Constant	-1.0338***	-0.9015***	-1.1929***	-1.1458***
	(-8.8833)	(-8.3768)	(-11.1897)	(-10.4283)
Industry	Yes	Yes		
Year	Yes	Yes		
Observations	23,391	27,383	27,383	27,383
R-squared	0.4140	0.4017	0.4005	0.3996

CONCLUSIONS AND IMPLICATIONS

Taking A-share listed corporations in China between 2007 and 2022 as study subjects, this paper uses Word2Vec neural network model based on natural language processing technology and machine learning to construct a dictionary about data asset disclosure level, and mines data asset information in the text of enterprise annual report to measure the information degree with respect of data asset disclosure, and empirically tests impact and mechanism in relation to disclosing data assets on capital market effectiveness. The findings illustrate that the capital market's effectiveness increases with the amount of data asset information given by businesses. Mechanism test analysis shows that data asset disclosure can promote capital market efficacy through three channels: enhancing corporate governance, increasing institutional investors holdings and improving stock liquidity. Heterogeneity analysis finds that in growing and mature firms, and in high-technology, state-owned, media-focused and investor-focused enterprises, data asset disclosure produces a particularly crucial effect on improving capital market effectiveness.

The article makes recommendations in light of aforementioned findings.: First, for government departments (1) Relevant departments should formulate and improve related laws and regulations on the disclosure of data assets information, clarify the content, format and standards of disclosure, and ensure that enterprises disclose data assets information in accordance with unified standards. (2) The government should strengthen the supervision of information disclosure of enterprise data assets, and punish the behavior of not disclosing or deliberately concealing information according to regulations, so as to ensure the fairness, justice and transparency of the capital market. (3) The government should promote the standardization process of data asset evaluation and transaction, establish a data asset evaluation system, standardize data trading behavior, and increase the market value of data assets. (4) The government should optimize the investment environment, encourage the development of institutional investors and increase their proportion in the capital market, as well as optimizing the stock trading mechanism to improve the liquidity of the stock market. Second, for listed companies (1) Enterprises should fully recognize the value and importance of data assets, lift its

management and protection, as well as ensuring its security and integrity. (2) Enterprises should establish a standardized data asset information disclosure process, specify the content, scope and time of disclosure to ensure the timeliness and accuracy of information. (3) Enterprises should establish a sound corporate governance structure, regulate their own earnings management behavior, and avoid misleading investors and markets by manipulating earnings, so as to improve company governance level, and then better promote the efficiency of capital market. (4) Enterprises should actively communicate with investors and the media, regularly disclose data asset information, respond to market concerns in a timely manner, and enhance market confidence and transparency. Thirdly, for investors (1) Investors should enhance the awareness of data assets investment, recognize the importance and potential of data assets in capital market. Meanwhile, they should attend to information disclosure of enterprise data assets, understand status and value of data assets, and actively pay attention to the investment opportunities related to data assets, in order to provide references for investment decisions. (2) When choosing investment targets, investors should raise risk awareness, be alert to enterprises with abnormal earnings fluctuations, pay attention to the governance structure and governance level of the company, as well as the volatility and risks of the stock market, and rationally allocate institutional investors to reduce investment risks by using their professional investment ability and rich market experience. (3) Investors should have an eye on enterprises' data assets information disclosure, which are in the growing and mature phases, as well as the high-tech or state-owned companies. Meanwhile, they can obtain more information about the corporate through media channels to protect their own rights and interests.

REFERENCES

[1] T. MIKOLOV ,I. SUTSKEVER, K. CHEN, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," Advances in neural information processing systems, vol. 26, 2013, pp. 3111-3119.
[2] L. VELDKAMP, "Valuing data as an asset," Review of Finance, vol. 27, 2023, pp. rfac073.
[3] K. BIRCH, D. COCHRANE, and C. WARD, "Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech," Big Data & Society, vol. 8, 2021, pp. 20539517211017308.
[4] G. CALIC and M. GHASEMAGHAEI, "Big data for social benefits: Innovation as a mediator of the relationship between big data and corporate social performance," Journal of Business Research, vol. 131, 2021, pp. 391-401.
[5] Z. M. Yuan, X. Yu, and M. Li, "Data Asset Information Disclosure, Institutional Investor Heterogeneity and Enterprise value," Modern Finance and Economics-Journal of Tianjin University of Finance and Economics, vol. 42, 2022, pp. 32-47.
[6] L. WU, L. HITT, and B. LOU, "Data analytics, innovation, and firm productivity," Management Science, vol. 66, 2020, pp. 2017-2039.
[7] C. ACCIARINI, F. CAPPA, P. BOCCARDELLI, and R. Oriani, "How can organizations leverage big data to innovate their business models? A systematic literature review," Technovation, vol. 123, 2023, pp. 102713.
[8] Y. R. Hu and X. C. Xu, "Research on the Statistical Measurement of Business Data Assets," Statistical Research, vol. 39, 2022, pp. 3-18.
[9] R. MORCK, B. YEUNG, and W. YU, "The information content of stock markets: why do emerging markets have synchronous stock price movements?" Journal of Financial Economics, vol. 58, 2000, pp. 215-260.
[10] S. ZHENG, X. ZHANG, and H. WANG, "Green credit policy and the stock price synchronicity of heavily polluting enterprises," Economic Analysis and Policy, vol. 77, 2023, pp. 251-264.
[11] M. KACPERCZYK, S. SUNDARESAN, and T. WANG, "Do foreign institutional investors improve price efficiency?" The Review of Financial Studies, vol. 34, 2021, pp. 1317-1367.
[12] T. L. DANG, M. DANG, L. HOANG, L. Nguyen, and H. L. Phan, "Media coverage and stock price synchronicity" International Review of Financial Analysis, vol. 67, 2020, pp. 101430.
[13] Z. BARKA, R. BENKRAIEM, T. HAMZA, and F. Lakhal, "How does family control affect stock price synchronicity?" Finance Research Letters, vol. 49, 2022, pp. 103092.
[14] H. REN, L. YE, and S. ZHENG, "Share repurchase and capital market pricing efficiency," Finance Research Letters, vol. 60, 2024, pp. 104904.
[15] X. GAN and P. HU, "Managerial Ability and Stock Price Synchronicity," Finance Research Letters, vol. 59, 2023, pp. 104124.
[16] I. GOLDSTEIN and L. YANG, "Good disclosure, bad disclosure," Journal of Financial Economics, vol. 131, 2019, pp. 118-138.
[17] K. D. WEST, "Dividend innovations and stock price volatility," Econometrica: Journal of the Econometric Society, vol. 56, 1988, pp. 37-61.
[18] A. ATMAZ and S. BASAK, "Belief dispersion in the stock market," The Journal of Finance, vol. 73, 2018, pp. 1225-1279.
[19] J. HU, Q. ZOU, and Q. YIN, "Research on the effect of ESG performance on stock price synchronicity: Empirical evidence from China's capital markets," Finance Research Letters, vol. 20, 2023, pp. 103847.

[20] W. CAI, J. LU, and Y. ZHAO, "Do corporate site visits impact idiosyncratic volatility? Evidence from China," Research in International Business and Finance, vol. 66, 2023, pp. 101992.
[21] Y. CHEN, B. KELLY, and W. WU, "Sophisticated investors and market efficiency: Evidence from a natural experiment," Journal of Financial Economics, vol. 138, 2020, pp. 316-341.
[22] N. Hu, F. J. Xue, and H. N. Wang, "Does Managerial Myopia Affect Long-term Investment? Based on Text Analysis and Machine Learning," Journal of Management World, vol. 37, 2021, pp. 139-156+11+19-21.
[23] Y. LECUN, Y. BENGIO, and G. HINTON, "Deep learning," nature, vol. 521, 2015, pp. 436-444.
[24] A. DURNEV, R. MORCK, B. YEUNG, and P. Zarowin, "Does greater firm ‐ specific return variation mean more or less informed stock pricing?" Journal of Accounting research, vol. 41, 2003, pp. 797-836.
[25] S. Lin and Y. Y. Ge, "Economic Policy Correlation Index, Corporate Governance and Investment Efficiency," Modern Economic Research, vol. 42, 2023, pp. 63-76.
[26] N. XU, K. C. CHAN, X. JIANG, and Z. Yi, "Do star analysts know more firm-specific information? Evidence from China," Journal of Banking & Finance, vol. 37, 2013, pp. 89-102.
[27] V. DICKINSON, "Cash flow patterns as a proxy for firm life cycle," The accounting review, vol. 86, 2011, pp. 1969-1994.
[28] C. G. BRUSH and R. CHAGANTI, "Cooperative Strategies in Non-High-Tech New Ventures: An Exploratory Study," Professor of Management Policy, Boston University Associate Professor of Management, Rider University, vol. 21, 1996, pp. 37-54.
[29] J. GAN, Y. GUO, and C. XU, "Decentralized privatization and change of control rights in China," The Review of Financial Studies, vol. 31, 2018, pp. 3854-3894.