Asia Pacific Science Press

Asia Pacific Economic and Management Review
ISSN(O):3005-9275
ISSN(P):3005-9267
Vol. 2 No.5 (2025)

The Effect of Readability of Listed Company Prospectuses on IPO Breaks: Evidence from the Shanghai Stock Exchange in China

Ruining Li[1]*, Jiangjiao Duan[2]

School of Management, University of Shanghai Science and Technology, Shanghai, 200000, China

Corresponding author: Ruining Li, 12295827002@163.com

Abstract: Using text analysis methods, this study analyzed the prospectuses of A-share listed companies on the Shanghai Stock Exchange that went public between 2014 and 2024, retrieved from the SSE website. By constructing readability metrics through an indicator-based approach, the relationship between the readability of Chinese text and the occurrence of initial public offering breaks on the first trading day was examined. The research found that, after controlling for industry and time fixed effects, there was a negative correlation between the readability of listed companies' prospectuses and the incidence of IPO breaks. This relationship was mediated through underwriter behavior. After a series of robustness checks, the above conclusion remained valid. Further research revealed that the degree of negative correlation between prospectus readability and IPO price breaks differed before and after the implementation of the registration-based IPO system, as well as when classified by the nature of listed enterprises. This finding provides a new perspective for research on factors influencing IPO pricing, which is beneficial for regulatory agencies to further standardize the information disclosure content of listed companies' prospectuses and also conducive to the development of capital markets.

Keywords: The Readability of Prospectus; IPO Break; Information Asymmetry

Published: Sept 15, 2025

DOI: https://doi.org/10.62177/apemr.v2i5.595

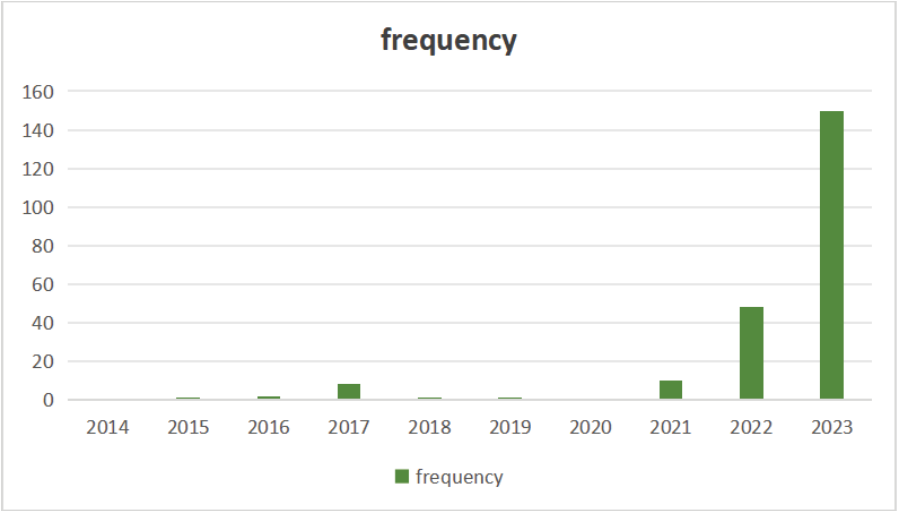
1.Introduction

The IPO system imparts vitality and resilience to the capital market, facilitating its ability to better serve the real economy. Under the backdrop of the comprehensive registration-based IPO reform, the adoption of market-oriented inquiry and underwriting mechanisms has resulted in greater freedom in the pricing of new shares, leading to prevalent phenomena such as IPO breaks. After a stock is issued, its price fluctuations are influenced by factors such as the market supply-demand relationship, investor sentiment, and company-related information. If the issuance price in the primary market is too high, or if the trading price in the secondary market is too low, it can lead to the stock's price falling below its issuance price, known as an IPO break. Figure 1 introduces the IPO breaks on the first day of enterprises listed on the Shanghai Stock Exchange.

Under the registration-based IPO system, the primary channel for information disclosure by listed companies during the IPO period is the prospectus. A prospectus is a document issued by a joint-stock company when offering shares, which discloses relevant matters related to the issuance and serves as an invitation to non-specific investors to purchase or sell shares. On the one hand, the prospectus is an important reference for underwriters in IPO pricing. If the IPO price is set too high and

the stock price subsequently returns to a reasonable level, it may lead to a situation where the stock price falls below the offering price. On the other hand, the prospectus is also a significant source of information for investors to understand the operating conditions of listed companies. If the readability of the prospectus is poor, investors may not be able to fully obtain the information disclosed by the listed company(Yu, 2022), leading to an increase in information asymmetry(Arnold et al,, 2010), which ultimately affects the price in the secondary stock market and may result in IPO breaks.

Figure1 the IPO breaks on the first day of enterprises on the SSE



Source: iFinD database

When investors obtain information about a company through a prospectus, they may face the issue of information asymmetry(Hanley and Hoberg, 2012). Complex information disclosure can impair investors' judgment, exclude them from the stock market, reduce investor participation, and gradually widen the gap among stock market participants (Tan et al., 2015). The increased cost and difficulty of obtaining information exacerbate the degree of information asymmetry in the IPO market (Grossman and Stiglitz，1980；Bloomfield，2002; Asay et al., 2018). Enhancing the readability of prospectuses can help mitigate information asymmetry. Fazzari et al. (1988) argue that problems such as information asymmetry exist in the market, rendering capital markets imperfect. In the process of stock issuance, the readability of the prospectus affects the degree of information asymmetry, thereby influencing IPO pricing（Li et al .，2018）. In China's stock market, retail investors constitute a large proportion. Consequently, an increase in information asymmetry raises their cost of information processing（Dyck et al., 2008）, leading to a decrease in demand for the stock. This causes a decline in prices in the secondary market and increases the ratio of IPO breaks. Therefore, this paper examines the impact of prospectus readability on IPO breaks based on information asymmetry.

Since the implementation of the registration-based IPO system, the number of IPO breaks in China's stock market has gradually increased. However, existing literature on pricing issues in the stock market has primarily focused on underpricing（Liu et al.，2011；Dan Bernhardt et al.，2022; Laura et al., 2023）, with limited studies on IPO breaks. Since the implementation of the registration-based IPO system in 2019, the number of IPO breaks has increased significantly in the STAR Market. Therefore, research on stock pricing cannot be limited solely to underpricing. The research in this paper can effectively fill the gap in this field.

Specifically, this paper adopts the method of text analysis, taking the prospectuses of A-share listed companies on the Shanghai Stock Exchange from 2014 to 2024 as the analysis object, and using the index method as a measure of readability to investigate the relationship between the readability of Chinese financial texts and the break of stocks. The higher the readability of the textual information in the prospectus, the lower the likelihood of IPO breaks. This conclusion remains valid after robustness testing. Furthermore, this paper also concludes that the relationship between the readability of a listed company's prospectus and IPO breaks is mediated through the behavior of underwriters. Companies with higher readability of prospectuses tend to have more accurate stock pricing after listing, which more truly reflects their intrinsic value. This is

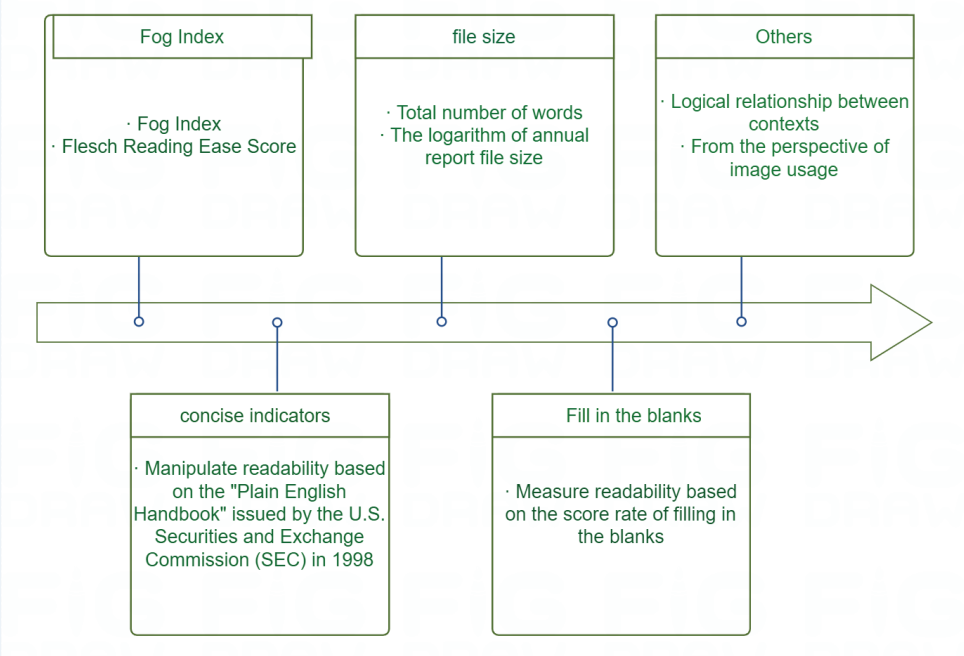
also conducive to the full play of the role of direct financing in the capital market, improving the efficiency of capital use, and leading to better performance of listed companies and less break ratio of stocks.

The main contributions of this paper are embodied in the following aspects: Firstly, our model delineates the transmission process through which the readability of prospectuses affects IPO breaks, and constructs readability indicators using both textual and metric methods, applying text mining techniques to the construction of these indicators. Secondly, for the first time, we identify the readability of prospectuses as a major factor leading to IPO breaks in the stock market, analyzing the impact of different actors on stock prices from the perspective of information asymmetry. Finally, we conduct regression analysis using the Logistic model, which serves as a valuable supplement to research on China's stock market, which is still in the stage of emerging capital markets.

2.Literature review and hypothesis

So far, a set of more established financial readability measures has been introduced, such as the Fog Index, Flesch Kincaid Indices, and Flesch Reading Ease, among others(Li, 2008; Biddle et al，2009；Lehavy et al，2011). Li (2008) was the first to apply the Misty Index to measure the readability of listed companies' annual reports. Dale and Chall (1949) defined readability from the perspective of communication effectiveness. They proposed that readability refers to the sum of all interacting factors in a given text that affect whether readers can read, understand, and be interested in it at an optimal speed. Klare et al. (1955), on the other hand, defined text readability as the ease or difficulty of reading a text due to its writing style. Baker and Kare (1992) used the complexity and length of the text to measure the readability of annual reports, and found that annual reports are very difficult for most investors to understand, requiring people with a college education or above to comprehend their content. Loughran and McDonald (2014) argued that in the composition of the Fog Index, the average sentence length is difficult to measure, and the definition of complex vocabulary does not conform to the characteristics of financial texts, so this index cannot effectively measure the readability of financial texts. Lingli Yu (2020) finally introduced a model tailored specifically for Chinese readability, potentially enhancing the comprehension of information disclosure in the Chinese market. Haiyu Yan(2024) utilizes AR indicator to explain the readability of MD&A. Figure 2 introduces the methods for measuring text readability up to now.

Figure2 the methods for measuring text readability up to now



The above discussion primarily focuses on the measurement of annual report readability. Many scholars have also explored the impact of annual report readability, with Management's Discussion and Analysis (MD&A) and financial statement notes being widely used in readability studies. Li (2008) investigated the relationship between the readability of MD&A and corporate earnings. Rjiba (2021) believes that when the disclosure tone is more negative or ambiguous, the impact of annual

report complexity on equity costs is greater. Minkwan Ahn (2023) has found that sustainability reports with higher readability tend to provide more company-specific information in the financial market.

It can be observed that the readability of information disclosure content can be influenced by the management of listed companies. If the management obscures information disclosure content to cover up poor business performance (Lo et al, 2017), then the readability of the company's information disclosure content will decrease(MerklDavies at el.，2011), increasing the degree of information asymmetry between listed companies and investors (Rameezdeen and Rajapakse, 2007). The relationship between the readability of annual reports and the profitability of listed companies has become a research direction for many scholars（Courtis，1998）, aiming to analyze whether there is any potential for manipulation by management. Subramanian et al. (1993) examined that firms with superior operating results tend to have more readable annual reports. A study conducted by Bushee et al. (2018) concluded that when firms are performing well, managers tend to be more proactive in disclosing information. Bai et al. (2019) concluded that in English-speaking countries, listed companies with less readable annual reports experience delayed stock price movements, leading to a higher risk of future stock price crashes. Hasan (2020) argued that the readability of disclosures in financial reports is significantly and positively correlated with the managerial competence of executives of listed firms.

The phenomenon of IPO breaks was first discovered by Stigler (1964) when studying industrial enterprises. Subsequently, scholars researching stock market performance successively uncovered this abnormal phenomenon and attributed it to a relatively high degree of uncertainty. Benveniste (1989), referencing the US book-building pricing mechanism, proposed an optimal pricing mechanism selection model, and argued that the main reason for IPO breaks is the information asymmetry between securities underwriters and investors.

Therefore, the rationale for IPO breaks lies in the information asymmetry existing in the market, and the readability of the prospectus significantly influences this information asymmetry. Here, we propose the first hypothesis:

H1: The higher the readability of the prospectus of a listed company, the lower the occurrence rate of IPO breaks on the first day.

As China's capital market gradually becomes more market-oriented, the number of IPO breaks has increased. Currently, research on "IPO breaks" phenomenon is still not thorough enough, and this study aims to fill that gap. Based on early research on IPO breaks, Stoll and Curley (1970) examined the price trends of 205 U.S. stocks and found that they exhibited a long-term downward trend. Sung Gyun Mun and Jang Soo Cheong (2019) concluded that restaurant businesses with lower profitability would bear greater financial burdens, ultimately leading to IPO breaks until they were delisted.

The phenomenon of IPO breaks is due to investors being overly optimistic about the stock's valuation and unreasonable pricing at the time of issuance, resulting from an excessively high pricing of the stock (Derrien and Kent, 2003). Khurshed (2009) found that issuers of stocks tend to increase the offering price in order to raise more funds, and as a result, they may release related false information, which leads to IPO breaks in the later stages. Kevin K. (2016) conducted a study on IPOs and found that the conducts of underwriters significantly influenced pricing decisions. Therefore, it can be concluded that overpricing by underwriters is one of the main factors contributing to the phenomenon of IPO breaks for listed companies. The prospectus is the primary channel for underwriters to obtain information about a company's operations. When the degree of information asymmetry between the company and the market increases, the pricing set by the underwriter may deviate from the true value of the stock. Motivated by the desire to earn higher underwriting fees, underwriters have an incentive to increase the IPO price (Cook et al.,2006), ultimately resulting in an overpriced IPO and a break of the issue price.

Therefore, it can be concluded that overpricing by underwriters is one of the important factors contributing to the phenomenon of IPO breaks for listed companies. Based on this, we propose the second hypothesis:

H2a: The readability of the prospectus can influence the incidence rate of IPO breaks by affecting the behavior of underwriters in the primary market.

When investors decide whether to purchase a stock, they first refer to the prospectus of the stock. If the prospectus is highly readable, investors can more easily obtain relevant information about the company, thereby making a judgment on whether the stock is worth investing in(Boulton et al ., 2011, Hong et al., 2014). Conversely, if the prospectus is less readable,

investors may have limited access to officially disclosed information (Yu, 2022) or limited access to information about the listed company through other channels(Tetlock, 2007). This may lead investors to be cautious in purchasing the stock(Daniel, Hirshleifer, & Subrahmanyam, 1998 ; Hirshleifer, 2001), reducing the demand for it (Zhang ,2006a; Zhang ,2006b), and ultimately causing the stock price to fall.

Therefore, the sentiment of secondary market investors is also one of the important factors influencing the underpricing of listed companies' stocks. Based on this, we propose the third hypothesis:

H2b: The readability of the prospectus will influence the incidence rate of IPO breaks by affecting the sentiment of secondary market investors.

3.Data and methodology

3.1 Sample acquisition and screening

This paper selects relevant data and prospectuses of all listed companies on the Main Board and STAR Market of the Shanghai Stock Exchange from 2014 to 2024 as the research objects. The financial data of listed companies and IPO market data used in the research are sourced from the CSMAR database and iFinD database. The prospectuses used in this paper were batch-downloaded from the designated sections of the Main Board and STAR Market of the Shanghai Stock Exchange using a Python web crawler program. The initial sample comprises 1,358 listed companies.

The initial sample underwent the following processing: all ST and *ST listed companies were excluded; due to the specificity of financial listed companies, they were also excluded; additionally, 80 companies with prospectuses that could not be correctly parsed due to encryption or coding issues were removed, resulting in a final set of 1,232 observations. This paper utilizes Python to conduct statistics on these prospectuses, including information such as word count, average sentence length, whether they contain a directory, and total number of pages.

3.2 Definition of variables

3.2.1 Dependent variable

In empirical analysis, Y can be defined as a binary variable indicating whether a stock has experienced IPO break, with Y=1 representing that the stock has suffered from an IPO break and Y=0 indicating that the stock has not experienced an IPO break, as shown in Equation (1).

$$Y = \mathrm{Logit}(BD) = Ln(\frac{BD}{1-BD}) \tag{1}$$

The explained variable, denoted as BD, is calculated based on the logit model and represents the odds of a newly listed stock experiencing an IPO break, as shown in Equation (2).

$$BD = \frac{\exp(\bar{y})}{1+\exp(\bar{y})} \tag{2}$$

3.2.2 Independent variables

The formula-based readability metrics consist of two calculations. The first is based on the readability measurement model proposed by Senter and Smith (1967). This metric employs two of the most fundamental and crucial indicators for measurement. Here, the AR metric is inversely correlated with readability; a higher value indicates that the text is more difficult to understand and less readable. Therefore, for consistency in the interpretation of the specification, the negative value of this indicator is used to represent the readability of the text. The construction of this indicator is shown in (3).

AR=-[4.17*(Vocabulary length/Word count)+0.39*(Word count/Sentence Count)-21.43] (3)

Secondly, we utilize the Flesch-Kincaid (FK) formula to assess the readability of a text based on the reading ability required by the intended audience. This indicator measures the ease of reading, where a higher score indicates lower requirements for reading ability. To ensure consistency in interpretation, we multiply the FK score by (-1), so that a higher absolute value corresponds to lower readability. The construction of this modified indicator is illustrated in (4).

FK = -[0.39*(Word count/Sentence Count)+11.8*(Vocabulary length/Word count)-15.9] (4)

3.2.3 Control variables

The control variables selected for this study comprise the logarithmic value of total sponsorship fees(BJ), the logarithmic value of assets raised(RM), the industry price-earnings ratio on the listing date(PE), the logarithmic value of pre-listing company's total assets(TA), return on assets(ROA), the pre-listing company's gearing ratio(AD), the turnover rate on the first day of listing(TR), a dummy variable indicating whether the company is state-owned(XZ), and another dummy variable indicating whether the company is ranked in the top ten of equity underwriters(SY). Additionally,control for the year (Year) and industry (Industry) effects.

Table 1 Variable definition table.

Variable type	Variable name	Variable symbol	Variable definition
Dependent variables	the ratio of IPO breaks occurrences	BD	BD=exp($\bar{y}$)/(1+exp($\bar{y}$))
Independent variables	Readability measures1	AR	-ln[4.17*(Vocabulary length/Word count)+0.39*(Word count/Sentence Count)-21.43]
	Readability measures2	FK	-ln[0.39*(Word count/Sentence Count)+11.8*(Vocabulary length/Word count)-15.9]
Control variable	total sponsorship fees	BJ	he logarithmic value of total sponsorship fees
	Total assets raised	RM	the logarithmic value of assets raised
	Price Earnings Ratio	PE	he industry price-earnings ratio on the listing date
	pre-listing company's total assets	TA	the logarithmic value of pre-listing company's total assets
	return on assets	ROA	Return on assets before the company went public
	Pre-listing gearing ratio	AD	the pre-listing company's gearing ratio
	the turnover rate	TR	the turnover rate on the first day of listing
	Nature of business	XZ	Whether it is a state-owned enterprise, 1 for yes, 0 for no
	Underwriter reputation	SY	Whether the equity underwriters are ranked in the top ten, yes 1, no 0
	Year	Year	Annual fixed effects
	industry	Industry	industry fixed effect

3.2.4 Model construction

This study employs the Logistic model for regression analysis. Given that the sample consists of only one observation per IPO company, namely the prospectus, it is not feasible to control for individual fixed effects. Instead, the study controls for industry and year effects. The baseline regression model is shown in Equation (5).

$$Logit(BD) = \ln(\frac{BD}{1-BD}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_i x_i + \varepsilon_{i,t} + Year + Industry \tag{5}$$

In this model, BD represents the incidence of IPO breaks, β_0 is the coefficient of the constant term in the regression, x_i denotes the factors influencing IPO breaks, and β_i represents the regression coefficient for the ith variable. Based on the Logistic regression model, the regression models constructed with the incidence of IPO breaks as the dependent variable are presented in Equations (6) and (7).

$$\begin{aligned} BD = {} & \beta_0 + \beta_1 AR + \beta_2 BJ + \beta_3 RM + \beta_4 PE + \beta_5 TA + \beta_6 ROA + \\ & \beta_7 AD + \beta_8 TR + \beta_9 XZ + \beta_{10} SY + Year + Industry + \varepsilon_{i,t} \end{aligned} \tag{6}$$

$$BD = \beta_0 + \beta_1 FK + \beta_2 BJ + \beta_3 RM + \beta_4 PE + \beta_5 TA + \beta_6 ROA + \beta_7 AD + \beta_8 TR + \beta_9 XZ + \beta_{10} SY + Year + Industry + \varepsilon_{i,t} \quad (7)$$

4.Results

4.1 Empirical analysis

4.1.1 Descriptive statistics

This paper utilizes web crawling technology to obtain the prospectuses of listed companies on the Shanghai Stock Exchange. After excluding all ST, *ST, and financial companies, it employs Python software to calculate the total number of pages, total number of words, total number of tables, total number of images, and total number of sentences in these prospectuses, as shown in Table 2.

Table2 Data Overview of Prospectuses

	Max	Min	Mean
Pages	2800	300	497.41
Words	713442	319	264389.83
Tables	653	142	247.70
Images	15081	0	263.61
Sentences	7587	35	2807.53

Table 3 presents the descriptive statistics of the data. The standard deviation of the AR, FK, and BD indicators for the broken group are 0.331, 0.248, and 0.136 respectively, while for the unbroken group, they are 0.340, 0.259, and 0.212, suggesting that the relatively large volatility in the readability of prospectuses from listed companies that have experienced IPO breaks explains why the mean value for the broken group is relatively low.

Table 3 Summary statistics.

	(1)	(2)		(3)		(4)		(5)	
variables	Obs	Mean		Std.		Min		Max	
Group	N	broken group	unbroken group	broken group	unbroken group	broken group	unbroken group	broken group	unbroken group
BD	1,222	0.923	0.822	0.136	0.212	0.172	0.121	1.000	1.000
AR	1,231	-3.787	-3.759	0.331	0.340	-4.451	-4.681	-2.349	-2.917
FK	1,231	-5.113	-5.092	0.248	0.259	-5.650	-5.806	-4.159	-4.480
BJ	1,231	0.318	0.791	1.297	1.949	0	0	6.215	6.397
RM	1,231	11.99	11.20	0.637	0.783	10.81	9.749	14.61	13.82
PE	320	32.55	30.66	16.06	15.01	1	3.130	75.12	71.72
TA	1,223	20.01	20.59	1.325	1.165	17.28	18.38	24.19	24.52
ROA	1,223	-0.0444	0.0860	0.327	0.0934	-1.738	-0.303	0.514	0.361
AD	1,224	54.45	36.09	38.69	25.72	0	0	211.6	98.70
TR	1,232	0.0801	0.0640	0.0430	0.0732	5.00e-05	3.00e-05	0.148	0.184
XZ	1,232	0.0698	0.119	0.256	0.324	0	0	1	1
SY	1,232	0.698	0.527	0.462	0.499	0	0	1	1

4.1.2 Correlation analysis

Given the large number of variables in this paper, there may exist strong correlations among them, which can lead to the issue of multicollinearity. To prevent this issue, a correlation analysis is conducted prior to regression. First, the correlation coefficients between variables are calculated using Stata software.As shown in Table 4, the correlation coefficients between the control variables are all below 0.4, indicating that the correlations among the control variables are weak, and there is no issue of multicollinearity.

Table 4 Correlation analysis.

Variables	BD	AR	FK	BJ	RM	PE	TA	ROA	AD	TR	XZ	SY
BD	1											
AR	-0.67***	1										
FK	-0.67***	0.98***	1									
BJ	-0.007	-0.013	-0.02	1								
RM	0.08	0.16***	0.16***	-0.09*	1							
PE	0.106	-0.07	-0.06	0.08	0.02	1						
TA	-0.34***	0.18***	0.18***	-0.01	0.37***	-0.17	1					
ROA	0.01	-0.07	-0.08	0.01	-0.19***	-0.08	0.001	1				
AD	-0.05	0.07	0.08	-0.03	0.22***	-0.04	0.20***	-0.29***	1			
TR	-0.09	-0.02	-0.01	-0.02	0.19***	0.06	-0.38***	-0.12***	0.11***	1		
XZ	-0.25***	0.08	0.08	0.004	0.17***	-0.06	0.39**	-0.04	0.08	-0.04	1	
SY	-0.15***	0.03	0.03	-0.01	0.23***	0.05	0.10**	-0.097*	0.12***	0.08	0.04	1

4.1.3 regression

Based on the regression results presented in Table 5, the overall R-squared values for the two models are 0.821 and 0.813, respectively, indicating that these models have good fit effects. Furthermore, the P-values for both models are 0.000, suggesting that they are statistically significant overall, and the regression results are reliable. When the readability indicator for companies is AR, the AR indicator passes the test at the 1% significance level, and the regression coefficient is negative, which aligns with the expected hypothesis. The higher the readability of a listed company's prospectus, the lower the proportion of IPO breaks.

It should be noted that the readability indicator is designed such that a lower score indicates higher readability. For the convenience of our research, the original indicator was multiplied by (-1). Therefore, when the text readability is higher, the score of the readability indicator is also higher. This conclusion is consistent with the expected hypothesis H1, which states that the higher the readability of a listed company's prospectus, the lower the incidence of IPO breaks.

Table 5 Regression results

	(1)	(2)
VARIABLES	BD	BD
AR	-0.421***	
	(-22.65)	
FK		-0.541***
		(-21.94)
BJ	-0.002	-0.002

VARIABLES	(1) BD	(2) BD
	(-0.52)	(-0.52)
RM	0.115***	0.115***
	(14.71)	(14.40)
PE	-0.000	-0.000
	(-0.49)	(-0.44)
TA	-0.070***	-0.071***
	(-13.40)	(-13.21)
ROA	-0.011	-0.009
	(-0.47)	(-0.40)
AD	0.000	0.000
	(0.89)	(0.88)
TR	-1.237***	-1.245***
	(-11.84)	(-11.67)
XZ	-0.054***	-0.053***
	(-3.59)	(-3.47)
SY	-0.049***	-0.050***
	(-4.53)	(-4.44)
Constant	-0.559***	-1.731***
	(-3.26)	(-8.36)
Observations	320	320
R-squared	0.821	0.813
Industry FE	YES	YES
Year FE	YES	YES

Note(s): t-statistics are in parentheses. ***, **, and * indicate significance at the 1 %, 5 %, and 10 % levels, respectively.
Source(s): Authors' work

4.2 Mediation effect analysis

4.2.1 conduct of underwriters

The behavior of underwriters can be represented by taking the logarithm of the total underwriting fee (CX) (Lewellen, 2010). IPO pricing behavior can be viewed as a form of information transmission, where underwriters' judgments on IPO prices reflect their assessments of the companies. The prospectus of a company serves as the primary basis for underwriters to assess the operating conditions of a company that is soon to be listed. Therefore, the readability of the prospectus will affect the degree to which underwriters can obtain information, further influencing the IPO price.

If the readability of a company's prospectus is low, underwriters will have limited access to relevant information about the company, leading to deviations in IPO pricing from its accurate value. In such cases, underwriters may intentionally raise the offering price for their own benefit. Once the stock is traded in the secondary market and its price returns to a rational level, an IPO break will occur. Therefore, we consider the behavior of underwriters as a transmission mechanism through which the readability of the prospectus affects IPO breaks. Improving the readability of the prospectus can reduce the degree of information asymmetry, enabling underwriters to price IPOs more accurately and thus lowering the incidence of IPO breaks.

The mechanism test employs a two-step approach, with the model setup as shown in Equations (8) and (9).

$$CX = e_1 + cAR\ （FK）+ \beta_1 Controls + Industry + Year + \varepsilon_{it} \quad (8)$$

$$BD = e_2 + aCX + \alpha_t Controls + Industry + Year + \varepsilon_{t} \quad (9)$$

4.2.2 investor sentiment

Investor sentiment can be represented by the turnover rate (TR). On the one hand, if the readability of a listed company's prospectus is low, investors have limited access to specific company information, which can lead to decreased investment in that stock. On the other hand, if online media obtain limited information from the prospectus (Guldiken et al, 2017) or even convey inaccurate information to investors (Arnold et al, 2010), this can also result in reduced demand for the stock (Wang and Wu, 2015), causing a price decline in the secondary market. Therefore, investor sentiment in the secondary market can serve as a transmission mechanism for the impact of prospectus readability on IPO breaks. Enhancing the readability of the prospectus reduces the degree of information asymmetry in the market, allowing secondary investors to access more information about the relevant company and increasing their demand for the listed company's stock, as illustrated in Equations (10) and (11).

$$TR = e_1 + cAR\ （FK）+ \beta_1 Controls + Industry + Year + \varepsilon_{it} \quad (10)$$

$$BD = e_2 + aTR + \alpha_t Controls + Industry + Year + \varepsilon_{t} \quad (11)$$

We can represent this transmission mechanism in Figure 3.

Figure 3 transmission mechanism

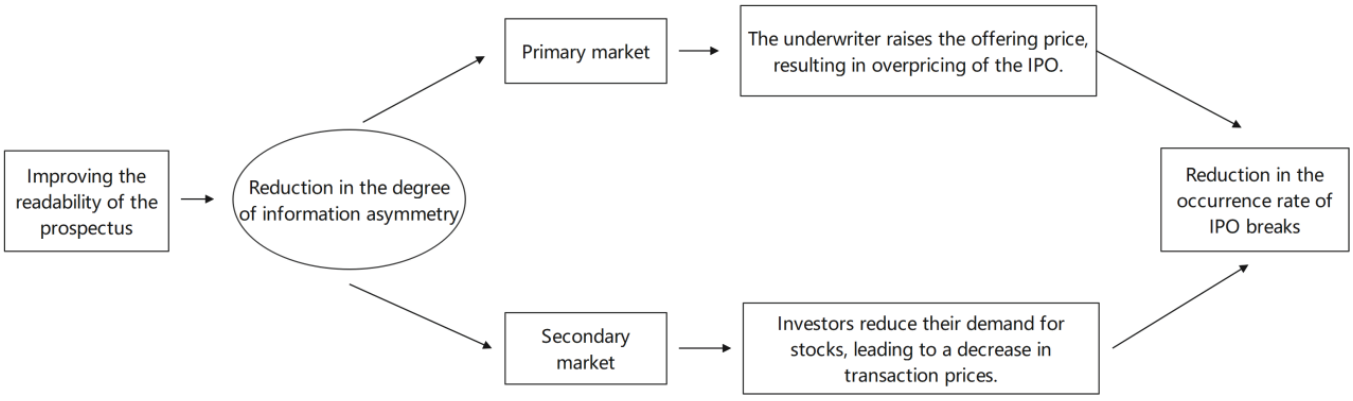


Table6 Mediation effect results

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	**CX**	**CX**	**BD**	**TR**	**TR**	**BD**
AR	-0.061**			-0.015***		
	(-2.33)			(-3.22)		
FK		-0.078**			-0.019***	
		(-2.27)			(-3.22)	
CX			0.053***			
			(2.72)			
TR						-0.358***
						(-3.18)
ROA	-0.049	-0.050	0.077	-0.014	-0.014	0.070
	(-0.65)	(-0.66)	(1.51)	(-1.04)	(-1.05)	(1.38)
AD	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000*
	(-0.52)	(-0.53)	(-1.53)	(-0.98)	(-0.99)	(-1.69)

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	CX	CX	BD	TR	TR	BD
RM	0.577***	0.577***	0.011	-0.002	-0.002	0.040***
	(45.68)	(45.66)	(0.78)	(-0.92)	(-0.91)	(4.78)
SY	0.028	0.028	-0.071***	0.000	0.000	-0.069***
	(1.57)	(1.57)	(-5.88)	(0.02)	(0.02)	(-5.76)
Constant	1.509***	1.344***	0.135	-0.045	-0.087**	0.235**
	(7.42)	(5.29)	(1.13)	(-1.27)	(-1.99)	(2.06)
Observations	1,222	1,222	1,221	1,222	1,222	1,221
R-squared	0.763	0.762	0.101	0.498	0.498	0.103
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES

Note(s): t-statistics are in parentheses. ***, **, and * indicate significance at the 1 %, 5 %, and 10 % levels, respectively.
Source(s): Authors' work

As stated in columns (1), (2), and (3) of Table 6, underwriter behavior can induce a negative correlation between the readability of the prospectus and IPO breaks. However, as derived from columns (4), (5), and (6), investor sentiment does not support this negative correlation. Therefore, with Hypothesis H2a confirmed, we can draw the following conclusion from the regression results: Through a two-step analysis, we can infer that an increase in the readability of the prospectus reduces the behavior of underwriters deliberately raising the issue price to obtain higher underwriting fees. This alleviates the degree of information asymmetry in the market, leading to more reasonable IPO issue prices and thus reducing the incidence of IPO breaks.

4.2.3 Bootstrap Test

Based on the results presented in Table 7, we can observe that the confidence intervals are all negative, with no numbers falling between positive and negative, or zero within these intervals. Therefore, we can conclude that there is a mediation effect: the readability of listing companies' prospectuses influences the occurrence of IPO breaks through the mechanism of underwriters' behavior.

Table 7 Bootstrap Test

	coefficient	std. err.	z	P>|z|	[95% conf. interval]	
_bs_1	0.0140258	0.0038252	3.67	0.000	0.0065287	0.021523
_bs_2	-0.4519545	0.0137623	-32.84	0.000	-0.478928	-0.424981
_bs_3	-0.4379287	0.0140974	-31.06	0.000	-0.465559	-0.4102983

4.3 Robustness analysis

Robustness test can be divided into four parts.

4.3.1 Only the last four years of sample data were retained

The STAR Market has experienced a large quantities breaks in IPOs since 2021 and has become the most affected area of A-share IPOs, shattering the perception that IPOs are always profitable. We can also clearly observe from Figure 1 that the occurrence rate of IPO breaks in the past four years is significantly higher than that in previous years.Therefore, only data from after 2021 is retained for the regression analysis, and the results are presented in Table 9 column(1) and column(2). As a result, we can draw a similar conclusion to our previous analysis. Specifically, the readability measures AR and FK coefficients have increased, indicating that in the last four years, the readability of financial statements has had a more pronounced impact on the incidence rate of IPO breaks for listed companies.

4.3.2 Add other control variables

We have incorporated the logarithm of total underwriting fees(LnCX) and net cash flow per share(LL) into the model to account for the influence of prospectus readability on the probability of breaks for listed companies. Based on the regression results, we can conclude that the model fits well, and after controlling for industry and time fixed effects, the readability of listed companies' prospectuses exhibits a negative correlation with the incidence of IPO breaks among listed companies. The results are presented in column(3) and column(4).

4.3.3 Substitution of explanatory variables

Readability can also be measured using a text-based method, which involves counting six standardized indicators: the presence or absence of a table of contents, the page number (standardized), the total number of words (standardized), the average sentence length (standardized and isotropic), the number of sentences (standardized and multiplied by 10), and the number of words (standardized). All text analysis is conducted within a Python environment, and the specific steps include: The first step is the preprocessing of prospectuses, converting all PDF files into TXT format. The second step involves obtaining data such as the total number of words and sentences in each prospectus, and calculating the average sentence length. The third step is to construct readability indicators by summing the standardized data results to calculate the final readability indicator value. The results are shown in Table 8.

Table 8 Textual Readability Metrics

Transaction Code	contents	page number	number of words	number of sentences	average sentence length	readability
600025	1	0.1972	0.4050	0.1277	0.9997	2.7296
600032	1	0.2244	0.6520	0.1392	0.9996	3.0152
.....	.....	.....	.....	.....	.....	.....
689009	1	0.038	0.6154	0.1245	0.9997	2.7776

According to column(5), we can conclude that, at the 10% significance level, a higher readability of a listed company's prospectus is associated with a lower incidence of break of the company's stock.

4.3.4 PSM test

The PSM test results are shown in column(6) and column(7), in order to avoid the possibility of endogenous disturbances in the model, the readability metrics AR as well as the median FK are used as benchmarks in this study. The data table greater than the median is labeled as 1 otherwise 0, which is indicated by the readability AR1 and FK1. Meanwhile, 1:1 nearest-neighbor matching is performed without replacement, and the samples with successful matches are returned again, and we can still conclude that after controlling for the fixed effects of industry and time, the readability of listed firms' prospectuses is responsibly correlated with the incidence of IPO breaks.

Table 9 Robustness analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	BD	BD	BD	BD	BD	BD	BD
readability					-0.114*		
					(-1.89)		
AR	-0.425***		-0.423***				
	(-22.67)		(-22.43)				
AR1						-0.176***	
						(-13.02)	
FK		-0.547***		-0.544***			

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	**BD**	**BD**	**BD**	**BD**	**BD**	**BD**	**BD**
		(-21.93)		(-21.72)			
FK1							-0.179***
							(-13.57)
BJ	-0.001	-0.001	-0.002	-0.002	-0.007	-0.001	-0.003
	(-0.39)	(-0.40)	(-0.53)	(-0.53)	(-1.28)	(-0.22)	(-0.64)
RM	0.118***	0.117***	0.107***	0.106***	0.114***	0.107***	0.115***
	(14.68)	(14.34)	(7.51)	(7.31)	(8.34)	(10.74)	(10.79)
PE	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
	(-0.34)	(-0.30)	(-0.40)	(-0.36)	(0.61)	(0.63)	(0.04)
TA	-0.071***	-0.071***	-0.070***	-0.070***	-0.084***	-0.076***	-0.082***
	(-13.28)	(-13.08)	(-12.54)	(-12.34)	(-8.68)	(-11.23)	(-11.49)
ROA	-0.011	-0.009	-0.013	-0.012	0.049	-0.011	-0.010
	(-0.47)	(-0.40)	(-0.57)	(-0.50)	(0.92)	(-0.32)	(-0.32)
AD	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.71)	(0.70)	(0.90)	(0.88)	(0.70)	(1.24)	(0.51)
TR	-1.245***	-1.254***	-1.243***	-1.251***	-1.226***	-1.215***	-1.258***
	(-11.77)	(-11.60)	(-11.87)	(-11.70)	(-6.89)	(-9.02)	(-9.04)
XZ	-0.056***	-0.055***	-0.051***	-0.050***	-0.042*	-0.052***	-0.030
	(-3.67)	(-3.54)	(-3.37)	(-3.24)	(-1.66)	(-2.89)	(-1.61)
SY	-0.051***	-0.051***	-0.049***	-0.049***	-0.058***	-0.044***	-0.045***
	(-4.60)	(-4.50)	(-4.43)	(-4.34)	(-3.16)	(-3.04)	(-3.14)
LnCX			0.012	0.013			
			(0.71)	(0.74)			
LL			0.004	0.004			
			(0.90)	(0.89)			
Constant	-0.688***	-1.870***	-0.594***	-1.773***	1.542***	1.100***	1.131***
	(-4.36)	(-9.50)	(-3.38)	(-8.40)	(5.46)	(6.61)	(6.65)
Observations	317	317	320	320	320	340	342
R-squared	0.822	0.814	0.822	0.814	0.504	0.676	0.679
Industry FE	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES

Note(s): t-statistics are in parentheses. ***, **, and * indicate significance at the 1 %, 5 %, and 10 % levels, respectively.
Source(s): Authors' work

5.Conclusion

This empirical study examines the relationship between the readability of listing prospectuses and the occurrence of IPO breaks among listed companies. Through Logistic model analysis, it is found that there is a negative correlation between the

readability of listing prospectuses and the incidence rate of IPO breaks. Taking the listing prospectuses of companies listed from 2014 to 2024 as the analysis sample, we employed text mining methods to construct readability indicators and verified the hypothesis through empirical analysis. The theoretical and practical significance of this paper are as follows:

Firstly, the model and empirical results of this paper indicate that the readability of prospectuses affects the incidence rate of IPO breaks in the stock market. When underwriters set IPO prices, they refer to the listing prospectus. An increase in readability can reduce the deviation of IPO prices from their intrinsic value, thereby decreasing the occurrence of IPO breaks and contributing to the stability of the stock market. Therefore, the higher the readability of the listing prospectus, the lower the incidence rate of IPO breaks. This conclusion demonstrates that high readability plays a crucial role in reducing the factors that contribute to instability in the stock market. This paper not only provides a reference for research on IPO breaks among listed companies in emerging capital markets but also offers new insights for promoting the stable development of the stock market.

Secondly, this study utilizes Python for text mining to construct readability indicators based on textual methods, enriching the research on the measurability of readability in financial information disclosure in Chinese text. Previous studies mostly used the Fog Index to measure readability, which is suitable for English text. However, this study provides a more accurate and reasonable measurement of readability, offering a reference for readability research in Chinese text.

Lastly, for government departments, it is crucial to further standardize the information disclosure content in listing prospectuses of listed companies(Coffee,1984). Improving the readability of listing prospectuses holds dual significance. On one hand, it assists underwriters in accurately pricing IPOs for listed companies(Carter et al.,1998); on the other hand, it reduces information asymmetry between companies and investors, enhancing the transparency of the capital market. Furthermore, for investors, the reduced cost of obtaining information from the stock market attracts more participants. Enhanced market liquidity facilitates accurate pricing of listed companies' stock trading prices, reducing the occurrence of market anomalies, and ultimately promoting the stable development of the Chinese stock market.

The limitation of this paper is that it does not consider the impact of the pandemic on IPO breaks. Since the outbreak of the pandemic in 2019, the phenomenon of IPO breaks has gradually increased. Therefore, this factor should be taken into account in subsequent research. This paper focuses on the first-day breaks, while the impact of investor sentiment in the secondary market on stock trading prices is limited. Future research can extend the period of IPO breaks, which will allow for a deeper influence of investor sentiment on stock trading prices and facilitate further research.

Funding

no

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

[1] Yu, Z. (2022), "Financial report readability and accounting conservatism", Journal of Risk and Financial Management, Vol. 15 No. 10, p. 454, Doi: 10.3390/jrfm15100454.

[2] Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998), "Investor psychology and security market under- and overreactions", Journal of Finance, 53(6), 1839–1885.

[3] Hirshleifer, D. (2001), "Investor psychology and asset pricing", Journal of Finance, 56(4), 1533–1597.

[4] Zhang, X. F. (2006a), "Information uncertainty and stock returns", Journal of Finance, 61(1), 105–137.

[5] Zhang, X. F. (2006b), "Information uncertainty and analyst forecast behavior", Contemporary Accounting Research, 23(2), 565–590.

[6] Boulton, T. J.; S. B. Smart; and C. J. Zutter., (2011), "Earnings Quality and International IPO Underpricing", Accounting Review, 86, 485–505.

[7] Hong, H. A.; M. Hung; and G. J. Lobo., (2014), "The Impact of Mandatory IFRS Adoption on IPOs in Global Capital Markets." Accounting Review, 89, 1365–1397.

[8] Dale, E., and J. S. Chall., (1949), "The concept of readability", elementary English, Pages: 19-26.

[9] Tinker, M. A., G. R. Klare, (1955), "Review of Know Your Reader", Journal of Educational Psychology, 46(5), 318–319.

[10] Eugene Baker, H. and Kare, D.D., (1992), "Relationship Between Annual Report Readability and Corporate Financial Performance", Management Research News, Vol. 15 No. 1, pp. 1-4.

[11] Stigler, G. J., (1964), "Public regulation of the securities markets", Bus. Law., 19, 721.

[12] François Derrien, Kent L. Womack, (2003), "Auctions vs. Bookbuilding and the Control of Underpricing in Hot IPO Markets", The Review of Financial Studies, Volume 16, Issue 1, Pages 31–61.

[13] Marc Goergen, Arif Khurshed, Luc Renneboog, (2009), "Why are the French so different from the Germans? Underpricing of IPOs on the Euro New Markets", International Review of Law and Economics, Volume 29, Issue 3, Pages 260-271.

[14] Kevin K. Boeh, Craig Dunbar, (2016), "Underwriter deal pipeline and the pricing of IPOs", Journal of Financial Economics, Volume 120, Issue 2, Pages 383-399, ISSN 0304-405X.

[15] Fazzari S, Hubbard R G, Petersen B. (1988), "Investment, financing decisions, and tax policy", The American economic review, 78(2): 200-205.

[16] Li, R., Liu, W., Liu, Y. and Tsai, S.-B. (2018), "IPO underpricing after the (2008), financial crisis: a study of the Chinese stock markets", Sustainability, Vol. 10 No. 8, p. 2844.

[17] Dyck, A., Volchkova, N. and Zingales, L. (2008), "The corporate governance role of the media: evidence from Russia", The Journal of Finance, Vol. 63 No. 3, pp. 1093-1135.

[18] Dan Bernhardt, Kostas Koufopoulos, Giulio Trigilia,(2022), "Separating equilibria, underpricing and security design", Journal of Financial Economics, Volume 145, Issue 3, Pages 788-801, ISSN 0304-405X.

[19] Xiaoding Liu, Jay R. Ritter, (2011), "Local underwriter oligopolies and IPO underpricing", Journal of Financial Economics, Volume 102, Issue 3, Pages 579-601, ISSN 0304-405X.

[20] Laura Xiaolei Liu, Ruichang Lu, Ann E. Sherman, Yong Zhang, (2023), "IPO underpricing and limited attention: Theory and evidence", Journal of Banking & Finance, Volume 154, 106932, ISSN 0378-4266.

[21] Arnold, T., Fishe, R. P., & North, D., (2010), "The effects of ambiguous information on initial and subsequent IPO returns", Financial Management, 39(4), 1497-1519.

[22] Hanley, K. W., and G. Hoberg,(2012), "Litigation risk, strategic disclosure and the underpricing of initial public offerings", Journal of Financial Economics ,103:235–54.

[23] Hun-Tong Tan, Elaine Ying Wang, Bo Zhou, (2015), "How Does Readability Influence Investors' Judgments? Consistency of Benchmark Performance Matters", The Accounting Review. 90 (1): 371–393.

[24] Bloomfield R. J., (2008), "Discussion of annual report readability, current earnings, and earnings persistence", Journal of Accounting and Economics, 45):248 ~ 252.

[25] Grossman S. J., Stiglitz J. E., (1980), "On the impossibility of informationally efficient markets", American Economic Review, 70): 393 ~ 408.

[26] H. Scott Asay, Robert Libby, Kristina Rennekamp, (2018), "Firm performance, reporting goals, and language choices in narrative disclosures", Journal of Accounting and Economics,Volume 65, Issues 2–3, Pages 380-398.

[27] Li F. (2008), "Annual report readability, current earnings, and earnings persistence", Journal of Accounting and Economics, 45 (2/3): 221-247.

[28] Biddle,G. C., G. Hilary, and R. S. Verdi, (2009), "How Does Financial Reporting Quality Relate to Investment Efficiency?", Journal of Accounting and Economics, 48(2-3): 112-131.

[29] Lehavy, R., F. Li, and K. Merkley, (2011), "The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecasts", Accounting Review , 86(3): 1087-1115.

[30] LOUGHRAN, T. and MCDONALD, B. (2014), "Measuring Readability in Financial Disclosures", The Journal of Finance, 69: 1643-1671.

[31] Haiyu Yan, (2024), "Environmental information disclosure, earnings quality and the Readability and Emotional

Tendencies of Management Discussion and Analysis", Finance Research Letters, Volume 60,104913,ISSN1544-6123.

[32] Lingli Yu and Q. Cao and Yunhan Mou and Hongyu Du, (2020), "Chinese FOG Index: the Readability of Information Disclosure in Chinese Listed Companies", Proceedings of the 2020 5th International Conference on Mathematics and Artificial Intelligence.

[33] Hatem Rjiba, Samir Saadi, Sabri Boubaker, Xiaoya (Sara) Ding, (2021), "Annual report readability and the cost of equity capital", Journal of Corporate Finance, Volume 67, 101902, ISSN 0929-1199.

[34] Minkwan Ahn, Doowon Jung, Jeong-Taek Kim, Woo-Jong Lee, Hee-Yeon Sunwoo,(2023), "Do more readable sustainability reports provide more value-relevant information to shareholders?", Finance Research Letters,Volume 57,1 04154, ISSN 1544-6123.

[35] Lo, K., F. Ramos, and R.Rogo, (2017), "Earnings Management and Annual Report Readability", Journal of Accounting and Economics, 63 (1): 1 - 25.

[36] Merkl-Davies D M, Brennan N M, McLeay S J., (2011), "Impression management and retrospective sense-making in corporate narratives", Accounting, Auditing & Accountability Journal, (2011), 24(3): 315-344.

[37] Rameezdeen, R., and C. Rajapakse. 2007. "Contract Interpretation: The Impact of Readability", Construction Management and Economics, 25(7) : 729-737.

[38] Courtis, J.K., (1998), "Annual report readability variability: tests of the obfuscation hypothesis", Account. Audit. Account, J. 11 (4), 459–472.

[39] Subramanian, R., Insley, R. G., & Blackwell, R. D. (1993), "Performance and Readability: A Comparison of Annual Reports of Profitable and Unprofitable Corporations", The Journal of Business Communication, 30(1), 49-61.

[40] BUSHEE, B.J., GOW, I.D. and TAYLOR, D.J., (2018), "Linguistic Complexity in Firm Disclosures: Obfuscation or Information?", Journal of Accounting Research, 56: 85-121.

[41] Mostafa Monzur Hasan, (2020), "Readability of Narrative Disclosures in 10-K Reports: Does Managerial Ability Matter?", European Accounting Review, 29:1, 147-168.

[42] Minkwan Ahn, Doowon Jung, Jeong-Taek Kim, Woo-Jong Lee, Hee-Yeon Sunwoo,(2023), "Do more readable sustainability reports provide more value-relevant information to shareholders?", Finance Research Letters, Volume 57, 104154, ISSN 1544-6123.

[43] Xuelian Bai, Yi Dong & Nan Hu, (2019), "Financial report readability and stock return synchronicity", Applied Economics, 51:4, 346-363.

[44] Mun G S, Jang S., (2019), "Restaurant firms' IPO motivations and post-IPO performances", International Journal of Contemporary Hospitality Management, 31(9): 3484-3502.

[45] Stoll HR, Curley AJ., (1970), "Small Business and the New Issues Market for Equities", Journal of Financial and Quantitative Analysis, 5(3):309-322.

[46] François Derrien, Kent L. Womack, (2003), "Auctions vs. Bookbuilding and the Control of Underpricing in Hot IPO Markets", The Review of Financial Studies, Volume 16, Issue 1, Pages 31–61.

[47] Lawrence M. Benveniste, Paul A. Spindt, (1989), "How investment bankers determine the offer price and allocation of new issues", Journal of Financial Economics, Volume 24, Issue 2, Pages 343-361.

[48] Marc Goergen, Arif Khurshed, Luc Renneboog, (2009), "Why are the French so different from the Germans? Underpricing of IPOs on the Euro New Markets", International Review of Law and Economics, Volume 29, Issue 3, Pages 260-271.

[49] Lewellen, J., Nagel, S., & Shanken, J.,(2010), "A skeptical appraisal of asset pricing tests", Journal of Financial economics, 96(2), 175-194.

[50] Guldiken, O., Tupper, C., Nair, A. and Yu, H. (2017), "The impact of media coverage on IPO stock performance", Journal of Business Research, Vol. 72, pp. 24-32.

[51] Arnold, T., Fishe, R.P.H. and North, D. (2010), "The effects of ambiguous information on initial and subsequent IPO returns", Financial Management, Vol. 39 No. 4, pp. 1497-1519.

[52] Wang, C.Y. and Wu, J.W. (2015), "Media tone, investor sentiment and IPO pricing", Journal of Financial Research, Vol. 9, pp. 174-189.

[53] Cook, D.O., Kieschnick, R. and Ness, R.A.V. (2006), "On the marketing of IPOs", Journal of Financial Economics, Vol. 82 No. 1, pp. 35-61.

[54] Tetlock, P.C. (2007), "Giving content to investor sentiment: the role of media in the stock market", The Journal of Finance, Vol. 62 No. 3, pp. 1139-1168.

[55] Senter, R. J., & Smith, E. A. ,(1967), "Automated readability index (pp. 1-14)", Technical report, DTIC document.

[56] Coffee, J. C., Jr. (1984), "Market failure and the economic case for a mandatory disclosure system", Virginia Law Review, 70:717–53.

[57] Carter R B, Dark F H, Singh A K., (1998), "Underwriter reputation, initial returns, and the long-run performance of IPO stocks", Journal of Finance, 53(1): 285-311.