Journal of Behavioral and Experimental Finance 37 (2023) 100733

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Finance

journal homepage: www.elsevier.com/locate/jbef

Full length article

The impact of the disclosure characteristics of the application material on the successful listing of companies on China's Science and Technology Innovation Board[☆]

Chen Han, Chengliang Wu, Lu Wei [*]

School of Management Science and Engineering, Central University of Finance and Economics, Beijing 100081, China

A R T I C L E I N F O

Article history:
Received 29 March 2022
Received in revised form 26 July 2022
Accepted 28 July 2022
Available online 4 August 2022

JEL classification:
G14
G18
G39

Keywords:
STAR Market
Registration statement
Inquiry-Reply letter
Disclosure characteristic
Text analysis

A B S T R A C T

The registration statement, the inquiry letter, and the reply letter are the main application materials for companies wanting to list on the Science and Technology Innovation Board (STAR) need to submit to regulatory agencies In this paper, we aim to study the impact of these three kinds of application materials on the successful listing of companies on STAR market in China through six text characteristics, including Words, Boilerplate, Fog Index, HardInfoMix, Redundancy, and Specificity for the first time. In the empirical analysis, we collect the registration statements and the inquiry-reply letters of 220 listed companies and 64 unlisted companies from June 13, 2019 to January 31, 2021 to perform the regression analysis. The empirical results show that, for registration statements, higher Words and Boilerplate will improve the success rate for listing, but higher Redundancy will lead to the failed listing. For the inquiry-reply letter, only the number of questions contained in the inquiry letter is negatively significantly associated with the initial public offering (IPO) success rate, while the text characteristics of the reply letter have little to do with the IPO success rate.

1. Introduction

To reform the structure of China's capital market registration system and enhance the financing support for technological innovation enterprises, the Science and Technology Innovation Board (STAR Market) was established in 2019 and the registration system was applied for the first time. It is independent of the main board market, and its main investment directions are emerging technology industries such as new generation of information technology, new materials, energy-saving and environmental protection, and biomedicine. On this board, Shanghai Stock Exchange (SSE) and China Securities Regulatory Commission (SCRC) have launched a registration-based initial public offering (IPO) system.

Under the registration-based system, the IPO pass rate is affected by many factors. In this paper, we study the effect from the perspective of the application material, including the registration statement and the inquiry-reply letter.[1] The registration statement is an application document submitted by the company for issuance and listing, which contains information about the company's basic situation, risk factors, technical status, and financial status, etc. After the company submits the registration statement, SSE will issue an inquiry letter based on the contents of the registration statement, and the company needs to answer the questions accordingly, which is called the reply letter. Finally, SCRC and SSE decide whether the company can list or not through the contents disclosed. Therefore, the textual application materials determine whether the company can be successfully listed on the STAR Board. Thus, the research on the impact of these three types of application material in STAR Market has important theoretical value and practical significance (see Fig. 1).

Recently, the valuable information contained in text data is getting more and more attention in research. The disclosure of textual information on corporate operating conditions is conducive to investors for a timely and comprehensive understanding of the company's internal operations, which plays an important role in investment decision-making (Manela and Moreira, 2017; Ouyang et al., 2021). In addition, information disclosure

☆ **Acknowledgments**
This work was supported by grants from the National Natural Science Foundation of China (72001223).

* Correspondence to: School of Management Science and Engineering, Central University of Finance and Economics, No. 39 Xueyuan South Road, Haidian District, Beijing 100081, China.
E-mail addresses: hanshenmesen@163.com (C. Han), chengliang_wu@foxmail.com (C. Wu), weilu2014ucas@163.com (L. Wei).

[1] Textual data source: Shanghai Stock Exchange, Available at: http://kcb.sse.com.cn/disclosure/.

https://doi.org/10.1016/j.jbef.2022.100733
2214-6350/

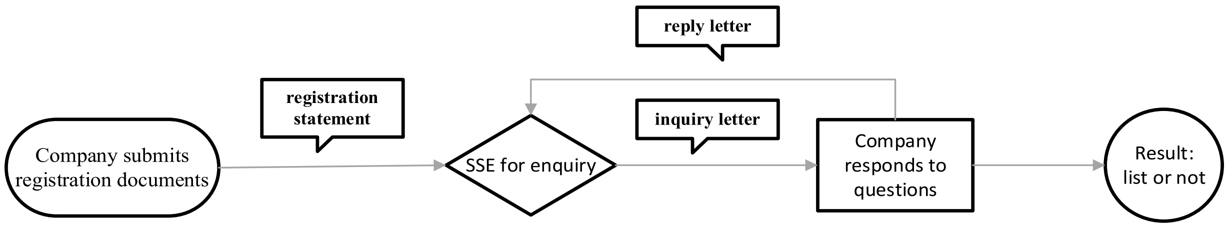

Fig. 1. SSE audit flow chart.

can also reduce the asymmetry of financial market information, optimize resource allocation, and improve the efficiency of the capital. The analysis of the textual information disclosed by listed companies can make up for the lack of financial data, thereby predicting the company's value and future development (Renault, 2017; Semiromi et al., 2020).

Among text documents about the company situation, the financial statement is one of the most important sources of information for investors to understand companies, which contains the financial and risk information of the company, providing valuable information for investors and regulators. Researchers have realized the importance of textual information reported in financial statements and there have been studies focusing on analyzing the qualitative textual disclosure of financial statements (Li et al., 2020). For example, Li (2010) studied the information content of the company's annual report and found that it can predict the company's future profitability and stock liquidity. Campbell et al. (2014) found that risk disclosure information can be used to predict the stock return volatility. Chircop and Tarsalewska (2015) studied the impact of the company's annual report on the efficiency of mergers and acquisitions and found that the longer the annual report of the acquired party, the greater the market response of the acquirer when the M&A announcement was issued. Wei et al. (2019) incorporated textual risk disclosures reported in financial statements into bank risk aggregation and obtain more reasonable aggregate risk results than traditional quantitative financial data. Li et al. (2020) studied the risk dependence between energy corporations based on textual risk disclosures in annual financial reports and found that the text-based risk dependence between the energy corporations is informative about their future stock co-movement.

With the development of computer processing technology, text analysis provides new methods and tools for the structured processing of text information (Qc et al., 2022). Some researchers focus on studying the disclosure characteristics of financial statements. Particularly, Klynveld Peat Marwick Goerdeler (KPMG),[2] one of the four major international accounting firms, concluded that too long length will reduce the readability of the text. And the Financial Accounting Standards Board (FASB) found that the generic and standardized disclosure (referred to as "Boilerplate") was often used by companies in 10-K reports.[3] The United States Securities and Exchange Commission (SEC) had urged firms to evaluate boilerplate disclosure and indicated that redundancies disclosure should also be paid more attention.[4] Mark and Lorien (2015) linked the use of boilerplate to decreased text quality. And Cazier and Pfeiffer (2017) showed that redundant disclosure will lead to less efficient price discovery.

As for specific disclosures, SEC had expressed concern that text disclosure has become increasingly vague and less likely to be supported by quantitative data.[5] Hope et al. (2016) found that more specific disclosures would cause greater market reactions and better risk assessments, leading to a rise in the stock price. Besides, Blankespoor (2019) found that after the adoption of XBRL (eXtensible Business Reporting Language) requirements, companies would increase their quantitative disclosures, which eventually improved the specificity. Dyer et al. (2017) summarized the above five text characteristics and added the Fog Index into the text characteristics and studied their dynamic changes over time in 10-K reports. So far, previous research has summarized 6 text characteristics in 10-K reports, which can comprehensively reflect the disclosure characteristics of the financial text.

However, there are few studies that focus on the text characteristics of the application materials of the STAR market and their impact on IPO success rate in the STAR Market. In this paper, we first comprehensively capture the textual disclosure characteristics of STAR Board application materials, including Words, Boilerplate, Fog Index, HardInfoMix, Redundancy, and Specificity, which are used to further study their impact on the IPO success rate. In the empirical analysis, we collect the registration statements and the inquiry-reply letters of 220 listed companies and 64 unlisted companies from June 13, 2019 to January 31, 2021 in the STAR Market to perform empirical analysis. Our research has important theoretical and practical implications. By first founding the relationships between textual disclosure characteristics of STAR Board application materials and the IPO success rate, we can provide suggestions on how to write application materials for enterprises to improve the success rate of listing; For regulators, our research findings can help them improve the writing requirements of enterprise listing application materials and screen out high-quality enterprises suitable for listing on STAR Market.

The structure of this paper is as follows. In Section 2, we present the methods of calculating the disclosure characteristics and the regression model. Section 3 provides the empirical results. And Section 4 offers recommendations and concludes the paper.

2. Methodology

In this section, we present the definition and calculation methods of six text disclosure characteristics and the regression model in the empirical part.

2.1. The definition of six text characteristics

1 Words

The length of the registration statement is undoubtedly important. Prior research believes too long disclosure words will lead to

[2] Klynveld Peat Marwick Goerdeler, Disclosure overload and complexity: hidden in plain sight. Available at: http://www.kpmg.com/US/en/IssuesAndInsights/ArticlesPublications/Documents/disclosure-overload-complexity.pdf.

[3] Financial Accounting Standards Board, Disclosure framework: Invitation to comment. Norwalk, CT, 2012.

[4] Securities and Exchange Commission, Disclosure effectiveness: remarks before the American bar association business law section spring meeting. Available at: https://www.sec.gov/news/speech/2014-spch041114kfh, 2014.

[5] Securities and Exchange Commission, A plain English handbook: How to create clear SEC disclosure documents. Available at: https://www.sec.gov/pdf/handbook.pdf, 1998.

Table 1
The definition of characteristics.

Text characteristics	Description
Words	The number of words used in the text
Boilerplate	The percent of sentences containing words that shared by at least 75% of all firms in a given periodic year
Fog Index	The Fog Index = 0.4[(total words/total sentences) + 100*(difficult words/total words)], where difficult words are the words out of Chinese common dictionary
HardInfoMix	The percent of informative numbers (price, tax, product quantity, etc.)
Redundancy	The percent of words in sentences that are repeated verbatim in other portions of the whole text
Specificity	The percent of entities (locations, people, organizations, dollar amounts, percentages, dates, or times)

less readability and more redundancy, and insufficient disclosure is often proved to be poor quality of disclosure (Li et al., 2020).

In this paper, we consider the length of the disclosure, measured by the number of words of the registration statement. Generally speaking, the longer the article, the more informative the information disclosure. However, too long text disclosure will also lead to the increase of redundancy and boilerplate, hindering investors from acquiring the valuable information, and eventually leading to an impact on the IPO underpricing.

2 Boilerplate

We measure boilerplate by counting sentences that contain words that are extremely common among registration statements. It represents how many template words are used among the enterprises.

For standard format, simply increasing the length of disclosure is not helpful if the content does not contain any useful information. Moreover, it may even provide opportunities to hide information and reduce informativeness (Mark and Lorien, 2015). Generally speaking, registration statements containing more boilerplate tend to be less informative. Therefore, the increase in boilerplate will reduce the unique information and affect the IPO underpricing.

3 Fog Index

The Fog Index is used to measure the readability of text based on the length of sentences and the proportion of complex terminology. It shows how many years of education it takes to understand the text and expresses how easy the text is to read.

The Fog is a common indicator of text readability. If it is too high, readers may find it difficult to understand the text. If it is too low, the professionalism of the text may be insufficient. Therefore, The Fog Index can affect the difficulty for investors to obtain information, affecting the investment and IPO underpricing (Li, 2008).

4 HardInfoMix

The percent of numbers is also one of the important elements in the quality of text disclosure, but it inevitably contains various useless numbers, such as omitting dates, section numbers, etc. HardInfoMix stands for the percent of valuable numbers in the text, such as the price, tax, product quantity, etc. Valuable numbers can play an important role in investors' decisions, which present the status of the business and affect the IPO underpricing (Dyer et al., 2017).

The following numbers will not be within our count: (1) 4-digit number without commas (such as 2009); (2) the date (such as 2008.01.01); (3) the citation notes or chapters (such as Note 7, Item 9, Section 2) (see Table 1).

5 Redundancy

Redundancy is defined as the proportion of invalid words in the disclosures. The redundant words stand for the words that are repeated verbatim in other portions of the text. And redundancy is the percent of redundant words in a text. It reflects the amount of useful information in the registration statement.

Excessive redundancy will often lead to a decline in readability, and it can change with the purpose of the registration statement (Cazier and Pfeiffer, 2017). Redundant disclosure can be linked to the purpose of writing the text. For example, giving redundant information in the corresponding part will hide certain risk information. Thus, investors' choices will be influenced as a result, eventually affecting the IPO underpricing.

6 Specificity

Specificity, as the percent of specific words or phrases conveying specific information relevant to the disclosing firm, is divided by the number of total words. The more entity names, the stronger credibility of the text, which will attract people to invest and influence the IPO underpricing (Hope et al., 2016).

For example, the specific name of "Shanghai Stock Exchange" contains more idiosyncratic details than using the general word of "institution". The category of entity names includes (1) names of persons, (2) names of locations, (3) names of organizations, (4) other specific entities, (5) times, and (6) dates.

2.2. Regression model

To study the impact of disclosure characteristics of registration statements on the IPO success rate, post-IPO performance, and stock price volatility, we establish the corresponding logistic regression models. The logistic regression model is a non-linear probability model. The independent variables do not need to obey the assumption of normal distribution, and the variables in the model can be continuous, discrete, or dummy variables. The dependent variable of the model is a binary variable, which can only take two values of 0 and 1. The regression model can ensure that the probability value obtained is meaningful, and it is often used in fields such as data mining, automatic disease diagnosis, and economic forecasting.

The logistic regression model in this paper is expressed as follows:

$$y = \alpha + \sum_{i=1}^{n} \beta_i x_i$$

The value of y is discrete (0 or 1), which represents if the company is successfully listed (1) or not (0) on the STAR Market; α is the constant term; β_i is the coefficient to be estimated; x_i is the explanatory variable, including six characteristics and control variables.

When studying the link between text characteristics and IPO success rate, we mainly control two types of variables, IPO-related characteristics and corporate characteristics. Table 2 present the following 7 control variables, based on the study of Yang (2013), Krishnan et al. (2008), and Yuan et al. (2019), and the availability of data from the database.

3. Empirical results

3.1. Data description

Our initial sample includes 220 listed companies and 64 unlisted companies from June 13, 2019 to January 31, 2021 in the STAR Market. The STAR Market was officially launched on June 13,

Table 2
The description of control variables.

Notation	Description
LnTA	The logarithm of the company's total assets in the year before the IPO
A	Years from company establishment to listing
LEV	Leverage ratios
ROA	Return on asset
GROWTH	Year-on-year main business income growth rates
CR	Current ratio
U	Whether the underwriter is among the top ten

Table 3
Descriptive statistics of the six text characteristics in successful IPO companies.

Text characteristics	Mean	Coefficient of variation	Skewness	Kurtosis
Words	229 019	0.252	1.03	3.82
Fog Index	18.85	0.057	4.70	42.41
Redundancy	0.9586	0.007	0.09	2.48
HardInfoMix	0.0182	0.193	0.39	5.07
Specificity	0.0046	0.323	4.57	40.98
Boilerplate	0.7690	0.233	−1.08	2.45

2019, so our initial sample includes all companies from the beginning of the board in this period. The data is obtained from the SSE website and the Wind Economy Database. Companies with incomplete IPO and financial data were eliminated, and the continuous variables were reduced by 1% before and after. Through the above steps, finally we obtained 257 samples, including 193 successful companies and 64 unsuccessful companies.

SSE may conduct several rounds of inquiry about the companies, and the samples we collected here are all the first-round inquiry-reply letters. For the registration statement and the reply letter, we use the 6 text characteristics to measure their textual disclosure. For the inquiry letter, we use the number of questions contained in the inquiry letter to measure the disclosure characteristic.

3.1.1. The registration statements

The registration statement is the most important document submitted by the company when listing in the STAR Market. For the companies which executed successful IPO, their registration statements may have something in common. In other words, text characteristics may present some common ground to some extent. Table 3 gives the descriptive statistics of the text characteristics and their distributions are visually shown in Fig. 2.

As can be seen, for the registration statements of the companies which succeed in IPO, the average number of words is 229,019. And the mean values of Fog Index, Redundancy, HardInfoMix, Specificity, and Boilerplate are 18.85, 0.9586, 0.0182, 0.0046, and 0.7690 respectively. Meadows (1986) found the Fog Index of national daily newspapers in the UK is about 12 and papers on scientific journals is about 19, and Dyer et al. (2017) also found that Fog Index of form 10-K is 21.34, which shows that the reading difficulty of registration statements of STAR market is similar to that of scientific journals but lower than 10-K reports. They also found the average HardInfoMix of 10-K reports is 0.0187, which is very close to that of the STAR market registration statements.

The coefficient of variation (CV) is a standardized measure of the degree of data dispersion. The actual value of the CV is independent of the unit in which the measurement has been taken, so it is a dimensionless number. For comparison between data sets with different units or widely different means, we should use the coefficient of variation instead of the standard

Table 4
Descriptive statistics of the six text characteristics in failed IPO companies.

Text characteristics	Mean	Coefficient of variation	Skewness	Kurtosis
Words	203 905	0.247	1.60	5.85
Fog Index	18.74	0.046	0.14	3.25
Redundancy	0.9565	0.006	0.54	3.37
HardInfoMix	0.0179	0.222	0.20	3.23
Specificity	0.0045	0.246	4.57	2.52
Boilerplate	0.6314	0.291	0.24	1.21

Table 5
Comparison between listed companies and unlisted companies in the registration statement.

Text characteristics	Mean	Coefficient of variation	Skewness	Kurtosis
Words(successful)	229 019	0.252	1.03	3.82
Words(failed)	203 905	0.247	1.60	5.85
Fog Index(successful)	18.85	0.057	4.70	42.41
Fog Index(failed)	18.74	0.046	0.14	3.25
Redundancy(successful)	0.9586	0.007	0.09	2.48
Redundancy(failed)	0.9565	0.006	0.54	3.37
HardInfoMix(successful)	0.0182	0.193	0.39	5.07
HardInfoMix(failed)	0.0179	0.222	0.20	3.23
Specificity(successful)	0.0046	0.323	4.57	40.98
Specificity(failed)	0.0045	0.246	4.57	2.52
Boilerplate(successful)	0.7690	0.233	−1.08	2.45
Boilerplate(failed)	0.6314	0.291	0.24	1.21

deviation. A larger CV means more dispersed data. We can see that in the six text characteristics, Specificity (0.323) is the most dispersed, closely followed by Words (0.252), Boilerplate (0.233), and HardInfoMix (0.193). The Fog Index degree of dispersion is low (0.057), and Redundancy is the least dispersed one (only 0.007). In the registration statements of the entire market, the difference of Specificity is the biggest, closely followed by Words, Boilerplate, HardInfoMix, and Fog Index, the smallest one is Redundancy.

Similarly, for the companies which failed in IPO, the text characteristics of the registration statements may also have some similarities. Table 4 gives the descriptive statistics of the text characteristics and their distributions are visually shown in Fig. 3.

As can be seen, for the companies which failed in IPO, the average number of words is 203,905. And the mean values of Fog Index, Redundancy, HardInfoMix, Specificity, and Boilerplate are 18.74, 0.9565, 0.0179, 0.0045, and 0.6314 respectively.

In the CV, different from the firms which succeed in IPO, Boilerplate (0.291) is the most dispersed, closely followed by Words (0.247), Specificity (0.246), and HardInfoMix (0.222). The Fog Index degree of dispersion is low (0.046), and Redundancy is the least dispersed one (only 0.006). In the registration statements of the firms which failed in IPO, the difference of Boilerplate is the biggest, closely followed by Words, Specificity, HardInfoMix, and Fog Index, the smallest one is Redundancy.

Is there any difference in text characteristics between successful and failed companies? The comparison between them is shown in Table 5. We can see that there are relatively large differences between successful IPO and failed IPO companies in some text characteristics (such as Words and Boilerplate). But some text characteristics (especially Specificity) are nearly identical between successful IPO and failed IPO companies. However, this conclusion is only a simple observation, more accurate results need to be determined by regression.

3.1.2. The inquiry letter

The inquiry letter refers to the document issued by SSE in response to the registration statement, where contains questions about the problems in the company's registration statement.

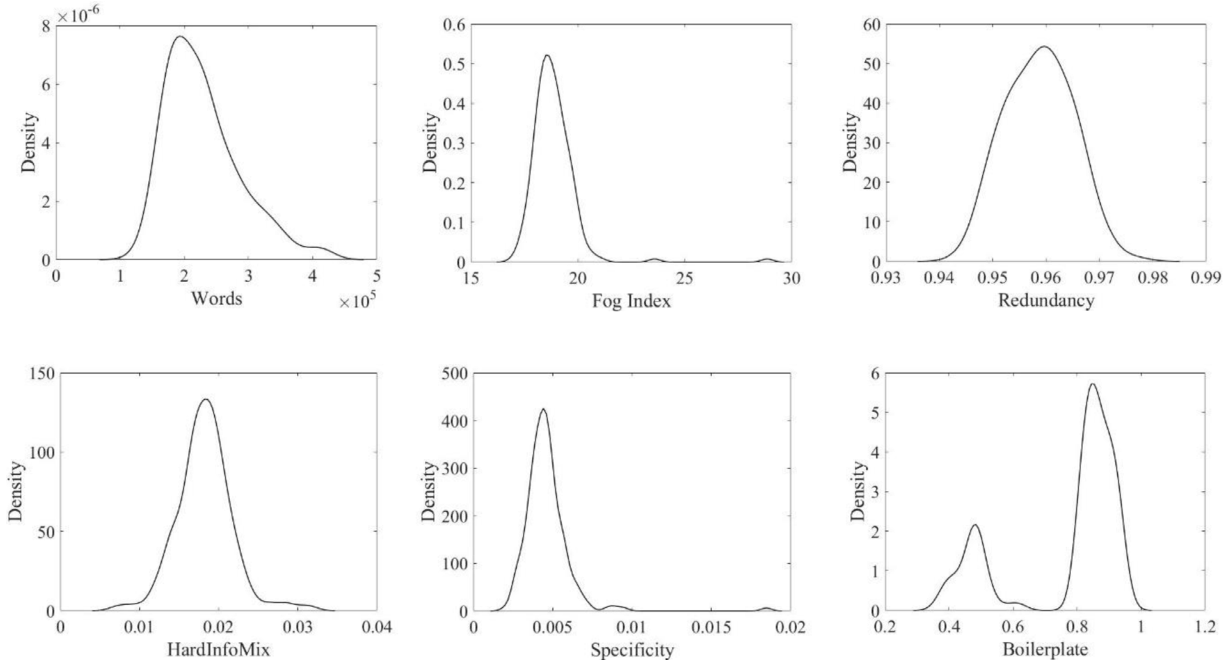

Fig. 2. The distribution of six characteristics of text disclosure in listed companies.

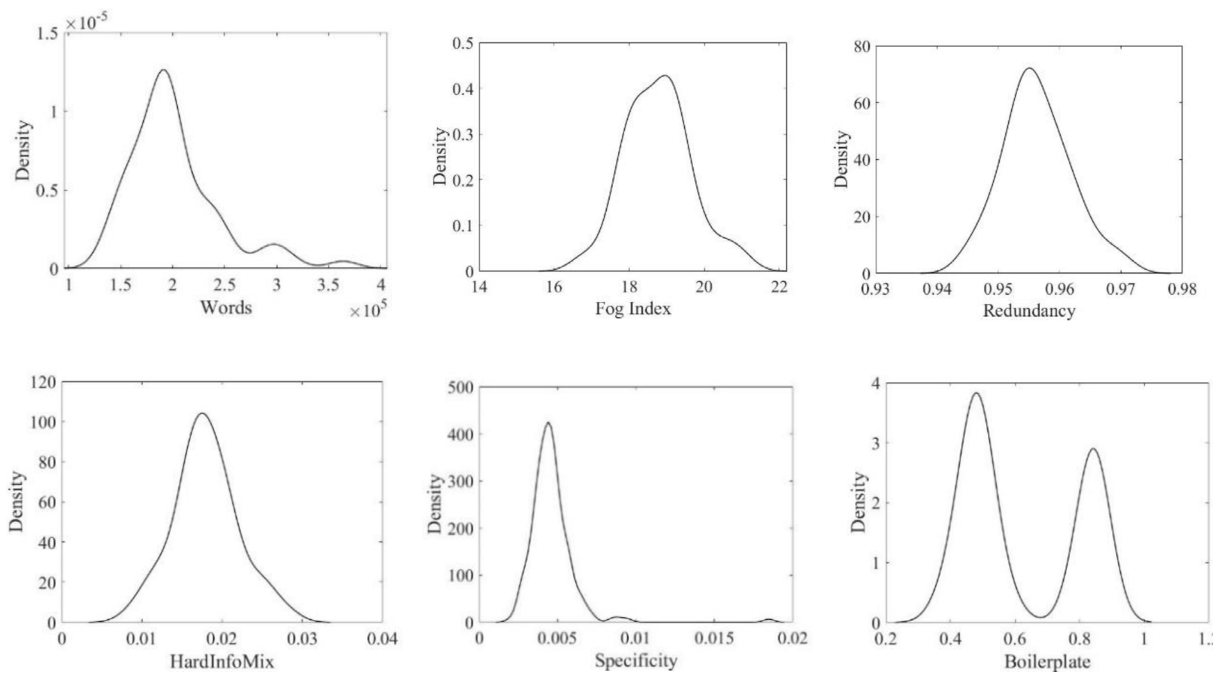

Fig. 3. The distribution of six characteristics of text disclosure in listed companies.

The samples we collected in the paper are all the first-round inquiry-reply letters. Since it is written by the regulator, its textual characteristics are relatively less important. We choose the number of questions rather than the text characteristics as our measure criterion. Table 6 gives the descriptive statistics of the question number for listed and unlisted companies and their distributions are visually shown in Fig. 4.

As can be seen, the average number of questions of successful listed companies is 48.3121, while that of failed listed companies is 51.1818. On the whole, listed companies are asked fewer questions than unlisted companies, which is consistent with our expected result. If the content disclosed by the company in the registration statement is more reliable and the quality of the disclosure is better, regulators will raise fewer questions, which will finally lead to a higher IPO success rate.

Table 6
Comparison between listed companies and unlisted companies in the question number.

Question Number	Mean	Standard deviation	Skewness	Kurtosis
Listed companies	48.3121	11.3494	0.615	0.685
Unlisted companies	51.1818	10.1271	0.355	−0.342

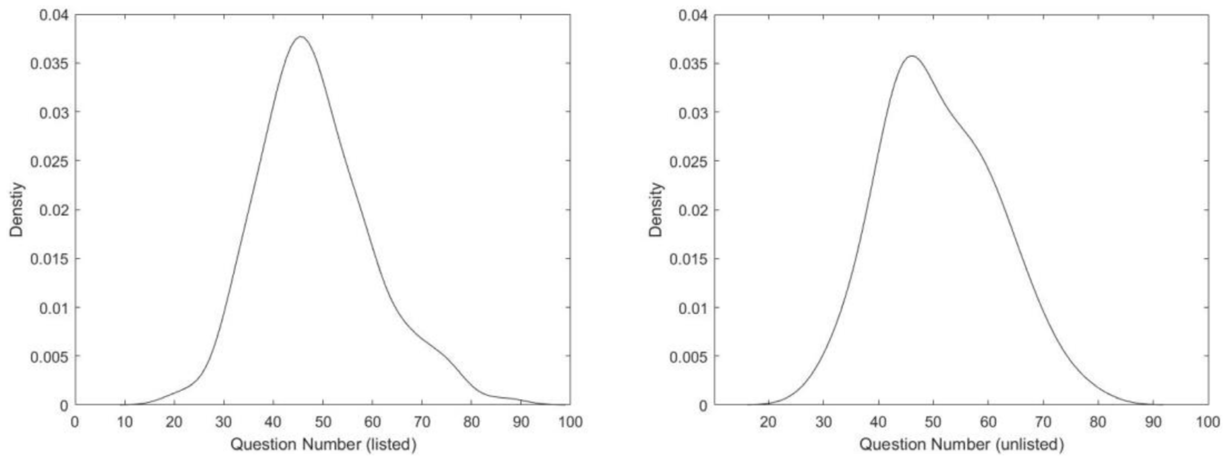

Fig. 4. The distribution of the question number.

Table 7
Descriptive statistics of the six text characteristics of the reply letter in listed companies.

Text characteristics	Mean	Coefficient of variation	Skewness	Kurtosis
Words	205 935	0.304	0.84	1.22
Fog Index	18.48	0.045	0.49	0.58
Redundancy	0.9595	0.008	0.07	1.09
HardInfoMix	0.0042	0.421	2.34	12.12
Specificity	0.0175	0.457	1.96	7.00
Boilerplate	0.6562	0.090	−0.19	−0.47

Table 8
Descriptive statistics of the six text characteristics of the reply letter in unlisted companies.

Text characteristics	Mean	Coefficient of variation	Skewness	Kurtosis
Words	242 018	0.317	0.66	−0.24
Fog Index	18.53	0.043	0.13	0.49
Redundancy	0.9636	0.007	−0.57	0.09
HardInfoMix	0.0039	0.322	1.70	3.78
Specificity	0.0157	0.309	1.43	3.83
Boilerplate	0.6524	0.090	−0.09	−0.90

The standard deviation of the number of questions of successful listed companies is 11.3494, while that of the failed listed companies is 10.1271, indicating that the number of questions of successful companies is more dispersed, which may be due to the larger number of successful samples.

3.1.3. The reply letter

The reply letter is a kind of official document used by companies to respond to the questions in the inquiry letter. After the regulator's several rounds of inquiries, it finally decides whether the company can go public. Table 7 gives the descriptive statistics of the text characteristics of listed companies and their distributions are visually shown in Fig. 5.

As can be seen, for the reply letters of the listed companies, the average number of words is 205 935, and the mean values of Fog Index, Redundancy, HardInfoMix, Specificity, and Boilerplate are 18.48, 0.9595, 0.0042, 0.0175, and 0.6562 respectively. For the CV, we can see that in the six text characteristics, Specificity (0.457) is the most dispersed, closely followed by HardInfoMix (0.421) and Words (0.304). The Boilerplate (0.090) and the Fog Index (0.045) is relatively low, and Redundancy is the least dispersed one (only 0.008).

Similarly, for the companies which failed in IPO, Table 8 gives the descriptive statistics of the text characteristics and their distributions are visually shown in Fig. 6.

As can be seen, for the reply letters of the unlisted companies, the average number of words is 242 018, and the mean values of Fog Index, Redundancy, HardInfoMix, Specificity, and Boilerplate are 18.53, 0.9636, 0.0039, 0.0157, and 0.6524 respectively. For the CV, we can see that in the six text characteristics, HardInfoMix (0.322) is the most dispersed, closely followed by Words (0.317) and Specificity (0.309). The Boilerplate (0.090) and the Fog Index (0.043) is relatively low, and Redundancy is the least dispersed one (only 0.007).

Table 9
Comparison between listed companies and unlisted companies in the reply letter.

Text characteristics	Mean	Coefficient of variation	Skewness	Kurtosis
Words(successful)	205 935	0.304	0.84	1.22
Words(failed)	242 018	0.317	0.66	−0.24
Fog Index(successful)	18.48	0.045	0.49	0.58
Fog Index(failed)	18.53	0.043	0.13	0.49
Redundancy(successful)	0.9595	0.008	0.07	1.09
Redundancy(failed)	0.9636	0.007	−0.57	0.09
HardInfoMix(successful)	0.0042	0.421	2.34	12.12
HardInfoMix(failed)	0.0039	0.322	1.70	3.78
Specificity(successful)	0.0175	0.457	1.96	7.00
Specificity(failed)	0.0157	0.309	1.43	3.83
Boilerplate(successful)	0.6562	0.090	−0.19	−0.47
Boilerplate(failed)	0.6524	0.090	−0.09	−0.90

To study the text characteristics of reply letters from listed companies and unlisted companies, the comparison between them is shown in Table 9. We can see that the average Words of the reply letter of listed companies is significantly less than that of failed companies, and the differences of other text characteristics are relatively small. Similarly, this conclusion is also only a simple observation, more accurate results still need to be determined by regression.

3.2. Link between the registration statement and IPO success rate

Next, we will use the binary logistic regression model to study the relationship. The explained variable, SUCCESS, is a dummy variable that equals 1 if the company is successfully listed and 0 if the company is rejected. We use LnTA, A, LEV, ROA, GROWTH, CR, and U as control variables, and Words, Fog Index, Redundancy, HardInfoMix, Specificity, and Boilerplate in registration statement texts as explanatory variables.

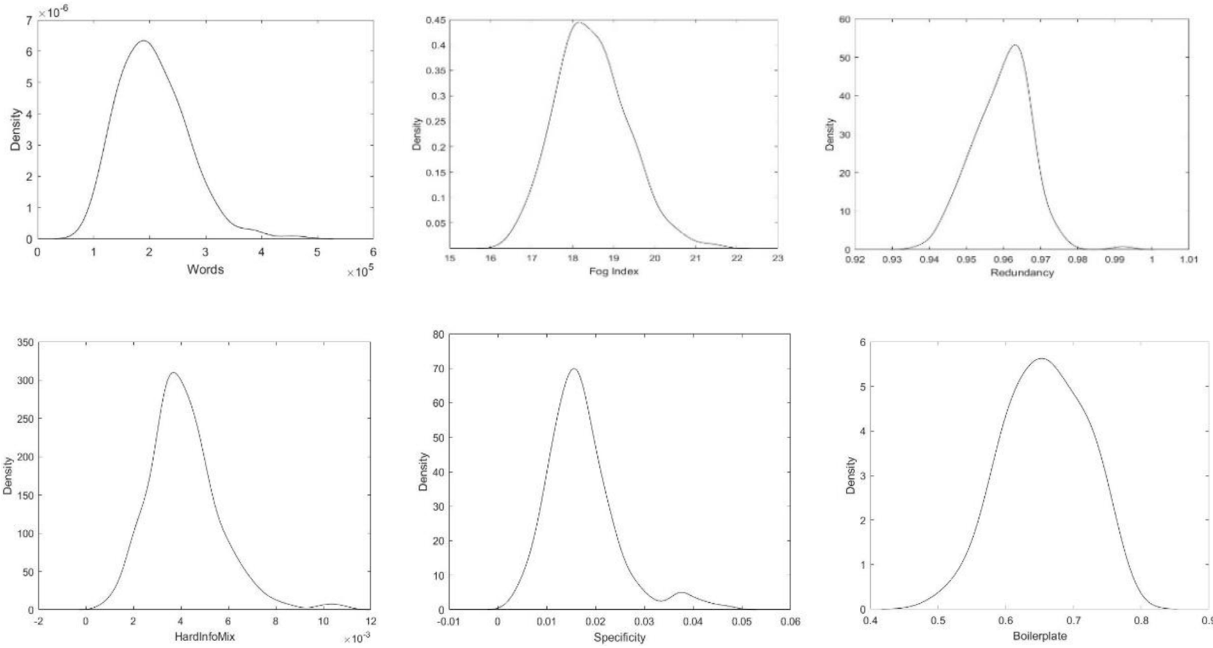

Fig. 5. The distribution of six characteristics of the reply letter in listed companies.

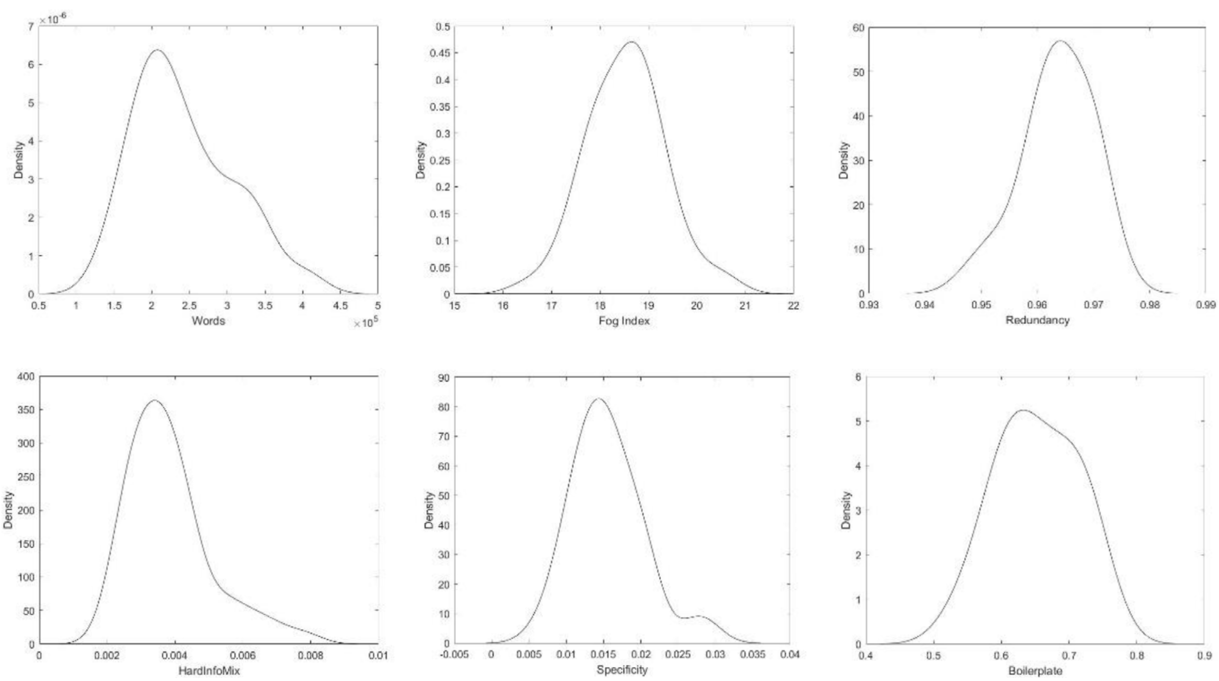

Fig. 6. The distribution of six characteristics of the reply letter in unlisted companies.

We estimate the logistic model in the following equation to model the influence of different types of text characteristics on the probability of IPO success:

$$SUCCESS = \alpha + \beta_1 \cdot Words + \beta_2 \cdot Fog\ Index + \beta_3 \cdot Redundancy$$

$$+\beta_4 \cdot HardInfoMix + \beta_5 \cdot Specificity + \beta_6 \cdot$$

$$Boilerplate + \beta_7 \cdot LnTA + \beta_8 \cdot A + \beta_9 \cdot LEV + \beta_{10} \cdot ROA$$

$$+\beta_{11} \cdot GROWTH + \beta_{12} \cdot CR + \beta_{13} \cdot U$$

where α captures unobserved company-specific effects, ε is the random error term, and $\beta_1 \sim \beta_6$ are the coefficient of the explanatory variables while $\beta_7 \sim \beta_{13}$ are the model coefficient of the control variables.

We first estimate a logistic model of SUCCESS on all control variables. Their coefficients are presented in Table 10, columns (2). As can be seen, the control variables have some explanatory power, and their coefficients have the expected signs. Among them, GROWTH and U are significantly positively associated with SUCCESS, LEV and ROA are significantly negatively associated with SUCCESS.

We next estimate a full cross-sectional model by further including all variables. Their coefficients are presented in Table 10, columns (1). As can be seen, Words, Redundancy, Boilerplate are significantly associated (positively, negatively and positively,

Table 10
Estimation result of multiple linear regression analysis.

Variable	Coefficient (1)	Coefficient (control-only) (2)
Words	2.29E−05** (0.028)	
Fog Index	−0.2972 (0.343)	
Redundancy	−191.0972** (0.023)	
HardInfoMix	−75.2550 (0.310)	
Specificity	113.8125 (0.544)	
Boilerplate	3.8639*** (0.004)	
LnTA	0.5243 (0.141)	0.7575*** (0.010)
A	0.0219 (0.681)	0.2304 (0.818)
LEV	−8.8407*** (0.000)	−9.2000*** (0.000)
ROA	−25.8880*** (0.000)	−25.564*** (0.000)
GROWTH	1.3898*** (0.000)	1.3336*** (0.000)
CR	4.0607 (0.628)	2.1357 (0.773)
U	1.1055** (0.037)	0.9115* (0.063)
Intercept	183.6058** (0.020)	1.7391 (0.169)

Note. p-values in parentheses.
*Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.

respectively) with SUCCESS. Coefficients of other variables do not have significant explanatory power. For the three text characteristics which have statistically explanatory power, the coefficient of Words is 2.29×10^{-5}. We can find that in the STAR market, a longer registration statement can improve the success rate of IPO, it is consistent with our intuition. A longer registration statement often means more useful information is disclosed, some of which are necessary for the IPO audit. Similarly, Chircop and Tarsalewska (2015) found that the length of the company's annual report on the efficiency of mergers and acquisitions can affect the market response positively.

The coefficient of Redundancy is −191.0972, which means that a lower Redundancy can improve the success rate of IPO. Cazier and Pfeiffer (2017) pointed out that higher redundant information disclosure would lead to less efficient price discovery. One possible explanation is that higher Redundancy means the registration statement has more repetition parts. Lots of repetition is ineffective and increases the cost of information discovery, bringing a negative impact on the approval of IPO audit.

The coefficient of Boilerplate is 3.8639, which means that a higher Boilerplate can improve the success rate of IPO. Higher Boilerplate means that registration statements use more phrases that are widely used in the same industry. The high boilerplate can be considered that the template language of its declaration is similar to that of other companies in the industry, which means that the "standard template" is used. This means that for passing IPO audit, it is beneficial to refer to the writing routines and templates of successful IPO companies in the same industry. At the same time, it is noted that Mcclane (2019) believes that the higher Boilerplate in the application documents will reduce the time of amendments to the application documents, which is conducive to the approval of IPO audit.

3.3. Link between inquiry-reply letter and IPO success rate

In this experiment, the explained variable is the same as above. The difference from the previous experiment is that our explanatory variable here has one more Question Number (which stands for the number of questions in the inquiry letter), and the six text characteristics are derived from the company's reply letter. The meaning of each coefficient is also the same as the previous experiment. Finally, we get the following logistic model to study the influence of the inquiry letter and the reply letter on the probability of IPO success:

$$SUCCESS = \alpha + \beta_1 \cdot Words + \beta_2 \cdot FogIndex + \beta_3 \cdot Redundancy + \beta_4 \cdot HardInfoMix + \beta_5 \cdot Specificity + \beta_6 \cdot Boilerplate + \boldsymbol{\beta_7} \cdot \textbf{\textit{Question Amount}} + \beta_8 \cdot LnTA + \beta_9 \cdot A + \beta_{10} \cdot LEV + \beta_{11} \cdot ROA + \beta_{12} \cdot GROWTH + \beta_{13} \cdot CR + \beta_{14} \cdot U$$

Table 11
Estimation result of multiple linear regression analysis.

Variable	Coefficient (1)	Coefficient (control-only) (2)
Words	0.0003 (0.218)	
Fog Index	−0.0551 (0.860)	
Redundancy	−97.9281 (0.247)	
HardInfoMix	97.8860 (0.150)	
Specificity	107.3751 (0.694)	
Boilerplate	2.7304 (0.624)	
Question Number	−0.061** (0.033)	
LnTA	0.7042** (0.047)	0.6721** (0.025)
A	−0.0146 (0.803)	−0.0025 (0.818)
LEV	−10.5902*** (0.000)	−9.7890*** (0.000)
ROA	−29.0060*** (0.000)	−28.2496*** (0.000)
GROWTH	1.6340*** (0.000)	1.4296*** (0.000)
CR	1.637 (0.844)	1.1775 (0.886)
U	−0.959 (0.109)	−0.8950* (0.092)
Intercept	95.6521 (0.239)	3.426** (0.021)

Note. p-values in parentheses.
*Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.

Similarly, we first estimate a logistic model of SUCCESS on all control variables. Their coefficients are presented in Table 11, columns (2). As can be seen, the control variables have some explanatory power, and their coefficients have the expected signs. Among them, LnTA and GROWTH are significantly positively associated with SUCCESS, and A, LEV, ROA, and U are significantly negatively associated with SUCCESS.

Next, we estimate a full cross-sectional model by further including all variables. Their coefficients are presented in Table 11, columns (1). The result shows only Question Number is negatively significantly associated with SUCCESS. Coefficients of other variables do not have significant explanatory power.

The coefficient of Question Number is −0.061, which means that a lower Question Number can improve the success rate of IPO. The more questions regulator raises in the inquiry letter, the lower the success rate for the company's listing, which is consistent with our expected results. If the regulator raises fewer questions to the company, it means the better quality of the information disclosed in the company's registration statement.

All of the six text characteristics have no significant explanatory power, which means that the disclosure quality of the reply letter has little to do with the success rate of listing. One possible explanation is that the regulator's attention to the reply letter focuses on the company's explanation of the problems, rather than the text disclosure quality. It only requires the company to clarify the problem clearly and does not need to make too many supplements, so the text characteristics of the reply letter will not be so valuable.

4. Conclusion

To our knowledge, this is the very first attempt to examine the impact of the main application materials of companies, including the registration statements, inquiry letters, and reply letters on their IPOs in the STAR Market. The main contribution of this paper is to analyze the impact of the three kinds of disclosed text on the IPO success rate through the six text characteristics, including Words, Boilerplate, Fog Index, HardInfoMix, Redundancy, and Specificity.

In the empirical analysis, we collect the registration statements and the inquiry-reply letters of 220 listed companies and 64 unlisted companies from June 13, 2019 to January 31, 2021 in the STAR Market. The empirical results show that for registration statements, more Words and higher Boilerplate will improve the success rate for listing, but higher Redundancy will lead to the failed listing. A longer registration statement often means more useful information is disclosed, which is necessary for the IPO audit. Besides, the higher Boilerplate means that the registration statement uses more standard template phrases that are widely used in the same industry, and this kind of standardized disclosure is conducive to the IPO audit in the STAR Market. However, a higher Redundancy will reduce the success rate of IPO. Lots of repetition is ineffective and increases the cost of information discovery, bringing a negative impact on the approval of IPO audit.

For the inquiry-reply letter, only the number of questions contained in the inquiry letter is negatively significantly associated with the IPO success rate. In the first round of inquiries, if the company is questioned more questions, it shows there are more problems in the registration statement disclosed by the company, thereby the success rate of listing will be lower. As for the reply letter, all of the six text characteristics have no significant explanatory power, which shows that the text characteristics of the reply letter have no significant impact on the IPO success rate.

Nevertheless, this study has certain limitations. One is that we do not analyze the impact of disclosure content of the application materials on the IPO success rate in China's STAR market. This paper mainly focuses on analyzing the textual disclosure characteristics of STAR market application materials. Therefore, in future research, we will further study the impact of disclosure content contained in the applications materials on the companies' listing success rates in STAR market.

CRediT authorship contribution statement

Chen Han: Conceptualization, Methodology, Writing – original draft. **Chengliang Wu:** Data curation, Visualization, Investigation, Software. **Lu Wei:** Supervision, Software, Validation, Writing – review & editing.

References

Blankespoor, E., 2019. The impact of information processing costs on firm disclosure choice: evidence from the XBRL mandate. J. Account. Res. 57, 919–967.

Campbell, J.L., Chen, H., Dhaliwal, D.S., Lu, H.M., Steele, L.B., 2014. The information content of mandatory risk factor disclosures in corporate filings. Rev. Account. Stud. 19 (1), 396–455.

Cazier, R.A., Pfeiffer, R.J., 2017. 10-K disclosure repetition and managerial reporting incentives. J. Financ. Report. 2 (1), 107–131.

Chircop, J., Tarsalewska, M., 2015. Annual report length and M & A efficiency. Eur. J. Finance 26 (6), 532–553.

Dyer, T., Lang, M., Stice-Lawrence, L., 2017. The evolution of 10-K textual disclosure: Evidence from latent dirichlet allocation. J. Account. Econ. 64 (2–3), 221–245.

Hope, O.K., Hu, D., Lu, H., 2016. The benefits of specific risk-factor disclosures. Rev. Account. Stud. 21, 1005–1045.

Krishnan, C., Masulis, R.W., Singh, A.K., 2008. Does venture capital reputation matter? Evidence from subsequent IPOs. In: American Finance Association Annual Meeting.

Li, F., 2008. Annual report readability, current earnings, and earnings persistence. J. Account. Econ. 45 (2–3), 221–247.

Li, F., 2010. The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach. J. Account. Res. 48 (5), 1049–1102.

Li, J., Li, J., Zhu, X., 2020. Risk dependence between energy corporations: A text-based measurement approach. Int. Rev. Econ. Finance 68, 33–46.

Manela, A., Moreira, A., 2017. News implied volatility and disaster concerns. J. Financ. Econ. 123 (1), 137–162.

Mark, L., Lorien, S.T., 2015. Textual analysis and international financial reporting: large sample evidence. J. Account. Econ. 60, 110–135.

Mcclane, J., 2019. Boilerplate and the impact of disclosure in securities dealmaking. Vanderbilt Law Rev. 72 (1), 191–293.

Meadows, A.J., 1986. The readability of physics papers. Czech. J. Phys. 36 (1), 89–91.

Ouyang, Z.S., Yang, X.T., Lai, Y., 2021. Systemic financial risk early warning of financial market in china using attention-lstm model. North Am. J. Econ. Finance 56 (3), 101383.

Qc, B., Hz, A., Dy, A., Lh, C., 2022. How does investor attention matter for crude oil? Evidence from time-frequency quantile causality analysis. 59, 101581.

Renault, T., 2017. Intraday online investor sentiment and return patterns in the US stock market. J. Bank. Financ. 84, 25–40.

Semiromi, H.N., Lessmann, S., Peters, W., Beladi, H., 2020. News will tell: forecasting foreign exchange rates based on news story events in the economy calendar. North Am. J. Econ. Finance 52, 101181.

Wei, L., Li, G., Li, J., Zhu, X., 2019. Bank risk aggregation with forward-looking textual risk disclosures. North Am. J. Econ. Finance 50, 101016.

Yang, Z., 2013. Do political connections add value to audit firms? Evidence from IPO audits in China. Contemp. Account. Res. 30 (3), 891–921.

Yuan, H., Zhang, C., Kong, D., Shi, H., 2019. The consequence of audit failure on audit firms: evidence from ipo approval in china. China J. Account. Stud. 7 (2), 245–269.