Annals of Operations Research (2024) 334:761–797
https://doi.org/10.1007/s10479-022-05012-8

ORIGINAL RESEARCH

Cross-influence of information and risk effects on the IPO market: exploring risk disclosure with a machine learning approach

Huosong Xia[1,2] · Juan Weng[1] · Sabri Boubaker[3,4,5] · Zuopeng Zhang[6] · Sajjad M. Jasimuddin[7]

Accepted: 27 September 2022 / Published online: 22 October 2022

Abstract
The paper examines whether the structure of the risk factor disclosure in an IPO prospectus helps explain the cross-section of first-day returns in a sample of Chinese initial public offerings. This paper analyzes the semantics and content of risk disclosure based on an unsupervised machine learning algorithm. From both long-term and short-term perspectives, this paper explores how the information effect and risk effect of risk disclosure play their respective roles. The results show that risk disclosure has a stronger risk effect at the semantic novelty level and a more substantial information effect at the risk content level. A novel aspect of the paper lies in the use of text analysis (semantic novelty and content richness) to characterize the structure of the risk factor disclosure. The study shows that initial IPO returns negatively correlate with semantic novelty and content richness. We show the interaction between risk effect and information effect on risk disclosure under the nature of the same stock plate. When enterprise information transparency is low, the impact of semantic novelty and content richness on the IPO market is respectively enhanced.

Keywords Risk disclosure · Information effect · Risk effect · Enterprise information transparency · Unsupervised machine learning algorithm

✉ Sajjad M. Jasimuddin
sajjad.jasimuddin@kedgebs.com

1 School of Management, Wuhan Textile University, Wuhan 430073, China

2 Research Center of Enterprise Decision Support, Key Research Institute of Humanities and Social Sciences, Wuhan 430073, China

3 EM Normandie Business School, Métis Lab, France

4 International School, Vietnam National University, Hanoi, Vietnam

5 Swansea University, Swansea, UK

6 Coggin College of Business, University of North Florida, Jacksonville, FL 32224, USA

7 Kedge Business School, Marseille, France

1 Introduction

Initial public offerings (IPO) have become one of the core strategic means of raising capital for cash-hungry companies (Grover & Bhullar, 2021). In the recent years, the risk of IPO has brought disasters to the operation of China's capital market (Saci & Jasimuddin, 2021). The research of information disclosure plays an essential role in promoting the effective operation of the capital market (Michelon et al., 2019). The exposure of fraud scandals (e.g., the Luckin fraud in China) challenges the bottom line of supervision and also reflects the situation that the existing information disclosure system is difficult to effectively contain fraud in the financial market. In fact, in the year of the IPO fraud, the company attracted more media attention and more positive media attitudes (Sun et al., 2021). The existing research on risk disclosure is mainly divided into two perspectives, i.e., risk effect and information effect (Adam-Müller & Erkens, 2020; Hope et al., 2016; Hussein et al., 2020; Li et al., 2019a, 2019b; Xia et al., 2022). According to the view of information effect, risk information can explain known risk factors and emergencies increase information transparency and reduce the risk perception of information users (Ada et al., 2021; Benamati et al., 2021; Hope et al., 2016). It is also to be argued that the more risk disclosure, the more risk perception of investors. At this time, investors are willing to see less risk disclosure. That is, investors are willing to see more risks disclosed. Under the risk effect perspective, information disclosure can enhance communication with the market and reduce information asymmetry between listed companies and investors (Adam-Müller & Erkens, 2020; Li et al., 2019a, 2019b). Although regulators have been improving the information disclosure system (Hoque & Mu, 2019; Zimmer et al., 2010), there are still problems such as vague information disclosure, false statements, and misleading fraud (Lobo & Zhao, 2013; Que & Zhang, 2019). This paper intends to explore the characteristics of enterprise risk disclosure and the kind of market reaction that different risk disclosure could have.

Several scholars (Hussein et al., 2020; Wasiuzzaman et al., 2018) have examined the link between IPO initial returns and information disclosure on the IPO prospectus. For the IPO opening day, the initial returns are the IPO underpricing rate. In similar literature, academics and researchers use the terms initial returns and IPO undervaluation rate interchangeably (Ritter & Welch, 2002). A large number of studies have studied the factors influencing the efficiency of the IPO capital market from the perspectives of IPO price limit policy (Kim et al., 2013), IPO venture capital (Que & Zhang, 2019), investor attention (Chang & Kwon, 2020), and directed share program (DSP) (Chong & Liu, 2020); however, they have not taken the risk information text into account. Although the existing literature has focused on the information value of risk disclosure text to some extent (e.g., Wei et al., 2019), there are still some deficiencies.

First, the risk information disclosed can provide accurate and valuable information to estimate the company's position and forecast future development (Alshirah et al., 2022). The existing literature on risk disclosure is still lacking, especially the analysis of risk disclosure text from investors' perspectives. Through manual semantic analysis of the content, we can find out and compare the norms of intangible information disclosure (Catalfo & Wulf, 2016). At the same time, the content of information disclosure also impacts decision-making (Zhang & Liu, 2020). Some studies have evaluated the degree of risk disclosure from the risk scope, simplicity, and uniqueness of the text (Sheng et al., 2021). Therefore, this paper analyzes the degree of information disclosure from a new perspective, that is, content and semantics.

Second, the existing research measures the risk disclosure either by manual reading and labeling or by exchange rating, which is time-consuming, labor-consuming, and subjective.

Many scholars have confirmed the importance of data mining and analysis methods for crisis management (Akter & Wamba, 2017). Due to the unpredictability of global events such as the COVID-19 pandemic, governments are using data mining techniques to prepare more innovative and proactive crisis management strategies for future uncertain crises (Park, 2021). Data mining technology has the advantages of being fast and accurate, which has opened up a new development direction for enterprise financial analysis and played an important role in financial crisis management (Shang et al., 2021). This paper uses data mining technology to analyze the prospectus and mine the risk disclosure in the report. This paper uses an unsupervised machine learning algorithm to solve this problem. Thirdly, studies have shown that improving enterprise information transparency can increase the stock market's liquidity (Choi & Jung, 2021). At the same time, the transparency of enterprise information also affects the trust of investors (Han et al., 2021). Investors have increased their demand for information transparency because they can make better decisions based on the disclosures provided by their companies (Zia-ur-Rehman et al., 2021). This paper takes the regulatory role of enterprise information transparency into consideration.

Finally, this paper further analyzes the characteristics of risk disclosure of illegal enterprises in the future from the perspective of risk aversion. Risk aversion arises from a perceptual bias that, given the limitations of the mental representation of the situation, represents the optimal decision rule (Khaw et al., 2021). However, most decision-makers tend to avoid risks, and the risk-neutral approach cannot meet their needs (Huang et al., 2021a, b). In the case of risk aversion, a preference function that is higher than the expected return and risk is usually required. The level of enterprise risk aversion impacts the optimal decision and its profit (Kouvelis et al., 2021). Therefore, the appeal of safer options depends on decision-makers' risk aversion (Calsamiglia et al., 2021).

As risk disclosure serves as the only official information channel for IPO companies to open to investors, this study explores the following research questions:

1. What are the characteristics of the risk disclosure text of the prospectus?
2. Do they follow the information effect or the risk effect, respectively?
3. What are the characteristics of the risk disclosure of illegal enterprises in the future?

To answer these questions, this paper uses a text analysis method to analyze the risk information of the prospectus and empirically tests the influence of the text content and semantic features of risk disclosure on the efficiency of the capital market.

This study contributes to the literature by empirically examining the role of risk disclosure texts in improving capital market efficiency and guiding value investment. Firstly, this paper empirically analyzes the impact of prospectus risk disclosure on capital market efficiency by calculating IPO pricing efficiency and regression of independent variables from two aspects of the semantic and content characteristics of the prospectus. This is conducive to a more comprehensive estimation of investors' response to the IPO enterprise risk disclosure text and deeply analyzes the effect of this response (information or risk effect). Secondly, as a supplementary channel of information resources, enterprise information transparency plays a moderating role in the "information effect" and "risk effect" of risk disclosure. We find that risk disclosure is cross-influenced by information effect and risk effect, and enterprise information transparency directly affects enterprise risk disclosure degree. Thirdly, from the perspective of risk measurement, this study breaks through the existing manual reading, labeling, or complex machine learning classification methods (Cheng et al., 2021; Srivastava & Eachempati, 2021). Based on an unsupervised machine learning algorithm, this paper constructs an unsupervised feature extraction model of risk disclosure text, which provides a new method for the research of feature extraction of risk disclosure text. Fourth, this paper

suggests that regulators should adopt different ways to require the risk disclosure degree of newly listed enterprises according to their specific information environment. According to the IPO performance of future illegal enterprises, a regression model of independent variables is constructed. We find that the richness of information disclosure of future illegal enterprises can be reduced to avoid the inspection of regulators.

The remainder of this paper is organized as follows. Section 2 is the literature review and research hypotheses. Section 3 outlines the research design, including sample selection measures, variable design, and the establishment of the measurement model. Section 4 introduces the detailed analysis of empirical results. Section 5 shows the results of the robustness test. Section 6 discusses the empirical results. Section 7 highlights the practical and theoretical contributions of the paper as well as its limitations.

2 Literature review and research hypothesis

2.1 IPO pricing efficiency

As an important part of IPO, pricing has always been one of the key issues of the capital market (Gao et al., 2019; He et al., 2019). The existing literature mainly has studied the reasons for IPO pricing efficiency from the perspectives of information asymmetry, enterprise nature, behavioral finance, and government regulation (Huang et al., 2019a, 2019b; Jog et al., 2019; Liu et al., 2019; Rathnayake et al., 2019; Xuan et al., 2019). According to the view of information asymmetry, poor information disclosure may interfere with the information environment, resulting in a decline in pricing efficiency (Kao et al., 2020); and good information disclosure is conducive to pricing efficiency in the IPO market (Zhou & Sadeghi, 2019).

Although the previous studies have explored the causes of the poor pricing efficiency from the perspectives of investor attention (Chang & Kwon, 2020), regulatory interactions (He & Fang, 2019), and market maker competition (Farooq & Hamouda, 2016), there still exists a lack of in-depth analysis of risk disclosure. Yao and Zhao (2016) argue that the underpricing rate on the first day of listing can effectively reflect the efficiency of asset pricing. The efficiency of good asset pricing represents positive feedback from investors on information disclosure. The lower the underpricing rate is, the higher the pricing efficiency is, and the more positive the market reaction is. Therefore, from the perspective of prospectus risk disclosure, we can understand the internal influence mechanism of IPO pricing efficiency.

2.2 Risk disclosure

In the capital market, information disclosure is an important way to mitigate information asymmetry. Careful investigation of risk disclosure may help weaken investors' risk choices (McGuinness, 2019). Güçbilmez and Briain (2020) contend that investors with more information should get higher returns. But different information disclosures have different effects. There is no consistent conclusion on the role of risk information disclosure in the existing literature. The current research is mainly divided into two perspectives: risk effect and information effect (Adam-Müller & Erkens, 2020; Hope et al., 2016; Hussein et al., 2020; Li et al., 2019a, 2019b). To explore the true effect of risk disclosure, this article extracts and analyzes the characteristics of risk disclosure text from the perspectives of semantics and content.

The risk effect view is that the disclosure of "bad news" could bring vicious feedback. However, if companies disclose information honestly, but the gains are not worth the loss, listed companies would choose to hide "bad news" when disclosing information (Jin et al., 2021), which could lead to "adverse selection" problems in the IPO market and exacerbate risk information. This is why many listed companies use "good news, but not bad news" to whitewash and disclose information (Lo et al., 2017). Besides, hiding risk information can also effectively avoid the punishment of the third-party supervision mechanism (Nefedova & Pratobevera, 2020). Risk disclosure in the prospectus impacts initial IPO returns (Hussein et al., 2020). Therefore, this paper studies the impact of risk disclosure on IPO underpricing from the perspective of risk effect. This paper puts forward the following hypothesis for risk disclosure from the perspective of risk effect, as shown in Fig. 1.

H_1A: When the semantic novelty of prospectus risk disclosure is higher, the risk perceived by investors is lower, and the underpricing rate will be lower.
H_1B: When the risk disclosure content of the prospectus is richer, the risk perceived by investors is lower, and the underpricing rate will be lower.

The information effect view holds that the disclosure of risk information will bring positive feedback (Kamal, 2021). According to the information effect viewpoint, the content of risk disclosure helps reveal known risk factors, alleviating information asymmetry and enabling investors to have specific risk estimations (Li et al., 2019a, 2019b). The view of the information effect of risk disclosure holds that risk disclosure increases the supply of information, reduces the asymmetry of information, and easily wins the trust of investors, which may trigger a positive market response (Huo et al., 2022). The risk effect view of risk

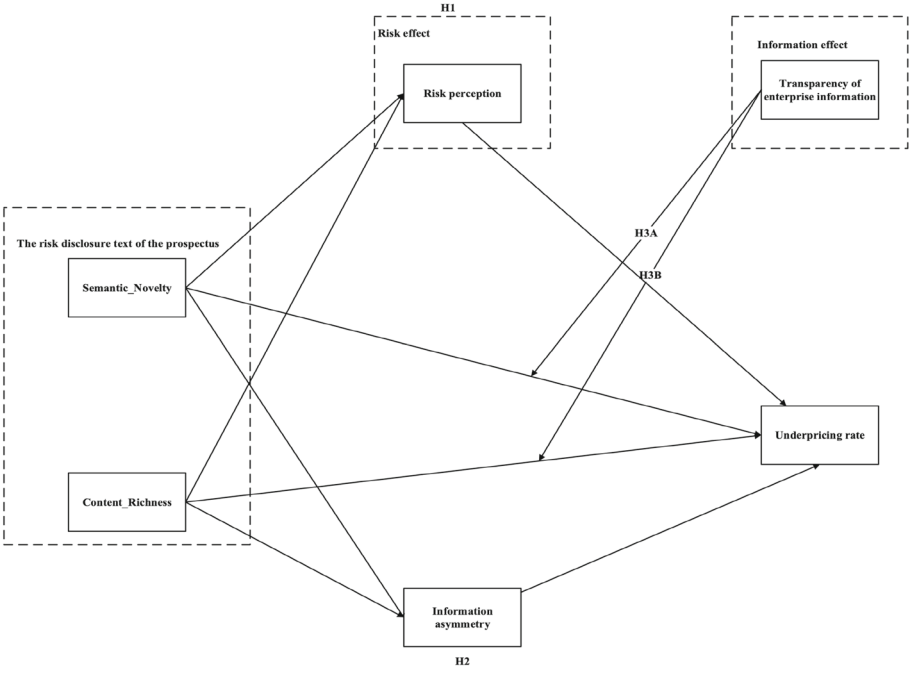

Fig. 1 Research model

disclosure holds that risk information can reveal unknown risk factors, enhance investors' risk perception, and trigger their fear of unknown risks (Campbell et al., 2014). Regulators such as the Securities and Exchange Commission have also shown interest in the quality of security risk disclosure (Cheong et al., 2021). Theoretical work usually predicts a negative correlation between disclosure and risk premium, where additional disclosure reduces estimated risk or information asymmetry (Ellahie et al., 2022). Therefore, this paper studies the impact of risk disclosure on IPO underpricing from the perspective of information effect. This paper puts forward the following hypothesis for risk disclosure from the perspective of information effect, as shown in Fig. 1.

H_2A: When the semantic novelty of prospectus risk disclosure is higher, the degree of information asymmetry will be lower, and the underpricing rate will be lower.
H_2B: When the risk disclosure content of the prospectus is richer, the degree of information asymmetry will be lower, and the underpricing rate will be lower.

2.3 Information environment

Transparency of the information environment refers to the extent to which external investors have access to the company's information. To respond to the concerns of external investors, enterprises disclose themselves, thus affecting the company's information environment (Xue et al., 2020). Loh and Stulz (2018) show that investors are more dependent on analysts' research when the market is uncertain. It has been found that risk information is often obscure and needs to be interpreted by professional analysts (Spence et al., 2020). Analysts' prediction of some quantitative indicators of the tracked enterprises can bring additional information to investors (Gu et al., 2019; Call et al., 2013). When more analysts track an enterprise, it could receive more attention. As a result, more information about the company can be revealed to the outside world, and the company's information transparency would be higher. Therefore, this paper tracks the number of analysts as an indicator to measure the transparency of enterprise information.

Many studies have shown that the information environment of a company is the main cause of information transmission (Farooq & Hamouda, 2016), and a good information environment is useful in stabilizing the financial market (Papadamou et al., 2017). As a matter of fact, there are certain prerequisites for the impact of information disclosure on the capital market (Albring et al., 2020). Huang et al., (2019a, 2019b) find that the high transparency of the market information environment is conducive to easing information asymmetry. When the information environment is uniquely examined, the negative impact of ESG ratings on IPO underpricing is more pronounced in countries with more transparent financial disclosure, higher liability standards, and stronger shareholder protection (Baker et al., 2021). Therefore, we assume that corporate information affects the degree of information disclosure. We study enterprise information transparency as a moderating variable and study the impact of underpricing rates on risk disclosure. Accordingly, the following hypotheses are proposed.

H_3A: When the transparency of enterprise information is low, the underpricing rate reaction to the semantic novelty of risk disclosure in the prospectus will be significantly strengthened.
H_3B: When the transparency of enterprise information is low, the underpricing rate reaction to the richness of risk disclosure's content in the prospectus will be significantly strengthened.

Based on the above assumptions, this paper discusses how the information and risk effects of risk disclosure play their respective roles, reveals the relationship between initial IPO returns and semantic novelty and content richness, and the impact of corporate transparency on semantic novelty and content richness. The research model is shown in Fig. 1.

3 Research design

3.1 Sample selection and data source

Since the reform and opening up, China's economy has witnessed sustained and rapid development, national wealth has grown rapidly, and GDP has ranked second in the world. China's capital market has made some achievements in development over the past 40 years. From the perspective of the proportion of China's capital market in the global market, the share of stocks, bonds, and asset management markets reached 13%, 15%, and 9% at the end of 2020, respectively, second only to the United States. The international status of the capital market has increased rapidly and gradually, matching China's economic status. As the second largest stock market in the world, it is reasonable to select the data as a research sample. In terms of representativeness, China's stock market has initially formed a multi-level capital market, with a variety of trading platforms, including small and medium-sized boards, main board, equity trading market, gem, etc. Although it started late, it has developed rapidly and has become relatively mature. From the perspective of uniqueness, due to the late start, there still exists a lag in the legal system construction. The credit system is, to a certain extend, imperfect and the economic systems are different, which makes China's stock market unique. From the above two aspects, the data of the Chinese stock market selected in this study has its research value.

The prospectus and related variables are from website "Oriental Wealth" and "Financial Circle". This paper uses prospectuses from 2009 to 2019 as risk disclosure samples. Respectively, the prospectuses for Oriental Fortune are from 2010 to the present, and the prospectuses for Financial Circle are from 2009 to the present. Corporate governance-related data and corporate financial data are extracted from the China Center for Economic Research (CCER) financial database. The company characteristic data comes from the China Stock Market & Accounting Research (CSMAR) database and is proofread with CCER database data to ensure the accuracy of IPO enterprise data.

Referring to the existing literature, we deal with the initial samples according to the following principles: (1) eliminating the missing samples after data matching, (2) removing samples of outliers, (3) eliminating ST-listed companies because the financial data of ST listed companies can only be disclosed after certain processing, which has no reference value, (4) eliminating A-share listed companies that issue H shares to avoid the impact of various regulatory rules, referring to the research of Chen et al. (2018), (5) eliminating backdoor listed companies as Lee et al. (2019) found that the performance of backdoor listed companies in China was significantly better than that of other IPO companies before and after listing, and (6) defining the sample industry as the manufacturing industry, considering the impact of different industry characteristics and information disclosure regulations. After the above treatment, 1297 observations are obtained. Considering the influence of plate differences, 826 observations are obtained. Considering the characteristics of the unsupervised model, 101 companies' prospectus risk disclosure text data is selected. There are 21 companies listed in the Shenzhen A-share market and 80 companies listed in the Shanghai A-share market.

3.2 Measurement of risk disclosure level

Information disclosure of listed companies can be divided into quantitative data and text. Unstructured text is of great significance to the analysis of the stock market and financial decision-making (Chan & Chong, 2017). There are five ways to measure the quality of information disclosure. First, the text length is used to measure the quality of enterprise information disclosure. When companies discuss more content in the financial statements, the quality of information disclosure is relatively high. Nowadays, the capital market risk disclosure text is old-fashioned and whitewashed (Lo et al., 2017). The length of the text has been difficult to measure the level of information disclosure. The second is to build an indicator system based on specific data to calculate the information disclosure index (Al-Hadi et al., 2019). Because of the complexity of specific indicators, indirect measurement of the level of information disclosure through specific indicators has lost the accuracy and effectiveness of the disclosure text itself. The third is to measure the quality of information disclosure by using the evaluation scores of information disclosure on platforms such as exchanges (Adam-Müller & Erkens, 2020). On the one hand, third-party organization scoring plays an important role in information valuation (Grassa et al., 2020). On the other hand, it ignores the subtle differences between listed companies. Besides, the market environment is changing with each passing day, and the quantitative index system of information disclosure needs to be verified, so it is lack of accuracy. The fourth is to manually read the risk disclosure of the prospectus or annual report and manually measure and mark the risk information (Shivaani et al., 2020; Yao & Zhao, 2016). Although this approach eliminates the barriers of other approaches to the semantic understanding of risk disclosure, it consumes too much labor. At the same time, it is subjective and time-consuming. Fifthly, many researchers use computer softwares to extract risk keywords (Ibrahim & Hussainey, 2019), mood and nature (Shivaani & Agarwal, 2020), semantic tone (Gonzalez et al., 2019), and other features to measure the intensity of risk information disclosure. In practice, a sentence may contain a lot of subjective information or intentions (Mai & Le, 2020) and different risk characteristics and require different risk management strategies (DuHadway et al., 2017). Therefore, it is challenging to express risk information with a single feature. This paper refers to the fifth way, using unsupervised machine learning to deal with risk disclosure. The text's semantic novelty and content richness are measured to explore its impact on the efficiency of the capital market.

3.3 Related algorithm design

3.3.1 Text vectorization model based on Neural Network–Word2vec

Word2vec (Mikolov et al., 2013) and Doc2vec (Le & Mikolov, 2014) models represent deep text representation models. Compared with the traditional text representation model, it can transform words, sentences, or even paragraphs into fixed dimension vectors more fully combined with text features. This method overcomes the limitations of traditional methods in mining single text features and has been widely used in abnormal comment detection (Chang et al., 2018), candidate recommendation (Kim et al., 2019), and other fields. Word2vec, through training, can simplify the processing of text content into vector operation in k-dimensional vector space, and the similarity in vector space can be used to represent the semantic similarity of text. Therefore, the word vector output from Word2Vec can be used to

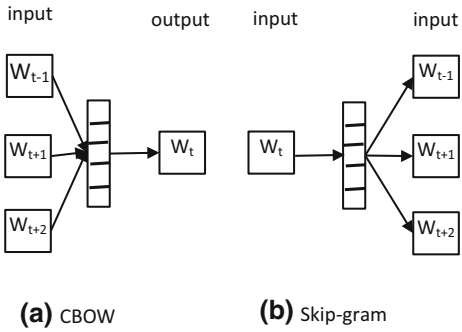

Fig. 2 CBOW and Skip-gram network structures of Word2vec model

do a lot of NLP-related work, such as clustering, finding synonyms, part of speech analysis, and so on. This section will explain the principle of the basic model word2vec.

Word2vec model is a word vector mapping model, which is mainly divided into two network structures: CBOW (Continuous Bag of Words) and skip-gram. CBOW is to predict the center word through context; Skip-gram is to predict the context through the center word. Thus, these two different approaches only change the way inputs and outputs are managed, but in any case, the network does not change, and the training always occurs between single pairs of words (as onehot in the inputs and outputs) (Di et al., 2021). Then, the Word2Vec model is used to use different hyperparameter training corpora, including vector dimension, context window size, and training iterations. The purpose of training with different hyperparameters is to fine-tune the model and determine the best embedding for synonym extraction (Al-Matham & Al-Khalifa, 2021). The principle is shown in Fig. 2.

3.3.2 Doc2vec

Based on Word2vec, Mikolov et al. (2013) developed an unsupervised method to map sentences or complete paragraphs to vector spaces of corresponding dimensions, namely the Doc2vec model. Doc2Vec has been used for sentiment analysis, though not as often as Word2Vec. We used Gensim's Doc2Vec implementation with the default hyperparameters (Mishra et al., 2019). Doc2Vec model is a common neural network embedding method in natural language processing, which is used to vectorize words and documents from context. Compared with TF-IDF, LDA, and Word2Vec models, Doc2Vec model has the highest accuracy in functional area classification (Niu & Silva, 2021a, b). The model includes PV-DM (Distributed Memory Model of Paragraph Vectors) and PV-DBOW (Distributed Bag Of Words Model of Paragraph Vectors), similar to CBOW and Skip-gram in Word2vec, respectively.

The PV-DM model is used through the context and the corresponding paragraph vector to predict the probability of the possible central word. The schematic diagram of the PV-DM model is shown in Fig. 3a. The objective function of the PV-DM model is the maximum mean log-likelihood function, as shown in formula 1:

$$\frac{1}{T}\sum_{t=k}^{T-k}\log p(w_t|d_t, w_{t-k}, \ldots w_{t+k}), \tag{1}$$

where T is the number of words, k specifies the size of the sliding window, d_t is the paragraph vector of the current sentence, and the prediction task is completed by a softmax classifier. As shown in formulas 2 and 3,

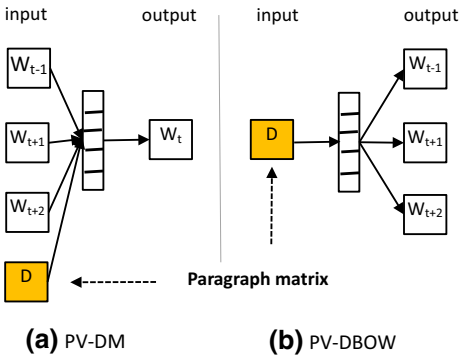

Fig. 3 PV-DM and PV-DBOW network structures of Word2vec model

$$p(w_t|d_t, w_{t-k}, \ldots w_{t+k}) = \frac{e^{y_{wt}}}{\sum_i e^{y_i}} \tag{2}$$

$$y = Uh(d_t, w_{t-k}, \ldots w_{t+k}; W, D) + b, \tag{3}$$

where the h function is the concatenation or average of context words, and b is the intercept.

The PV-DBOW model is another method of training paragraph vectors in Doc2vec to randomly extract text windows from which words are extracted as words to be predicted. The purpose is to complete a specific classification task through the known paragraph vector. The schematic diagram of the model is shown in Fig. 3b. The objective function of the model is as shown in formula 4,

$$\frac{1}{T}\sum_{t=k}^{T-k} \log p(w_{t-k}, \ldots w_{t+k}|d_t). \tag{4}$$

In general, the Doc2vec model can effectively extract text features and abstract the text content into a fixed dimension vector so that the similarity of the text on the semantic level can be calculated and expressed in the vector space. Therefore, We use the most advanced embedding algorithms Doc2Vec as learning techniques. The algorithm builds word and document embeddings in an unsupervised manner (Chen & Sokolova, 2021).

3.3.3 Cosine similarity

Common distance measurement methods include Euclidean distance and cosine distance. Considering the characteristics of the model studied in this paper, the cosine similarity still maintains the property of "1 when the same, 0 when orthogonal, and 1 when the opposite" in the high dimension (Alshammeri et al., 2021). The value of Euclidean distance is influenced by dimension. In addition, Euclidean distance represents the absolute difference in numerical value, while cosine distance represents the relative difference in direction. By comparing different similarity measurement methods, it is found that neural network technologies (Word2vec, Doc2vec, Law2Vec) can learn the embedding effect in the task as well as other technologies (Mandal et al., 2021). When calculating document similarity, we should consider the semantic similarity of the text, not the text itself. Therefore, this paper uses a Doc2vec model to calculate the similarity between risk disclosure documents (Niu & Silva, 2021a, b). The specific calculation method is shown in Formula 5 as follows:

$$Similarity(firmA, firmB) = \frac{\overrightarrow{\text{firm}A} \cdot \overrightarrow{\text{firm}B}}{(|\overrightarrow{\text{firm}A}| * |\overrightarrow{\text{firm}B}|)} \tag{5}$$

3.4 Variable design and model

3.4.1 Dependent variable: IPO pricing efficiency

Referring to the research of Yao and Zhao (2016), this paper uses the IPO underpricing rate on the first day as an index to measure IPO pricing efficiency. IPO underpricing rate on the first day is defined as the difference rate between the closing price and the issuing price of the stock on the first day. If the index is positive, it means IPO underpricing. If it is negative, it means IPO premium. Some scholars have pointed out that information asymmetry is an important reason for IPO underpricing (Kao et al., 2020). Therefore, this paper uses the IPO underpricing index as the explanatory variable to study the impact of risk information disclosure on the securities market. We use the pricing efficiency of existing shares IR = (P1−P0)/P0 (Chivianti & Sukamulja, 2021), where P1 is the closing price of shares on the issue date, and P0 is the issue price.

3.4.2 Independent variable: semantic novelty and content richness of risk disclosure texts

Semantic novelty Based on the vector space model and Python built-in a Gensim library of Python, this paper uses the doc2vec model to extract semantic vectors from risk disclosure texts after preprocessing and word segmentation. Compared with the traditional way of extracting text vectors, doc2vec brings word-order information into the model, which makes the results more accurate. Then we calculate the semantic cosine similarity between the risk information disclosure texts of the prospectus to measure the semantic novelty of the risk information disclosure texts (Alshammeri et al., 2021). For risk disclosure document set M and test document set N, this paper uses an unsupervised machine learning algorithm to calculate the semantic similarity between N and M. The semantic similarity list is constructed, and the overall similarity of the document set is obtained by processing the average value of the list value to describe the semantic novelty of the risk disclosure text. The process is shown in Fig. 4. The measurement of semantic similarity between any risk disclosure documents can be expressed by the cosine similarity of two vectors.

Content richness In this paper, the content richness of the risk disclosure of the prospectus is measured by the number of risk types in the risk disclosure text (Elshandidy & Zeng, 2022). Compared with the way of manual reading in the study of Yao and Zhao (2016), it is more intuitive and objective to describe the richness of the risk disclosure content of the prospectus according to the number of risk types of the disclosure text and the way of data acquisition is more convenient.

Control variable We control the company characteristic variables that may affect the pricing efficiency on the first day of listing, such as enterprise age, market system environment, registered capital, governance structure, loss situation, irrational emotions of investors, etc. Among them, the company size is the logarithm of total assets in the year of listing. The cost loss of large-scale institutions after financial fraud is far greater than that of small-scale

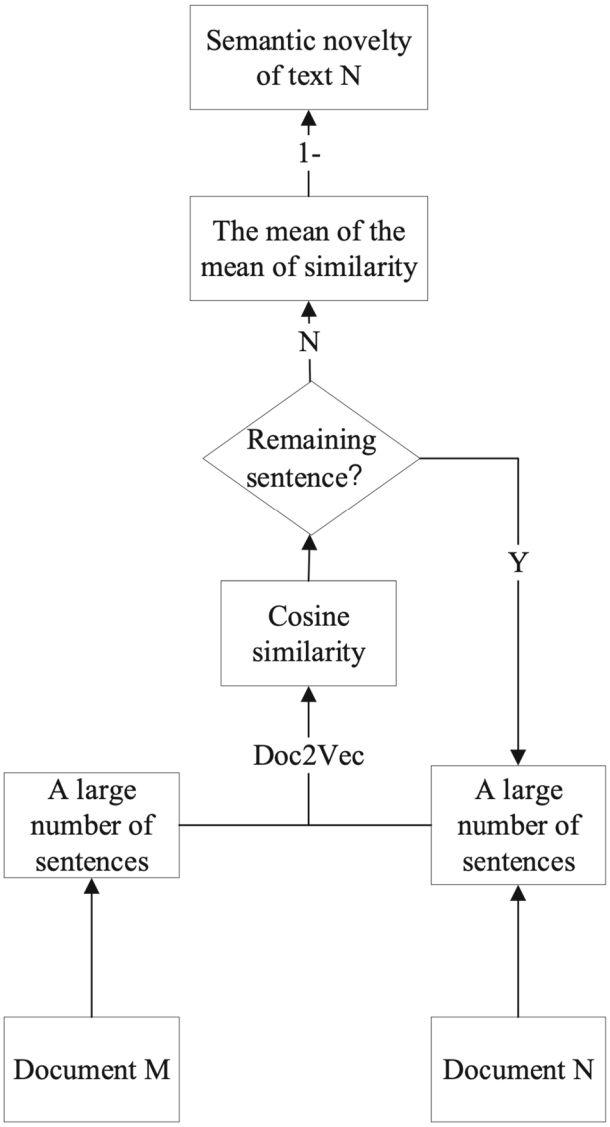

Fig. 4 Flow chart of the semantic novelty of text

institutions, so it may have an impact on the information disclosure effect of listed companies (Kabir et al., 2020).

In terms of governance structure, the nature of enterprise equity (State) is selected as the characteristic variable, and the state-owned holding is 1, while the non-state-owned holding is 0. Previous studies have shown significant economic differences in underpricing in different financial market environments, proving the importance of location selection in the listing process (Marcato et al., 2018). Bhardwaj and Imam (2019) also verify the importance of the external market environment for information disclosure. Therefore, for the institutional

market environment, this paper chooses the "market-oriented index of China's regions" in the 2013 report of China's provincial enterprise operating environment index compiled by Wang et al. (2013) as the alternative variable of the institutional market environment. All variables are defined and shown in Table 10 in Appendix.

After controlling the company characteristic variables that may affect the pricing efficiency on the first day of listing, such as the age of the enterprise, the market system environment, the registered capital, the governance structure, and the loss situation, this paper explores the cross-action mechanism of information effect and risk effect on the IPO market from the perspective of semantics and content. Therefore, the following models are built, as shown in formulas 6, 7, and 8:

$$\begin{aligned} Pricing\_Efficiency(Semantic) &= \alpha + \beta_1 Asset + \beta_2 Age + \beta_3 SOE \\ &+ \beta_4 Market\_Index + \beta_5 \mathrm{Re}gistered\_Capital \\ &+ \beta_6 Turnover\_rate + \beta_7 LOSS + \beta_8 Semantic\_Novelty + \beta_9 block + \sum Control + \varepsilon, \end{aligned} \tag{6}$$

$$\begin{aligned} Pricing\_Efficiency(Content) &= \alpha + \beta_1 Asset + \beta_2 Age + \beta_3 SOE \\ &+ \beta_4 Market\_Index + \beta_5 \mathrm{Re}gistered\_Capital \\ &+ \beta_6 Turnover\_rate + \beta_7 LOSS + \beta_8 Content\_Richness + \beta_9 block + \sum Control + \varepsilon. \end{aligned} \tag{7}$$

$$\begin{aligned} Pricing\_Efficiency(Semantic * Content) &= \alpha + \beta_1 Asset + \beta_2 Age + \beta_3 SOE \\ &+ \beta_4 Market\_Index + \beta_5 Registered\_Capital \\ &+ \beta_6 Turnover\_rate + \beta_7 LOSS + \beta_8 Semantic\_Novelty + \beta_9 Content\_Richness \\ &+ \beta_{10} Semantic\_Novelty * Content\_Richness + \beta_{11} block + \sum Control + \varepsilon. \end{aligned} \tag{8}$$

4 Analysis and results

4.1 Descriptive statistical analysis and text heterogeneity evaluation

From the perspective of semantics and content, this paper explores the cross-influence mechanism of information and risk effects on the IPO market. First, this paper makes descriptive statistics on some statistical variables. According to the statistical results in Table 1, the semantic novelty of the risk disclosure text of the IPO prospectus is generally high. In terms of text content, there are about 14 risk disclosure statements in the IPO prospectus, which is significantly higher than the seven risk categories required in the standards for the content and format of information disclosure by companies offering securities to the public No. 1—prospectus issued by CSRC in 2015. It can be seen that with the increasing demand of public investors for the openness of IPO enterprises, the willingness for risk disclosure of IPO enterprises gradually increases. We also classified according to different stock plates. Table 2 shows descriptive statistics of each variable. We found that there were differences in each variable of enterprises in different stock plates.

This paper uses an unsupervised machine learning algorithm to extract semantic novelty features of the risk disclosure document set. The distribution of semantic novelty of risk disclosure is depicted in Fig. 4, and the distribution of content richness of risk disclosure is shown in Fig. 5 and Fig. 6. According to the distribution diagram in Fig. 4, at the semantic

Table 1 Descriptive statistics of statistical variables

Variable	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
Semantic_Novelty	101	0.730	0.0360	0.649	0.848
Content_Richness	101	14.39	5.857	5	33
block	94	7.564	8.992	0	37
Asset	101	21.09	1.102	19.05	24.54
Age	101	12.29	6.034	1.126	38.10
SOE	101	0.881	0.325	0	1
Market_Index	101	3.120	0.0738	2.860	3.440
Registered_Capital	101	19.05	1.054	17.62	22.61
Turnover_rate	101	− 1.714	13.61	− 97	0.894
underprice	101	0.395	0.202	− 0.0755	1.364

level, the novelty distribution mainly focuses on the position of 0.70–0.75. It can be seen that the semantic novelty of the risk disclosure text of the IPO prospectus is generally high. Figure 5 implies that the risk categories of IPO enterprise risk disclosure mainly focus on 13–16, which is significantly higher than the seven risk categories required in the standards for the content and format of information disclosure by companies offering securities to the public No. 1—prospectus issued by CSRC in 2015. It can be seen that the willingness of IPO companies to disclose risks has gradually increased (Fig. 6).

4.2 Empirical results

Table 3 shows the results of risk disclosure on the IPO's first-day market performance. The results show that: At the semantic level, it can be seen from model 1 that semantic novelty (Semantic_Novelty(p = − 1.003**)) is inversely proportional to the underpricing rate. In other words, the higher the semantic novelty of IPO prospectus risk disclosure, the lower the risk perception of investors, the lower the information asymmetry between enterprises and investors, the lower the underpricing rate, and the lower the first-day market returns (Hussein et al., 2020). This result follows the risk effect and information effect of risk disclosure (Adam-Müller & Erkens, 2020; Hope et al., 2016; Hussein et al., 2020; Li et al., 2019a, 2019b). So this validates hypothesis 1A and hypothesis 2A. However, according to Model 1 and Model 2 in Table 4, taking a step to look at a level group of properties of the stock plate, we can see that semantic novelty has a significant negative impact on IPO underpricing rate, mainly for Shenzhen A-share listed enterprises, but has no significant impact on Shanghai A-share listed enterprises.

The above results may be caused by the different trading systems of the two stock markets. The trading rules of the Shenzhen stock Market are collective bidding, while the trading rules of the Shanghai Stock Market are continuous bidding. In the last 15 min, Shenzhen Stock Market has bidding time. As there is no bidding time in the Shanghai exchange market, investors in the Shenzhen exchange market can operate in the last fifteen minutes. Investors will feel tired of long-term decision-making, thus affecting the accuracy of investment (Ma et al., 2021), which leads to the instability of closing prices and increases the volatility of

Table 2 Descriptive statistics classified by the stock plates

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Shenzhen A shares					Shanghai A shares				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Semantic_Novelty	21	0.728	0.0300	0.697	0.821	80	0.731	0.0375	0.649	0.848
Content_Richness	21	15.52	3.803	9	23	80	14.09	6.271	5	33
block	21	11.19	9.169	0	27	73	6.521	8.726	0	37
Asset	21	20.25	0.676	19.05	21.94	80	21.31	1.088	19.52	24.54
Age	21	7.538	4.547	1.756	18.53	80	13.54	5.767	1.126	38.10
SOE	21	0.857	0.359	0	1	80	0.887	0.318	0	1
Market_Index	21	3.105	0.0912	2.980	3.440	80	3.124	0.0686	2.860	3.250
Registered_Capital	21	18.36	0.509	17.64	19.34	80	19.23	1.088	17.62	22.61
Turnover_rate	21	0.618	0.295	0.000240	0.894	80	− 2.326	15.26	− 97	0.873
underprice	21	0.432	0.341	− 0.0755	1.364	80	0.386	0.148	− 0.0725	0.639

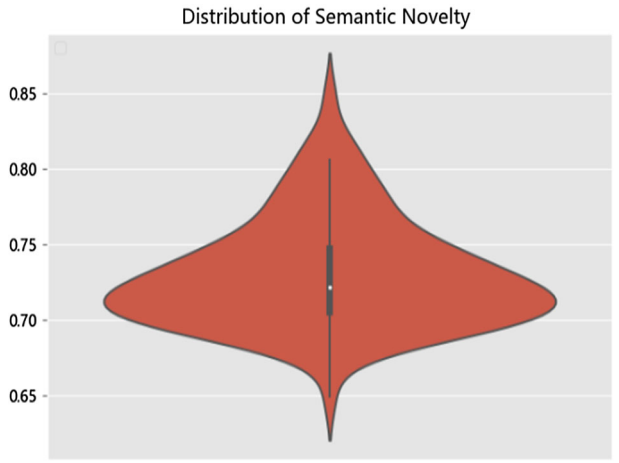

Fig. 5 Distribution of semantic novelty

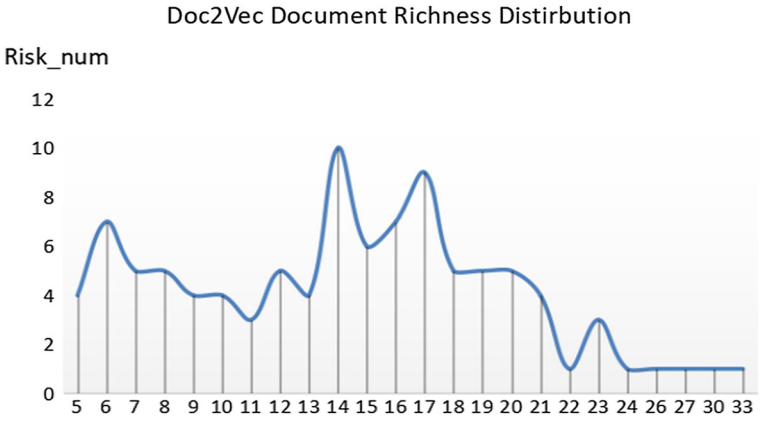

Fig. 6 Distribution of content richness

underpricing rate in the IPO market. As a result, the semantic novelty of Shenzhen A-share prospectuses has no significant effect on IPO underpricing.

At the content level, the results are consistent with those of Yao and Zhao (2016). It can be seen from model 2 in Table 3 that content richness (Content_Richness(− 0.009***)) is inversely proportional to the underpricing rate. In other words, the higher the richness of the risk disclosure in the prospectus, the lower the risk perception of investors, the lower the information asymmetry between enterprises and investors, the lower the underpricing rate, and the lower the first-day market returns (Ellahie et al., 2022). So this validates hypotheses 1B, and hypothesis 2B. The more types of risks disclosed in the IPO prospectus, the less information asymmetry between enterprises and investors (Elshandidy & Zeng, 2022). Hence, this leads to the easier gaining investor trust, the lower the underpricing rate (Li et al., 2019), and the more market returns on the first day. This result follows the information effect of risk disclosure. According to Models 3 and 4 in Table 4, taking a step to look by a level group

Table 3 Regression analysis of the novelty and richness of information disclosure to IPO underpricing rate

Variable	(1) IPO Underprice	(2) IPO Underprice
Semantic_Novelty	− 1.003**	
	(− 2.24)	
Content_Richness		− 0.009***
		(− 2.73)
Asset	− 0.006	0.011
	(− 0.16)	(0.30)
Age	0.008**	0.006**
	(2.21)	(2.08)
SOE	− 0.171	− 0.195*
	(− 1.55)	(− 1.85)
Market_Index	0.049	− 0.033
	(0.23)	(− 0.15)
Registered_Capital	− 0.072*	− 0.086**
	(− 1.73)	(− 2.21)
Turnover_rate	0.000	− 0.000
	(0.26)	(− 0.70)
LOSS	0.113**	0.019
	(2.56)	(0.56)
Constant	2.525**	2.117**
	(2.33)	(2.33)
Observations	101	101
R-squared	0.240	0.268
Result	Significant	Significant

Robust t-statistics in parentheses ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

of property of the stock plate, the richness of the risk disclosure statement has a significant negative effect on IPO underpricing rate for companies listed in Shenzhen and Shanghai A-shares. This further confirms that the richness of risk disclosure is negatively correlated with IPO underpricing rate.

According to Models 5 and 6 in Table 4, it can be concluded that the interaction between semantic novelty and content richness has a significant effect on IPO underpricing after grouping according to stock sectors. We conclude that the interaction between semantic novelty and content richness on IPO underpricing is significant across different groups. However, there are different interaction effects among unlisted stocks. In Shenzhen A-share listed companies, the interaction between semantic novelty and content richness has a positive correlation with IPO underpricing rate. In Shanghai A-share listed companies, the interaction between semantic novelty and content richness negatively correlates with IPO underpricing rate.

Moreover, this paper considers enterprise information transparency. We divide enterprise information transparency into high and low levels by the median enterprise information degree. According to Table 5, we also explore the moderating effect of enterprise information

Table 4 Grouping regression analysis of the novelty and richness of information disclosure to IPO underpricing rate

Variable	(1) IPO Underprice	(2) IPO Underprice	(3) IPO Underprice	(4) IPO Underprice	(5) IPO Underprice	(6) IPO Underprice
Semantic_Novelty	− 2.023	− 0.714*			− 21.349***	1.250**
	(− 0.71)	(− 1.99)			(− 3.17)	(2.31)
Content_Richness			− 0.040**	− 0.005**	− 0.845***	0.113***
			(− 2.18)	(− 2.15)	(− 3.51)	(3.96)
*Semantic_Novelty*Content_Richness*			− 0.101	0.042	1.100***	− 0.163***
			(− 0.91)	(1.21)	(3.36)	(− 4.11)
Asset	− 0.127	0.029	− 0.101	0.042	− 0.101	0.020
	(− 0.92)	(0.87)	(− 0.91)	(1.21)	(− 0.93)	(0.65)
Age	0.039	0.006*	0.028*	0.005*	0.017	0.005*
	(1.61)	(1.69)	(2.15)	(1.70)	(1.13)	(1.75)
SOE	− 0.323	− 0.017	− 0.330	− 0.030	− 0.109	− 0.058
	(− 1.00)	(− 0.26)	(− 1.47)	(− 0.47)	(− 0.47)	(− 1.14)
Market_Index	0.070	− 0.005	0.071	− 0.074	− 0.118	− 0.129
	(0.13)	(− 0.02)	(0.16)	(− 0.28)	(− 0.26)	(− 0.50)
Registered_Capital	0.008	− 0.073*	− 0.058	− 0.083**	− 0.098	− 0.058*
	(0.05)	(− 1.89)	(− 0.51)	(− 2.20)	(− 0.72)	(− 1.68)
Turnover_rate	0.418	− 0.000	0.075	− 0.001	0.059	− 0.000
	(1.51)	(− 0.59)	(0.26)	(− 1.40)	(0.22)	(− 1.03)
LOSS		0.082*		0.019		0.139***
		(1.93)		(0.68)		(3.38)
Constant	3.828	1.657*	3.966	1.337*	20.799**	0.629

Table 4 (continued)

Variable	(1) IPO Underprice	(2) IPO Underprice	(3) IPO Underprice	(4) IPO Underprice	(5) IPO Underprice	(6) IPO Underprice
	(1.23)	(1.88)	(1.46)	(1.69)	(3.06)	(0.67)
Observations	21	80	21	80	21	80
R-squared	0.489	0.285	0.609	0.290	0.752	0.371
Result	Non-significant	Significant	Significant	Significant	Significant	Significant

Robust t-statistics in parentheses ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$; Model 1,3 and 5 are the data of shenzhen A-share enterprises;Model 2,4 and 6 are the data of shanghai A-share enterprises

Table 5 Grouping regression results of the moderating effect of enterprise information transparency

Variable	(1) IPO Underprice	(2) IPO Underprice
Semantic_Novelty	− 1.065*	− 1.262
	(− 1.69)	(− 1.52)
Content_Richness	− 0.009*	− 0.012**
	(− 1.93)	(− 2.05)
Asset	− 0.035	0.040
	(− 0.63)	(0.57)
Age	0.005	0.006
	(1.58)	(0.92)
SOE	− 0.206	− 0.175
	(− 1.33)	(− 1.55)
Market_Index	− 0.083	0.119
	(− 0.40)	(0.26)
Registered_Capital	− 0.052	− 0.101
	(− 1.10)	(− 1.49)
Turnover_rate	0.016	0.000
	(0.11)	(0.32)
LOSS		0.131*
		(1.77)
Constant	3.402**	2.289
	(2.26)	(1.35)
Observations	50	51
R-squared	0.419	0.240
Result	Significant	Non-significant

Robust t-statistics in parentheses ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$; Model 1 is the company with low information transparency; Model 2 is the company with high information transparency

transparency. When corporate information transparency is low, market response to the semantic novelty and richness of risk disclosure in the prospectus will be significantly enhanced. When corporate information transparency is high, the market reaction to the semantic novelty and richness of risk disclosure in the prospectus has no significant effect. So this validates our hypothesis 3. We explain this as follows: corporate information transparency has a moderating effect on trust, and investors are significantly less interested in enterprises with high information degrees than those with low transparency (Zu et al., 2018). Therefore, the lower the transparency of corporate information, the fewer trust investors have in enterprises and the less trust they have in the documents disclosed by risks. As a result, when enterprises are listed in IPOS, investors will keep a conservative attitude when investing, which will restrain the IPO underpricing rate.

At the semantic level, risk disclosure is negatively correlated with IPO underpricing rate. At this point, the risk effect is significant. This means that when risk disclosure is improved, investors have more channels to obtain enterprise information and have certain predictions of enterprise risk. On the contrary, whitewash in semantics brings about positive effects (Lo

et al., 2017). At the content level, content richness has a negative correlation with the first-day underpricing rate, and the information effect is obvious. This means that when the richness of information disclosure is increased, investors are more inclined to make investment decisions according to the types of risks disclosed when there is no other way to obtain information.

4.3 Further analysis

The above part analyzes the internal influence mechanism of prospectus risk disclosure on IPO underprice from the perspective of cross-influence of risk effect and information effect. But why are companies willing to make full disclosure of risks before the relevant research results are available? Are enterprises with a better performance showing their good performance just to avoid undervalued evaluation in the lemon market? Or are the expedient measures adopted by companies with poor performance to avoid the regulatory risks caused by the decline of future performance? This paper will further study these questions next.

4.3.1 Risk disclosure and future violation risk

The existing research on financial market supervision almost focuses on punishment after the event. However, from the perspective of risk prevention, preventive supervision should play an important role. This paper examines the economic consequences of the opportunistic behavior of management from the perspective of the risk of corporate violations. Due to the delay of punishment result, date of illegal occurrence, and punishment judgment date, this paper defines whether the IPO company has illegal behavior fraud within three years after listing. We can find the possibility of fraud in the prospectus (Sun et al., 2021).

If the company's violations in that year, including false records, delayed disclosure, major omissions, fraudulent listing, insider trading, etc., are punished by the CSRC, the fraud is 1; otherwise, it is 0. According to the research results in Table 6, for enterprises with future irregularities, semantic novelty (Semantic_Novelty (P = 6.595)) has no significant effect on IPO underpricing rate, while content richness (Content_Richness (P = − 0.023)) has no significant impact on IPO underpricing rate. For enterprises with no future irregularities, semantic novelty (Semantic_Novelty (P = − 0.724*)) of the prospectus has a significant influence on the IPO underpricing rate. The content richness (Content_Richness (− 0.007***)) of the prospectus significantly affects the IPO underpricing rate. According to the regression results in Table 6, it is easier for future illegal enterprises to gain investor trust, improve pricing efficiency, and obtain market returns by manipulating risk disclosure information. This shows that in the future, illegal enterprises are more likely to obtain the reward of honest disclosure by manipulating the types of semantic novelty and risk disclosure (content richness) in the prospectus to avoid the risk and cover up the illegal issues. The semantic novelty and risk types (content richness) of the prospectus risk disclosure of excellent companies have a significant influence on the IPO underpricing rate. This shows that based on honest disclosure, to prevent the malignant market effect brought by semantic inflexibility, good enterprises will also improve semantic novelty, enhance the value content of information disclosure, and generate market returns.

Table 6 Grouping regression results of the classification of enterprises in violation of regulations in the future

Variable	(1) Underprice_punish	(2) Underprice_punish	(3) Underprice_no_punish	(4) Underprice_no_punish
Semantic_Novelty	6.595		− 0.724*	
	(1.79)		(− 1.83)	
Content_Richness		− 0.023		− 0.007***
		(− 0.07)		(− 2.85)
Asset	− 1.746	− 1.617	0.001	0.008
	(− 2.06)	(− 0.17)	(0.04)	(0.24)
Age	− 0.128	− 0.139	0.006*	0.005*
	(− 1.89)	(− 0.15)	(1.78)	(1.75)
SOE	− 1.788	− 1.498	− 0.042	− 0.081
	(− 3.15)	(− 0.39)	(− 0.51)	(− 1.00)
Market_Index	4.650	4.482	0.030	− 0.046
	(1.47)	(0.18)	(0.13)	(− 0.19)
Registered_Capital	2.505	2.403	− 0.060	− 0.068*
	(1.97)	(0.16)	(− 1.57)	(− 1.89)
Turnover_rate	− 1.396	− 1.458	− 0.000	− 0.000
	(− 2.41)	(− 0.17)	(− 0.19)	(− 0.83)
LOSS			0.091**	0.029
			(2.25)	(0.81)
Constant	− 26.383	− 21.608	1.892*	1.761*
	(− 1.55)	(− 0.14)	(1.80)	(1.96)
Observations	9	9	92	92

Table 6 (continued)

Variable	(1) Underprice_punish	(2) Underprice_punish	(3) Underprice_no_punish	(4) Underprice_no_punish
R-squared	0.967	0.933	0.195	0.225
Result	Non-significant	Non-significant	Significant	Significant

Robust t-statistics in parentheses ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

5 Robustness test

5.1 Changing the measurement of market pricing efficiency

To control the influence of the industry and the market, we need to redefine IPO pricing efficiency. Studies have shown that the return rate of the Shenzhen Stock Exchange component index from IPO pricing date to listing date is significantly different before and after the reform (Zhou et al., 2021). The P/E ratio is calculated using the closing price of the IPO day and earnings variables prior to the IPO (Makrominas & Yiannoulis, 2021). So, this paper adjusts the efficiency of IPO pricing, as shown in Formula 8, where RM is the weighted return rate of Shanghai and Shenzhen stock indexes on the first day of listing. According to Tables 7 and 8, after replacing the measurement method of market pricing efficiency, we find that the results are still significant, which shows the robustness of the results.

$$\mathrm{AdjIR} = \frac{1 + \mathrm{IR}}{1 + \mathrm{RM}} \tag{9}$$

Table 7 Robustness regression analysis of IPO pricing efficiency

Variable	(1) (Adj.) IPO Underprice	(2) (Adj.) IPO Underprice
Semantic_Novelty	− 0.031**	
	(− 2.23)	
Content_Richness		− 0.000**
		(− 2.61)
Asset	− 0.000	0.000
	(− 0.26)	(0.20)
Age	0.000*	0.000
	(1.76)	(1.54)
SOE	− 0.006	− 0.007*
	(− 1.63)	(− 1.89)
Market_Index	0.001	− 0.002
	(0.10)	(− 0.27)
Registered_Capital	− 0.002*	− 0.003**
	(− 1.79)	(− 2.26)
Turnover_rate	0.000	− 0.000
	(0.47)	(− 0.53)
LOSS	0.003**	0.000
	(2.27)	(0.12)
Constant	0.117***	0.104***
	(3.37)	(3.58)
Observations	101	101
R-squared	0.236	0.260
Result	Significant	Significant

Robust t-statistics in parentheses ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 8 Robust grouped regression analysis of IPO pricing efficiency

Variable	(1) (Adj.) IPO Underprice	(2) (Adj.) IPO Underprice	(3) (Adj.) IPO Underprice	(4) (Adj.) IPO Underprice	(5) (Adj.) IPO Underprice	(6) (Adj.) IPO Underprice
Semantic_Novelty	− 0.069	− 0.021*			− 0.680***	0.035**
	(− 0.78)	(− 1.93)			(− 3.21)	(2.12)
Content_Richness	(2)	(3)	− 0.001*	− 0.000**	− 0.027***	0.003***
			(− 2.16)	(− 2.01)	(− 3.54)	(3.66)
Semantic_Novelty * *Content_Richness*					0.035***	− 0.005***
					(3.39)	(− 3.80)
Asset	− 0.004	0.001	− 0.003	0.001	− 0.003	0.001
	(− 0.98)	(0.89)	(− 0.95)	(1.22)	(− 0.99)	(0.66)
Age	0.001	0.000*	0.001*	0.000*	0.001	0.000*
	(1.61)	(1.71)	(2.09)	(1.72)	(1.14)	(1.76)
SOE	− 0.010	− 0.001	− 0.010	− 0.001	− 0.003	− 0.002
	(− 0.98)	(− 0.43)	(− 1.44)	(− 0.61)	(− 0.44)	(− 1.23)
Market_Index	0.001	− 0.000	0.002	− 0.002	− 0.005	− 0.004
	(0.08)	(− 0.05)	(0.12)	(− 0.31)	(− 0.31)	(− 0.52)
Registered_Capital	0.001	− 0.002*	− 0.002	− 0.002**	− 0.003	− 0.002
	(0.12)	(− 1.89)	(− 0.46)	(− 2.18)	(− 0.66)	(− 1.66)
Turnover_rate	0.015	− 0.000	0.004	− 0.000	0.004	− 0.000
	(1.72)	(− 0.53)	(0.44)	(− 1.38)	(0.43)	(− 0.94)
LOSS		0.003**		0.001		0.004***

Table 8 (continued)

Variable	(1) (Adj.) IPO Underprice	(2) (Adj.) IPO Underprice	(3) (Adj.) IPO Underprice	(4) (Adj.) IPO Underprice	(5) (Adj.) IPO Underprice	(6) (Adj.) IPO Underprice
		(2.14)		(0.96)		(3.44)
Constant	0.154	0.081***	0.157*	0.071***	0.691***	0.051*
	(1.59)	(3.04)	(1.87)	(3.02)	(3.26)	(1.81)
Observations	21	80	21	80	21	80
R-squared	0.500	0.266	0.617	0.268	0.760	0.347
Result	Non-significant	Significant	Significant	Significant	Significant	Significant

Robust t-statistics in parentheses ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$; Model 1,3 and 5 are the data of shenzhen A-share enterprises;Model 2,4 and 6 are the data of shanghai A-share enterprises

5.2 Other robustness tests

In addition to changing the way to measure the efficiency of new share pricing, this paper also conducted many other ways of robustness tests. By changing the semantic training set of the unsupervised model (Niu & Silva, 2021a, b), for example, changing other IPO observation values (Massa & Zhang, 2021), multiple regression models (Rashidet al., 2014), and other ways to test the robustness, the verification results are consistent with our experimental results, which will not be described here. This paper finds that the risk disclosure semantics and content information of the prospectus has a significant impact on the IPO market performance on the first day.

For the important regulatory role of corporate information transparency as a channel for investors to obtain other information, the study on the regulatory role of corporate information transparency on IPO underpricing rate is significant (He & Fang, 2019). We re-classified the level of enterprise information transparency. 75% of enterprise information transparency level was delimited, as shown in Table 9 (Li & Zhu, 2021). After replacing the classification

Table 9 Robust grouped regression results of the moderating effect of enterprise information transparency

Variable	(1) IPO Underprice	(2) IPO Underprice
Semantic_Novelty	− 1.114**	− 0.007
	(− 2.30)	(.)
Content_Richness	− 0.009***	− 0.000
	(− 2.65)	(.)
Asset	0.015	− 0.000
	(0.39)	(.)
Age	0.006**	− 0.000
	(2.04)	(.)
SOE	− 0.219**	
	(− 2.11)	
Market_Index	0.082	
	(0.39)	
Registered_Capital	− 0.100**	0.000
	(− 2.45)	(.)
Turnover_rate	− 0.000	0.000
	(− 0.72)	(.)
LOSS	0.118***	
	(2.70)	
Constant	2.779**	0.450
	(2.64)	(.)
Observations	94	7
R-squared	0.332	1.000
Result	Significant	

Robust t-statistics in parentheses ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$; Model 1 is the company with low information transparency; Model 2 is the company with high information transparency

method of enterprise information transparency, we found that the results were still significant, indicating the robustness of the results.

Discussion

The paper examines whether the structure of the risk factor disclosure in an IPO prospectus helps explain the cross-section of first-day returns in a sample of Chinese initial public offerings (Grover & Bhullar, 2021). By constructing the IPO pricing efficiency model, the grouping regression analysis of risk disclosure and IPO pricing efficiency is carried out. The study finds that the structure of risk disclosure helps explain the first-day exchange returns of IPO (Sheng et al., 2021). The paper uses textual analysis to extract two aspects (semantic novelty and information richness) of language used in the risk disclosure segment of Chinese IPO prospectuses. From the perspective of risk effect, at the semantic level, the higher the semantic novelty of IPO prospectus risk disclosure, the lower the risk perception of investors, the lower the underpricing rate, and the lower the first-day market return. At the content level, the higher the richness of the risk disclosed in the IPO prospectus, the lower the risk perception of investors, the lower the underpricing rate, and the lower the first-day market return (Li et al., 2021a, b). From the perspective of the information effect, at the semantic level, the higher the semantic novelty of prospectus risk disclosure, the lower the degree of information asymmetry between enterprises and investors, the lower the underpricing rate, and the lower the first-day market return. In terms of content stratification, the richer the risk disclosure content of the prospectus is, the lower the information asymmetry between enterprises and investors, the lower the underpricing rate, and the lower the first-day market return (Hussein et al., 2020).

The paper examines the association of these two aspects with initial IPO returns. We also consider the interaction between the risk effect and the information effect. Under the same stock sector nature, the interaction between semantic novelty and content richness in risk disclosure prospectus and IPO underpricing rate is significant, but the adjustment direction of the interaction between different stock sector natures is different (Baker et al., 2021; Peng et al., 2021). This fills the previous interaction between risk effect and information effect on IPO underpricing. These results are interpreted as consistent with a risk effect and an information effect, respectively.

Based on an unsupervised machine learning algorithm, this paper explores the mechanism of information effect and risk effect on the short-term and long-term of the IPO market from the perspective of semantics and content (Engelen et al., 2020). To mark whether there are violations of future enterprises and explore the factors influencing the IPO underpricing rate of future enterprises. The prospectus is open to fraud (Sun et al., 2021). This paper constructs a regression model of the IPO underpricing rate of illegal enterprises in the future. It is found that the structure of risk disclosure helps to explain the characteristics of IPO prospectuses of future offending companies (Sheng et al., 2021). For enterprises with no future irregularities, the semantic novelty and content richness of prospectus significantly influence IPO underpricing rate, respectively. For enterprises with future irregularities, the prospectus's semantic novelty and content richness have no significant influence on IPO underpricing rate, respectively. Therefore, IPO underpricing rate is affected by the prospectus's semantic novelty and content richness in the short term. In the long run, whether enterprises violate rules is also influenced by semantic novelty and content richness. In the future, semantic novelty and content richness will have an effect on the IPO underpricing rate of non-offending enterprises, while semantic novelty and content richness have no effect on the IPO underpricing rate of offending enterprises (Chintya et al., 2020; Engelen et al., 2020).

First of all, based on the summary of the existing research on the risk effect and information effect of risk disclosure, this paper finds that these two views do not exist independently and

then explores the cross-action mechanism of the two effects (Adam-Müller & Erkens, 2020; Hope et al., 2016; Hussein et al., 2020; Li et al., 2019a, 2019b). The results show that at the semantic level, the risk disclosure text information of the IPO prospectus follows the risk effect of risk disclosure. At the content level, the results follow the information effect of risk disclosure. We find that risk effect and information effect interact with risk disclosure. Second, this paper considers the regulatory mechanism of enterprise information transparency (He & Fang, 2019). Under the same stock block nature, it is found that corporate information transparency can regulate the impact of semantic novelty and content richness of prospectus on IPO underpricing. This means that when the state regulates IPO underpricing, not only the structure of the prospectus can be reformed, but also the transparency of corporate information can be changed. The lower the transparency of the enterprise is, the higher the semantic novelty and content richness of the prospectus will be for investors, which reduces IPO underpricing. Investors are more inclined to make investment decisions according to the risk types of risk disclosure when there is no other access to information (Evans & Sun, 2021). Third, this paper classifies the samples to explore the risk disclosure characteristics of future illegal enterprises. It is found that the semantic and content of risk disclosure of future offending enterprises have no significant effect on IPO underpricing rate (Chintya et al., 2020; Engelen et al., 2020). This shows that future illegal enterprises are more likely to obtain the reward of honest disclosure by manipulating the content of the risk disclosure text of the prospectus to avoid the risk, cover up the problem of the violation, and then improve the market return. Based on honest disclosure, a good IPO company will also improve the level of semantic disclosure, increase the value content, and generate market returns to prevent the malignant market effect caused by a semantic template. Good companies generally disclose risk information voluntarily. Voluntary disclosure of enterprises has a better fund-raising capacity, so its own issue price is high, and the underpricing rate will be reduced (Bourveau et al., 2022).

The findings of this paper enrich the relevant research of agency theory, information asymmetry theory, and risk disclosure text analysis. First of all, from the perspective of short-term and long-term, this paper empirically studies the impact of prospectus risk disclosure on the efficiency of the capital market and complements the relevant research on risk disclosure, which is conducive to a more comprehensive and accurate estimation of the impact of IPO enterprise risk disclosure text on the efficiency of capital market in China (Engelen et al., 2020). Second, it summarizes the two viewpoints of the existing scholars (Adam-Müller & Erkens, 2020; Hope et al., 2016) and further explore the internal influence mechanism of risk disclosure text on the efficiency of the capital market, that is, how to play the role of risk effect and information effect, which help to alleviate the contradiction between frank communication and risk aversion. Third, as a supplementary channel of information resources, enterprise information transparency plays a regulatory role in the information effect and risk effect of risk disclosure (He & Fang, 2019). Fourth, from the perspective of risk measurement, it breaks through the existing manual reading, labeling, or complex machine learning classification (Catalfo & Wulf, 2016). Based on an unsupervised machine learning algorithm, this paper complements the research on quantitative feature extraction of risk disclosure text (Niu & Silva, 2021a, b).

The results of this study have some practical significance. Our research finds that the risk information of prospectus in China is uneven. This paper considers the significance of the IPO listing system and management from four aspects. For policymakers, the conclusion of this paper suggests that they need to readjust the structure of the prospectus and adopt a structure with high semantic novelty and high content richness to submit the prospectus and reduce the IPO underpricing rate of enterprises from the perspective of semantic novelty

and rich content (Grover & Bhullar, 2021). At the same time, enterprises are required to expand the submission of information on the transparency of corporate information in the prospectus (Choi & Jung, 2021). Although the willingness of IPO companies to disclose risks is gradually increasing, the phenomenon of manipulating the text of risk disclosure by whitewashing and other means still exists (Lo et al., 2017). This impacts the capital market's efficiency, especially the future illegal enterprises, which will lead to the "lemon market" effect. Therefore, the effective audit of third-party supervision should be continued based on further standardizing the content and form of risk disclosure of the IPO prospectus. For market regulation, the structure of the prospectus can be used to analyze whether an enterprise has been listed and whether it has the risk of future violations so as to prevent the short-term payment of huge funds by enterprises, to eliminate the arbitrage opportunities in the market, and to reduce the market stability and investors' property losses caused by the bankruptcy of enterprises (Sun et al., 2021). Future illegal enterprises can hide their problems by manipulating the risk text, which shows that the practice of risk disclosure in China does not reflect the original intention of the system (Jin et al., 2021), so it provides empirical evidence for the regulatory authorities to strengthen the supervision of risk disclosure of prospectus. The results of this paper can be used for reference for the regulatory authorities to supervise the information disclosure of the prospectus. Also, according to the specific information environment of newly listed enterprises, we should encourage them to reduce the degree of the risk disclosure model in different ways and transmit more information, which is conducive to the operation efficiency of the capital market (Baker et al., 2021). For portfolio managers, it is necessary to help investors reasonably analyze the future development of enterprises from corporate prospectuses, strengthen the promotion of new share issuance, and reduce the degree of information asymmetry between new shareasonably, select companies with high semantic novelty and content richness in the prospectus, and avoid pursuing companies with high IPO underpricing rates. For investors, from the perspective of risk effect and information effect, this paper makes use of semantic novelty and content richness in the prospectus to select enterprises with a low underpricing rate for investment (Wang & Song, 2021). It avoids speculative bubbles and improves the efficiency of capital allocation within the scope of the whole society. At the same time, but also consider the future of the enterprise, whether there is a violation of the impact of investment, to prevent their own property losses (Sun et al., 2021).

6 Conclusion and future research directions

This paper examines the content of risk disclosure of Chinese firms and how it affects IPO underpricing. The study uses the unsupervised machine learning approach to generate two major variables (Engelen et al., 2020): Semantic_Novelty and Content_Richness. Regression results indicate that Semantic Novelty is negatively related to IPO underpricing, and Content Richness is negatively associated with IPO underpricing. Therefore, the sentence novelty and richness of a good prospectus risk disclosure text are high. From the perspective of semantics and content, this paper explores the cross-influence mechanism of risk effect and information effect on the IPO market. At the semantic level, the risk disclosure text information of the IPO prospectus follows the risk effect of risk disclosure. At the content level, the results follow the information effect of risk disclosure. The interaction between risk effect and information effect on risk disclosure under the nature of the same stock plate (Adam-Müller & Erkens, 2020; Hope et al., 2016; Hussein et al., 2020; Li et al., 2019a, 2019b). In the future,

illegal enterprises can hide their own problems by manipulating the risk text. They can avoid punishment by improving the semantic novelty and richness of their prospectuses (Nefedova & Pratobevera, 2020).

Most especially, the study of this paper explores the economic effect of the risk disclosure text of the prospectus on China's capital market. The overall development of China's capital market is still a government-led model. According to the specific information environment of newly listed enterprises, different ways are adopted to encourage them to reduce the degree of risk disclosure mode so as to alleviate the information asymmetry between enterprises and investors (Huang et al., 2019a, 2019b). Our research can provide a theoretical basis for the Chinese government to reduce the low IPO price rate, stabilize the stock market, and provide a theoretical basis for the sustainable and healthy development of China's capital market and even the national economy.

There are still limitations in this research that are worthy of further study. First, this paper overcomes the subjectivity and high cost of manual reading and machine learning manual annotation and adopts an unsupervised machine learning algorithm to analyze the text, which is more objective and convenient in method. However, due to the sensitivity of the doc2vec model to the number of samples, there may be some errors in the results. Future research can explore the way of unsupervised text processing. Secondly, this paper reveals the important impact of unstructured text information on the market, which is helpful for us to fully and objectively understand the impact of risk disclosure text on the efficiency of market resource allocation. However, it is not enough to measure only novelty and risk content, which may be affected by many factors, and more factors can be taken into consideration in the future (Boroon et al., 2021; Changchit et al., 2021; Fu et al., 2021; Huang et al., 2021a, b; Islam et al., 2021; Li et al., 2021a, b; Vali et al., 2021; Wang et al., 2021; Zhang et al., 2021a, b; Zhang, Ye et al., 2021). Third, different market participants' perception levels of risk and investment decision-making ability may have an impact on the results. Therefore, the follow-up study can explore the impact of investors' risk perception level and investment decision-making ability on the capital market.

Acknowledgement This research has been supported by the National Natural Science Foundation of China:71871172, Model of Risk knowledge acquisition and Platform governance in FinTech based on deep learning; 71571139, Outlier Analytics and Model of Outlier Knowledge Management in the context of Big Data; We deeply appreciate the suggestions from fellow members of Xia's project team and Research Center of Enterprise Decision Support, Key Research Institute of Humanities and Social Sciences in Universities of Hubei province (DSS20180204).

Appendix

See Table 10.

Table 10 Variable definition and calculation method

Type	Variable	Meaning	Computing method
dependent variable	*Pricing_Efficiency*	IPO pricing efficiency	IPO pricing efficiency
independent variable	*Semantic_Novelty*	Semantic novelty	Unsupervised machine learning method doc2vec, cosine similarity
	Content_Richness	Content richness	Types of risk factors listed in the prospectus
control variable	*Asset*	Asset	Logarithm of total assets of the company in the current year
	Age	Company age	Annual value of time from incorporation to listing
	SOE	Nature of equity	1 for state-owned companies and 0 for non-state-owned companies
	Market_Index	Market system environment	Total index of marketization prepared by Wang et al. (2013)
	Registered_Capital	registered capital	The registered capital of the company at the time of offering, i.e. the total paid up capital, and the logarithm of it
	Turnover_rate	Irrational factors of investors	Turnover rate on the first day of listing
	LOSS	*LOSS*	The loss of the company in the current year is 1, otherwise it is 0
Regulatory variable	*Block*	Enterprise information transparency	Number of enterprise tracking analysts
Categorical variable	*Punish*	Whether the enterprise violates the regulations in the future	In the next three years, if the enterprise is punished for violating the rules, it will be 1, otherwise, it will be 0

References

Ada, E., Sagnak, M., Kazancoglu, Y., Luthra, S., & Kumar, A. (2021). A framework for evaluating information transparency in supply chains. *Journal of Global Information Management (JGIM), 29*(6), 1–22. https://doi.org/10.4018/JGIM.20211101.oa45

Adam-Müller, A. A., & Erkens, M. R. (2020). Risk disclosure noncompliance. *Journal of Accounting and Public Policy, 39*, 106739.

Akter, S., & Wamba, S. F. (2017). Big data and disaster management: A systematic review and agenda for future research. *Annals of Operations Research*. https://doi.org/10.1007/s10479-017-2584-2

Albring, S., Huang, S., Pereira, R., & Xu, X. (2020). Disclosure and liquidity management: Evidence from regulation fair disclosure. *Journal of Contemporary Accounting & Economics, 16*(3), 100205.

Alshammeri, M., Atwell, E., & Ammar Alsalka, M. (2021). Detecting semantic-based similarity between verses of the Quran with Doc2vec. *Procedia Computer Science, 189*, 351–358.
Alshirah, M. H., Alshira'h, A. F., & Lutfi, A. (2022). Political connection, family ownership and corporate risk disclosure: Empirical evidence from Jordan. *Meditari Accountancy Research*, *30*(5), 1241–1264. https://doi.org/10.1108/MEDAR-04-2020-0868
Al-Hadi, A., Al-Yahyaee, K. H., Hussain, S. M., & Taylor, G. (2019). Market risk disclosures and corporate governance structure: Evidence from GCC financial firms. *The Quarterly Review of Economics and Finance, 73*, 136–150.
Al-Matham, R. N., & Al-Khalifa, H. S. (2021). Synoextractor: A novel pipeline for Arabic synonym extraction using Word2Vec word embeddings. *Complexity*, *2021*, Article ID 6627434, p. 13. https://doi.org/10.1155/2021/6627434.
Baker, E. D., Boulton, T. J., Braga-Alves, M. V., & Morey, M. R. (2021). ESG government risk and international IPO underpricing. *Journal of Corporate Finance, 67*, 101913.
Bhardwaj, A., & Imam, S. (2019). The tone and readability of the media during the financial crisis: Evidence from pre-IPO. *International Review of Financial Analysis, 63*, 40–48.
Benamati, J. H., Ozdemir, Z. D., & Smith, H. J. (2021). Information privacy, cultural values, and regulatory preferences. *Journal of Global Information Management (JGIM), 29*(3), 131–164. https://doi.org/10.4018/JGIM.2021050106
Boroon, L., Abedin, B., & Erfani, E. (2021). The dark side of using online social networks: A review of individuals' negative experiences. *Journal of Global Information Management (JGIM), 29*(6), 1–21. https://doi.org/10.4018/JGIM.20211101.oa34
Bourveau, T., De George, E. T., Ellahie, A., & Macciocchi, D. (2022). The role of disclosure and information intermediaries in an unregulated capital market: Evidence from initial coin offerings. *Journal of Accounting Research, 60*(1), 129–167.
Call, A. C., Chen, S., & Tong, Y. H. (2013). Are analysts' cash flow forecasts naive extensions of their own earnings forecasts? *Contemporary Accounting Research, 30*(2), 438–465.
Campbell, J., Chen, L. H., Dhaliwal, D. S., Lu, H., & Steele, L. B. (2014). The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies, 19*(1), 396–455.
Catalfo, P., & Wulf, I. (2016). Intangibles disclosure in management commentary regulation in Germany and Italy: A semantic approach. *Journal of Intellectual Capital*, *17*(1), 103–119. https://doi.org/10.1108/JIC-09-2015-0083.
Chan, S. W. K., & Chong, M. W. C. (2017). Sentiment analysis in financial texts. *Decision Support Systems, 94*, 53–64.
Chang, W., Xu, Z., Zhou, S., & Cao, W. (2018). Research on detection methods based on doc2vec abnormal comments. *Future Generation Computer Systems, 86*, 656–662.
Chang, Y. B., & Kwon, Y. (2020). Attention-grabbing IPOs in early stages for IT firms: An empirical analysis of post-IPO performance. *Journal of Business Research, 109*, 111–119.
Changchit, C., Klaus, T., & Treerotchananon, A. (2021). Using customer review systems to support purchase decisions: A comparative study between the U.S. and Thailand. *Journal of Global Information Management (JGIM), 29*(6), 1–24. https://doi.org/10.4018/JGIM.20211101.oa51
Chen, Q., & Sokolova, M. (2021). Specialists, scientists, and sentiments: Word2Vec and Doc2Vec in analysis of scientific and medical texts. *SN Computer Science, 2*(5), 1–11.
Chen, Y. C., Hung, M., & Wang, Y. (2018). The effect of mandatory CSR disclosure on firm profitability and social externalities: Evidence from China. *Journal of Accounting and Economics, 65*(1), 169–190.
Cheng, L., Hu, H., & Wu, C. (2021). Spammer group detection using machine learning technology for observation of new spammer behavioral features. *Journal of Global Information Management (JGIM), 29*(2), 61–76.
Cheong, A., Yoon, K., Cho, S., & No, W. G. (2021). Classifying the contents of cybersecurity risk disclosure through textual analysis and factor analysis. *Journal of Information Systems, 35*(2), 179–194.
Chintya, N. M., Theodora, N., Evelyn, V., & Teja, A. (2020). Short-term and long-term effect of firms' IPO on competitors' performance. *Journal of Finance and Accounting Research, 2*(1), 114–139.
Choi, S., & Jung, H. (2021). Does early-life war exposure of a CEO enhance corporate information transparency? *Journal of Business Research, 136*, 198–208.
Chong, B. S., & Liu, Z. (2020). Issuer IPO underpricing and directed share program (DSP). *Journal of Empirical Finance, 56*, 105–125.
Duhadway, S., Carnovale, S., & Hazen, B. (2017). Understanding risk management for intentional supply chain disruptions: Risk detection, risk mitigation, and risk recovery. *Annals of Operations Research*. https://doi.org/10.1007/s10479-017-2452-0
Ellahie, A., Hayes, R., & Plumlee, M. A. (2022). Growth matters: Disclosure and risk premium. *The Accounting Review*, *97*(4), 259–286.

Elshandidy, T., & Zeng, C. (2022). The value relevance of risk-related disclosure: Does the tone of disclosure matter?. *Borsa Istanbul Review*, *22*(3), 498–514.

Engelen, P. J., Heugens, P., Van Essen, M., Turturea, R., & Bailey, N. (2020). The impact of stakeholders' temporal orientation on short-and long-term IPO outcomes: A meta-analysis. *Long Range Planning, 53*(2), 101853.

Evans, R. B., & Sun, Y. (2021). Models or stars: The role of asset pricing models and heuristics in investor risk adjustment. *The Review of Financial Studies, 34*(1), 67–107.

Farooq, O., & Hamouda, M. (2016). Stock price synchronicity and information disclosure: Evidence from an emerging market. *Finance Research Letters, 18*, 250–254.

Fu, S., Yan, Q., Feng, G. C., & Peng, J. (2021). Which review can make you engage?: The effect of reviewer-reader similarity on consumer-brand engagement. *Journal of Global Information Management (JGIM), 29*(6), 1–27. https://doi.org/10.4018/JGIM.20211101.oa50

Gao, S., Brockman, P., Meng, Q., & Yan, X. (2019). Differences of opinion, institutional bids, and IPO underpricing. *Journal of Corporate Finance, 60*, 101540.

Grassa, R., Moumen, N., & Hussainey, K. (2020). Is bank creditworthiness associated with risk disclosure behavior? Evidence from Islamic and conventional banks in emerging countries. *Pacific-Basin Finance Journal, 61*, 101327.

Grover, K. L., & Bhullar, P. S. (2021). The nexus between risk factor disclosures and short-run performance of IPOs: Evidence from literature. *World Journal of Entrepreneurship, Management and Sustainable Development*, *17*(4), 907–921.

Gu, Z., Li, Z., Yang, Y. G., & Li, G. (2019). Friends in need are friends indeed: An analysis of social ties between financial analysts and mutual fund managers. *The Accounting Review*, *94*(1), 153–181.

Güçbilmez, U., & Briain, T. Ó. (2020). Bidding styles of institutional investors in IPO auctions. *Journal of Financial Markets, 53*, 100579.

Han, H., Tang, J. J., & Tang, Q. (2021). Goodwill impairment, securities analysts, and information transparency. *European Accounting Review, 30*(4), 767–799.

He, P., Ma, L., Wang, K., & Xiao, X. (2019). IPO pricing deregulation and corporate governance: Theory and evidence from Chinese public firms. *Journal of Banking & Finance, 107*, 105606.

He, Q., & Fang, C. (2019). Regulatory sanctions and stock pricing efficiency: Evidence from the Chinese stock market. *Pacific-Basin Finance Journal, 58*, 101241.

Hope, O. K., Hu, D., & Lu, H. (2016). The benefits of specific risk-factor disclosures. *Review of Accounting Studies, 21*(4), 1005–1045.

Hoque, H., & Mu, S. (2019). Partial private sector oversight in China's A-share IPO market: An empirical study of the sponsorship system. *Journal of Corporate Finance, 56*, 15–37.

Huang, R., Qu, S., Yang, X., & Liu, Z. (2021a). Multi-stage distributionally robust optimization with risk aversion. *Journal of Industrial & Management Optimization, 17*(1), 233.

Huang, W., Li, J., & Zhang, Q. (2019a). Information asymmetry, legal environment, and family firm governance: Evidence from IPO underpricing in China. *Pacific-Basin Finance Journal, 57*, 101109.

Huang, Y., Ma, J., Wu, C., & Yang, S. (2021b). An emoji is worth a thousand words: The influence of face emojis on consumer perceptions of user-generated reviews. *Journal of Global Information Management (JGIM), 29*(6), 1–23. https://doi.org/10.4018/JGIM.20211101.oa2

Huang, Y. S., Li, M., & Chen, C. R. (2019b). Financial market development, market transparency, and IPO performance. *Pacific-Basin Finance Journal, 55*, 63–81.

Hussein, M., Zhou, Z. G., & Deng, Q. (2020). Does risk disclosure in prospectus matter in ChiNext IPOs' initial underpricing? *Review of Quantitative Finance and Accounting, 54*(3), 957–979.

Huo, X., Jasimuddin, S. M., Zheng, K., & Zhang, Z. (2022). Risks of supply chain financial warehouse receipts pledge: A structural equation approach. *Supply Chain Forum: An International Journal*, pp. 1–12. Taylor & Francis

Ibrahim, A. E. A., & Hussainey, K. (2019). Developing the narrative risk disclosure measurement. *International Review of Financial Analysis, 64*, 126–144.

Islam, M., Kang, M., & Haile, T. T. (2021). Do hedonic or utilitarian types of online product reviews make reviews more helpful?: A new approach to understanding customer review helpfulness on Amazon. *Journal of Global Information Management (JGIM), 29*(6), 1–18. https://doi.org/10.4018/JGIM.20211101.oa52

Jin, G. Z., Luca, M., & Martin, D. (2021). Is no news (perceived as) bad news? An experimental investigation of information disclosure. *American Economic Journal: Microeconomics, 13*(2), 141–173.

Jog, V., Otchere, I., & Sun, C. (2019). Does the Two-Stage IPO process reduce underpricing and long run underperformance? Evidence from Chinese firms listed in the U.S. *Journal of International Financial Markets, Institutions and Money, 59*, 90–105.

Kabir, H., Su, L., & Rahman, A. (2020). Firm life cycle and the disclosure of estimates and judgments in goodwill impairment tests: Evidence from Australia. *Journal of Contemporary Accounting & Economics, 16*(3), 100207.
Kamal, Y. (2021). Stakeholders expectations for CSR-related corporate governance disclosure: Evidence from a developing country. *Asian Review of Accounting*, *29* (2), 97–127. https://doi.org/10.1108/ARA-04-2020-0052.
Kao, L., Chen, A., & Krishnamurti, C. (2020). Outcome model or substitute model of D&O insurance on IPO pricing without information asymmetry before issuance. *Pacific-Basin Finance Journal, 61*, 101300.
Khaw, M. W., Li, Z., & Woodford, M. (2021). Cognitive imprecision and small-stakes risk aversion. *The Review of Economic Studies, 88*(4), 1979–2013.
Kim, H. J., Kim, T. S., & Sohn, S. Y. (2019). Recommendation of startups as technology cooperation candidates from the perspectives of similarity and potential: A deep learning approach. *Decision Support Systems, 130*, 113229.
Kim, K. A., Liu, H., & Yang, J. J. (2013). Reconsidering price limit effectiveness. *Journal of Financial Research, 36*(4), 493–518.
Kouvelis, P., Xiao, G., & Yang, N. (2021). Role of risk aversion in price postponement under supply random yield. *Management Science, 67*(8), 4826–4844.
Le, Q. V., & Mikolov, T. (2014). Distributed representations of sentences and documents. In *Proceedings of international conference on machine learning* (pp. 1188–1196).
Lee, C. M. C., Qu, Y., & Shen, T. (2019). Going public in china: Reverse mergers versus ipos. *Journal of Corporate Finance, 58*, 92–111.
Li, C., Liu, Y., & Du, R. (2021a). The effects of review presentation formats on consumers' purchase intention. *Journal of Global Information Management (JGIM), 29*(6), 1–20. https://doi.org/10.4018/JGIM.20211101.oa46
Li, H., & Zhu, F. (2021). Information transparency, multihoming, and platform competition: A natural experiment in the daily deals market. *Management Science, 67*(7), 4384–4407.
Li, M., Liu, D., Zhang, J., & Zhang, L. (2021b). Volatile market condition, institutional constraints, and IPO anomaly: Evidence from the Chinese market. *Accounting & Finance, 61*(1), 1239–1275.
Li, X., Wang, S. S., & Wang, X. (2019b). Trust and IPO underpricing. *Journal of Corporate Finance, 56*, 224–248.
Li, Y., He, J., & Xiao, M. (2019a). Risk disclosure in annual reports and corporate investment efficiency. *International Review of Economics & Finance, 63*, 138–151.
Liu, K., Tang, J., Yang, K., & Arthurs, J. (2019). Foreign IPOs in the U.S.: When entrepreneurial orientation meets institutional distance. *Journal of Business Research, 101*, 144–151.
Lo, K., Ramos, F., & Rogo, R. (2017). Earnings management and annual report readability. *Journal of Accounting and Economics, 63*(1), 1–25.
Lobo, G. J., & Zhao, Y. (2013). Relation between audit effort and financial report misstatements: Evidence from quarterly and annual restatements. *The Accounting Review, 88*(4), 1385–1412.
Loh, R. K., & Stulz, R. M. (2018). Is sell-side research more valuable in bad times? *The Journal of Finance, 73*(3), 959–1013.
Mai, L., & Le, B. (2020). Joint sentence and aspect-level sentiment analysis of product comments. *Annals of Operations Research*. https://doi.org/10.1007/s10479-020-03534-7
McGuinness, P. B. (2019). Risk factor and use of proceeds declarations and their effects on IPO subscription, price 'fixings', liquidity and after-market returns. *The European Journal of Finance, 25*(12), 1122–1146. https://doi.org/10.1080/1351847X.2019.1572023
Mikolov, T., Sutskever, I., Chen, K., et al. (2013). Distributed representations of words and phrases and their compositionality. *Neural Information Processing Systems, 26*, 3111–3119.
Nefedova, T., & Pratobevera, G. (2020). Do institutional investors play hide-and-sell in the IPO aftermarket? *Journal of Corporate Finance, 64*, 101627.
Niu, H., & Silva, E. A. (2021a). Delineating urban functional use from points of interest data with neural network embedding: A case study in Greater London. *Computers, Environment and Urban Systems, 88*, 101651.
Ma, X., He, J., & Liao, J. (2021). Does decision fatigue affect institutional bidding behavior? Evidence from Chinese IPO market. *Economic Modelling, 98*, 1–12.
Makrominas, M., & Yiannoulis, Y. (2021). IPO determinants of delisting risk: Lessons from the Athens Stock Exchange. In *Accounting forum* (pp. 1–25). Routledge.
Mandal, A., Ghosh, K., Ghosh, S., & Mandal, S. (2021). Unsupervised approaches for measuring textual similarity between legal court case reports. *Artificial Intelligence and Law*, *29*(3), 417–451.
Marcato, G., Milcheva, S., & Zheng, C. (2018). Market integration, country institutions and IPO underpricing. *Journal of Corporate Finance, 53*, 87–105.

Massa, M., & Zhang, L. (2021). Local investor horizon clientele and IPO underpricing. *Journal of Financial Markets, 54*, 100587.
Michelon, G., Rodrigue, M., & Trevisan, E. (2019). *The marketization of a social movement: Activists, shareholders and CSR disclosure* (p. 101074). Organizations and Society.
Mishra, S., Pappu, A., & Bhamidipati, N. (2019). Inferring advertiser sentiment in online articles using Wikipedia footnotes. In *Companion proceedings of the 2019 world wide web conference* (pp. 1224–1231).
Niu, H., & Silva, E. A. (2021b). Delineating urban functional use from points of interest data with neural network embedding: A case study in Greater London. *Computers Environment and Urban Systems, 88*, 101651.
Papadamou, S., Sidiropoulos, M., & Spyromitros, E. (2017). Interest rate dynamic effect on stock returns and central bank transparency: Evidence from emerging markets. *Research in International Business and Finance, 39*, 951–962.
Park, Y. E. (2021). Developing a COVID-19 crisis management strategy using news media and social media in big data analytics. *Social Science Computer Review*, 08944393211007314.
Peng, X., Jia, Y., & Chan, K. C. (2021). The impact of internationalization on IPO underpricing: A result of agency costs reduction, a certification effect, or a diversification benefit? *Finance Research Letters, 44*, 102059.
Que, J., & Zhang, X. (2019). Pre-IPO growth, venture capital, and the long-run performance of IPOs. *Economic Modelling, 81*, 205–216.
Rashid, R. M., Abdul-Rahim, R., & Yong, O. (2014). The influence of lock-up provisions on IPO initial returns: Evidence from an emerging market. *Economic Systems, 38*(4), 487–501.
Rathnayake, D. N., Louembe, P. A., Kassi, D. F., Sun, G., & Ning, D. (2019). Are IPOs underpriced or overpriced? Evidence from an emerging market. *Research in International Business and Finance, 50*, 171–190.
Ritter, J. R., & Welch, I. (2002). A review of IPO activity, pricing, and allocations. *The Journal of Finance, 57*(4), 1795–1828. https://doi.org/10.1111/1540-6261.00478
Saci, F., & Jasimuddin, S. M. (2021). Does the research done by the institutional investors affect the cost of equity capital? *Finance Research Letters, 41*, 101834.
Shang, H., Lu, D., & Zhou, Q. (2021). Early warning of enterprise finance risk of big data mining in internet of things based on fuzzy association rules. *Neural Computing and Applications, 33*(9), 3901–3909.
Sheng, J., Xu, N., & Zheng, L. (2021). Do mutual funds walk the talk? A textual analysis of risk disclosure by mutual funds. *A Textual Analysis of Risk Disclosure by Mutual Funds (February 8, 2021)*.
Shivaani, M. V., & Agarwal, N. (2020). Does competitive position of a firm affect the quality of risk disclosure? *Pacific-Basin Finance Journal, 61*, 101317.
Shivaani, M., Jain, P. K., & Yadav, S. S. (2020). Development of a risk disclosure index and its application in an Indian context. *Managerial Auditing Journal*, *35*(1), 1–23. https://doi.org/10.1108/MAJ-07-2016-1403.
Spence, P. R., Lin, X., Lachlan, K. A., & Hutter, E. (2020). Listen up, I've done this before: The impact of self-disclosure on source credibility and risk message responses. *Progress in Disaster Science, 7*, 100108.
Srivastava, P. R., & Eachempati, P. (2021). Intelligent employee retention system for attrition rate analysis and churn prediction: An ensemble machine learning and multi-criteria decision-making approach. *Journal of Global Information Management (JGIM), 29*(6), 1–29.
Sun, Y., Sun, X., & Wu, W. (2021). Who detects corporate fraud under the thriving of the new media? Evidence from Chinese-listed firms. *Accounting & Finance, 61*, 1313–1343.
Vali, H., Xu, J. D., & Yildirim, M. B. (2021). Comparative reviews vs. regular consumer reviews: Effects of presentation format and review valence. *Journal of Global Information Management (JGIM), 29*(6), 1–29. https://doi.org/10.4018/JGIM.20211101.oa7
Wang, X., & Song, D. (2021). Does local corruption affect IPO underpricing? Evidence from China. *International Review of Economics & Finance, 73*, 127–138.
Wang, X. L., Yu, J. W., & Fan, G. (2013). 2013 Report of China's provincial enterprise business environment index. *Journal of State Administration College, 4*, 24–34. (In Chinese).
Wang, Z., Chang, V., & Horvath, G. (2021). Explaining and predicting helpfulness and funniness of online reviews on the steam platform. *Journal of Global Information Management (JGIM), 29*(6), 1–23. https://doi.org/10.4018/JGIM.20211101.oa16
Wasiuzzaman, S., Yong, F. L. K., Sundarasen, S. D. D., & Othman, N. S. (2018). Impact of disclosure of risk factors on the initial returns of initial public offerings (IPOs). *Accounting Research Journal*, *31*(1), 46–62. https://doi.org/10.1108/ARJ-09-2016-0122.
Wei, L., Li, G., Zhu, X., Sun, X., & Li, J. (2019). Developing a hierarchical system for energy corporate risk factors based on textual risk disclosures. *Energy Economics, 80*, 452–460.

Xia, H., Wang, P., Wan, T., Zhang, Z., Weng, J., & Jasimuddin, S. M. (2022). Peer-to-peer lending platform risk analysis: An early warning model based on multi-dimensional information. *Journal of Risk Finance, 23*(3), 303–323.

Xuan, P., Xiongyuan, W., & Chan, K. C. (2019). Does customer concentration disclosure affect IPO pricing? *Finance Research Letters, 28*, 363–369.

Xue, X., Zhang, J., & Yu, Y. (2020). Distracted passive institutional shareholders and firm transparency. *Journal of Business Research, 110*, 347–359.

Yao, Y., & Zhao, M. (2016). Chinese style risk disclosure, disclosure level and market reaction. *Economic Research, 51*(07), 158–172. (In Chinese).

Zhang, H., Hong, X., Li, Q., Gong, Y., & Liu, S. (2021a). Exploring the intellectual structure and international cooperation in information management: A bibliometric overview using 2-tuple linguistic model. *Journal of Global Information Management (JGIM), 29*(6), 1–20. https://doi.org/10.4018/JGIM.294577

Zhang, Y., Ge, L., Xiao, L., Zhang, M., & Liu, S. (2021b). A bibliometric review of information systems research from 1975–2018: Setting an agenda for IS research. *Journal of Global Information Management (JGIM), 29*(6), 1–24. https://doi.org/10.4018/JGIM.287631

Zhang, Y. J., & Liu, J. Y. (2020). Overview of research on carbon information disclosure. *Frontiers of Engineering Management, 7*(1), 47–62.

Zhou, L., & Sadeghi, M. (2019). The impact of innovation on IPO short-term performance—Evidence from the Chinese markets. *Pacific-Basin Finance Journal, 53*, 208–235.

Zhou, Z. G., Hussein, M., & Deng, Q. (2021). ChiNext IPOs' initial returns before and after the 2013 stock market reform: What can we learn?. *Emerging Markets Review*, 100817.

Zia-ur-Rehman, M., Latif, K., Mohsin, M., Hussain, Z., Baig, S. A., & Imtiaz, I. (2021). How perceived information transparency and psychological attitude impact on the financial well-being: mediating role of financial self-efficacy. *Business Process Management Journal*, *27*(6), 1836–1853. https://doi.org/10.1108/BPMJ-12-2020-0530.

Zimmer, J. C., Arsal, R. E., Al-Marzouq, M., & Grover, V. (2010). Investigating online information disclosure: Effects of information relevance, trust and risk. *Information & Management, 47*(2), 115–123.

Zu, X., Yu, W., & Qiu, Y. (2018). Research on the impact of food safety crisis on brand trust: The mediating effect of enterprise information transparency. In *2018 International conference on economics, finance, business, and development (ICEFBD 2018)* (pp. 185–191).