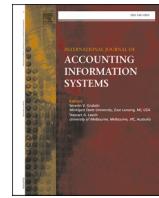




Contents lists available at ScienceDirect

International Journal of Accounting Information Systems

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



Impact of internal control quality on the information content of social responsibility reports: A study based on text similarity—Evidence from China



Pingping Huang^a, Yuehua Jiao^b, Sihai Li^{c,*}

^a School of Accounting, Nanjing Audit University, China

^b School of Accounting, Zhongnan University of Finance and law, China

^c School of Accounting, Zhongnan University of Finance and law School of Management & West Yunnan University of Applied Sciences, Dali, China

ARTICLE INFO

Keywords:

Internal control quality
Social responsibility reports
Information content
Text similarity

ABSTRACT

To date, the impact of internal control on text information social responsibility reports has been rarely explored. This study investigates the correlation between internal control and information content of social responsibility reports based on the similarity of text information of reports by listed enterprises in 2006–2017. Overall, high-quality internal control results in low text similarity of social responsibility and high information content of social responsibility reports. The internal control quality could enhance the information content of social responsibility reports by decreasing the enterprise's agency costs. For large-scale corporations, non-state enterprises and those with intensive public attention, effective internal control can enhance the information content of social responsibility reports. The agency cost can exert a partial mediating effect for large-scale corporations and enterprises with intensive public attention and complete mediating effect for non-state enterprises. Moreover, this study extends the scope of internal control to the quantitative dimension of text information and has crucial implications for listed enterprises, investors, and governments regarding internal control and disclosure of social responsibility information.

1. Introduction

Recently, adverse incidents like “injection fluid” of Fosun Pharma, “fake vaccines” of Changchun Changsheng, “poison capsules” of Xuizheng Pharma, “Rashomon” of Moutai, “plasticizer” of Jiugui Liquor, “over-high concentration of nitrates” of Beingmate, and “explosions” of Zijin Mining have drawn significant public attention to corporate social responsibility (CSR). In most cases, activities and accomplishments of CSR should be included in the CSR reports establishing a good reputation and market confidence. The CSR information has incremental values in assessments of future profitability and incomes of the enterprise by investors. Hence, the disclosure of social responsibility information, a crucial carrier of nonfinancial information, has garnered significant attention. On April 3, 2014, the China Securities Regulatory Commission released *Memorandum No. 21 of information disclosure: issues related to periodic reports*, which necessitates that CSR reports should provide full details about the current status, issues observed, and plans of social responsibility based on industrial characteristics, whereas selective disclosures of good news should be evaded.

* Corresponding author.

E-mail address: hpp551@126.com (P. Huang).

The “Internal Control-Integration Framework” updated in May 2013 highlighted that the reporting objectives to be attained by internal control comprise financial and nonfinancial reports, and cover a series of features like reliability, timeliness, and transparency needed by regulators, standard setters, or other policymakers. The COSO report (2013) extends the scope of internal control from the original external financial report to internal and external financial and nonfinancial reports. The external nonfinancial reports comprise internal control reports and sustainability reports. Social responsibility information disclosure is the extension of company’s financial information disclosure system. Internal control enhances the company’s operation efficiency and benefits and guarantees the authenticity of internal and external information communication (Bushman & Smith, 2001; Li, 2013) by augmenting the credibility of financial reports, identifying and evaluating related risks in time. Moreover, the enhancement of internal control helps to improve the level of corporate governance and ensure the legitimacy and integrity of information disclosure. Thus, effective internal control is the fundamental assurance for the quality of information disclosure and is closely related to the investors’ interests.

Disclosures of text information, typically voluntary, can provide unique hints on managers’ cognitive bias and are usually free from auditing. The language of text information depicts management’s subjectivity and behavioral characteristics (Lo et al., 2017; Tan et al., 2014). Thus, text information disclosure provides more forward-looking information than quantitative disclosure. Previous studies reported that the content of text information could project the future performance of an enterprise, cause substantial market reactions, and decrease IPO underpricing (Hanley & Hoberg, 2010; Price et al., 2012). With the advancement of natural language processing, the research method of text similarity emerged, which was adopted owing to previous research (Brown & Tucker, 2011) demonstrating that with an increase of the differences between MD&A texts, the reaction of investors would increase, and investors will respond to the new numerical information in 10-K filings; thus, investors will use the MD&A modification information. Thus, a negative correlation exists between information content and text similarity. Based on the frequency of words contained in social responsibility reports, the information of social responsibility reports was transformed into a vector form through text vectorization. The information content of text information was measured by evaluating the cosine similarity of text vectors. Cosine similarity identifies similar documents by comparing the relative word frequencies across documents (Lang & Stice-Lawrence, 2015). Owing to considerable text and chart information in the social responsibility report, which differs from the ordinary text information to be extracted directly, the chart with the need to be identified and separated from the text information, and the text information are extracted separately. The entire process takes longer time and more computing resources. At present, this method has not been wholly applied to the research and analysis of social responsibility reports. Besides, limited studies focused on the effects of internal control on text information on social responsibility reports. In this study, text similarity is used as a proxy of CSR reports information content. This study focuses on internal control as a crucial internal governance mechanism and whether its quality could affect social responsibility reports? What is the mechanism behind it? Furthermore, whether the property right nature of enterprises, the scale of enterprises, and the external media attention exert a heterogeneous impact. All these problems warrant theoretical analysis and empirical tests.

The capital market data from China were used for the following reasons. First, the language structure and form of Chinese differs from English, resulting in different algorithms and data processing in text processing from English. In the social responsibility text, word segmentation was performed on all texts through the “Jieba” Chinese word segmentation module in Python function library, and word frequency statistics were carried out. Per Runling Global Responsibility Rating (RKS), the number of listed companies that disclosed social responsibility reports increased from 518 in 2009 to 851 in 2018. In December 2017, China Securities Regulatory Commission (CSRC) published the No.2 Guidelines on Information Disclosure Contents and Formats of Companies Public Issuing Securities. Articles 42 and 43 of the Guidelines set relevant requirements for social responsibility information. China’s regulatory authorities are focusing more on the social responsibility information disclosure of listed companies. Second, the DIB database provides indicators to measure the quality of internal control of Chinese enterprises, including internal environment, risk assessment, control activities, information and communication, and internal supervision (Chen et al., 2019). Thus, this database provides data support for path detection in this study. Third, the property right nature has a unique system background in China. In non-state-owned enterprises, internal control exerts a more substantial positive impact on their social responsibility information disclosure. Using Chinese data to examine the impact of internal control on social responsibility text messages might have peculiar findings (Li & Zhang, 2017).

This study’s research contributions are primarily reflected in the following aspects: (i) expanding the relevant research of social responsibility information disclosure. For the quantitative assessment of social responsibility text information, the existing literature has demonstrated the tone intensity and semantic features of enterprise environment information disclosure based on semantic analysis methods (Zhang et al., 2016) and created the environment information disclosure index (Li, 2014; Shen, 2007). In addition, the existing literature has measured the level of social responsibility reports intonation through experiments and natural language processing (Huang & Li, 2020; Zhang et al., 2019). Besides, studies have been conducted from the quantitative perspective of text similarity of social responsibility. Despite studies on the similarity measurement of the text carrier of financial reports and MD&A (Brown & Knechel, 2016; Johnston & Zhang, 2018; Meng et al., 2017), limited studies are available on the measurement of the text carrier of social responsibility reports from the similarity dimension. (ii) Expanding research on the economic consequences of internal control. Some studies have measured the impact of internal control on manager information disclosure behavior, auditor behavior, analyst forecasts, and other stakeholders (Dong et al., 2017; Lee, 2018); however, limited studies are available on the effects of internal control text information disclosure.

2. Literature review

2.1. Measurement of social responsibility information disclosure

Dhaliwal et al. (2012) investigated the impact of social responsibility information disclosure on analyst forecasts, used virtual variables to assess social responsibility information disclosure, and determined whether the enterprise released social responsibility reports during the year, whether social responsibility reports were ensured by a third party, and whether the number of pages of social responsibility reports was higher than the median value of pages of all social responsibility reports. Based on the content analysis, Campopiano and De Massis (2015) explored how the organization of family firms affects the content of social responsibility reports, revealing some differences in the types and content of CSR reports between family and non-family firms through theoretical analysis. Based on the narrative intonation, readability, length, number of CSR reports, and the narrative content of social responsibility reports, Muslu et al. (2019) developed a disclosure score to examine the correlation between CSR disclosure scores and analyst forecasts, demonstrating that the content of CSR reports enables augmenting the accuracy of analyst forecasts; for CSR reports with richer content, this correlation is more noticeable. In addition, the environment information disclosure index was developed using the content analysis method, which was evaluated by the ratio of the sum of the listed enterprises' environment information disclosure item scores and the best disclosure environment information disclosure score (Li, 2014; Shen, 2007). Based on the semantic analysis method, a study described the tone intensity and semantic features of the enterprise environment information disclosure (Zhang et al., 2016). Using an experimental method, the intonations of social responsibility reports were identified, and the general statement of the social responsibility environment information content was defined as plain intonation. Moreover, words with apparent positive emotions, such as "obvious" and "huge," were defined as active intonation (Zhang et al., 2019). Huang and Li (2020) performed word segmentation of social responsibility reports text using Python development source "Jieba" Chinese word segmentation module and word frequency statistics, as well as constructed social responsibility reports intonation TONE = (POSPCT - NEGPCT)/(POSPCT + NEGPCT). Based on the agency theory and impression management theory, Ji et al. (2016) first used the average sentence length, number of pages, number of pictures, and whether a color is present in CSR reports cover as a measure of the readability of CSR reports; they established a significant positive correlation between the managerial power and CSR reports, and that listed enterprises have the phenomenon of manipulating CSR reports information disclosure. Based on the 2-POS model, Duan et al. (2017) used the text mining technology to measure the subjectivity and objectivity of listed enterprises social responsibility reports during 2009–2015; they demonstrated that the subjectivity score of social responsibility reports in China had been increasing over the years, and big differences existed between different industries and years. Sun (2012) used experimental methods to investigate the clarity of social responsibility reports and the impact of third-party assurance on investors decision-making. A social responsibility report only using words to describe the performance of its social responsibility is considered to be of low disclosure clarity. Meanwhile, the report that uses not only text descriptions but also clear data to describe the performance of its social responsibility is considered to be of high disclosure clarity.

2.2. Effects of internal control on social responsibility information disclosure

Al-Tuwaijri et al. (2004) reported that the management representing one of the internal control environments attaches great significance to environmental issues, exerting a substantial impact on corporate environment information disclosure. Reportedly, the more attention the management pays to environmental issues, the better the disclosure effect. Li and You (2013) constructed the environment information disclosure index and reported that the internal control scheme exerts significant positive impact on the disclosure level of the environment information, as well as irrespective of whether the marketization process is fast or slow, whether it is a heavily polluting enterprise, internal control, and environmental information disclosure are positively correlated (Li, 2014). In addition, the study found that internal control can produce economic benefits, and its environmental spillover effect is a crucial factor to effectively enhance the environmental quality of information disclosure. In addition, Li and Zhang (2017) reported that the internal control quality (ICQ) could enhance the social responsibility quality of information disclosure and is more prominent in non-state enterprises. Against the research background of the promulgation and implementation of "Enterprise Internal Control Application Guidelines" and "Environmental Protection Law," Qiao and You (2015) further revealed the correlation between the efficacy of internal control and the quality of environmental disclosure based on the objective of internal control, as well as reported that the higher the efficacy of internal control, the higher the environmental quality of information disclosure. Through research on listed enterprises in the A-share market during 2009–2013, Furthermore, Chen (2016) reported that enhancing internal control could help to increase the level of corporate environment information disclosure; however, influenced by the internal power structure, managerial power hampers this promotion.

3. Research on text information similarity

Some studies used similarity in text analysis to measure information content. Brown and Tucker (2011) used textual similarity in the disclosure of MD&A and demonstrated that changes in MD&A provided information to investors. Brown and Knechel (2016) and De Franco et al. (2016) used similarities in the textual annual financial reports to explore auditor selection and reported that companies with the same auditors had similar textual disclosures. In the context of internal control disclosures, Boritz et al. (2016) reported that texts drafted by auditors were more likely to contain boilerplate language. Hoberg and Phillips (2010, 2016) and Bills et al. (2018) used textual similarity in product descriptions and measured business similarity. Bills et al. (2018) specifically examined auditor selection

and reported that companies tended to select the same auditor when product descriptions were similar. Johnson and Zhang (2018) used the previous text similarity measure to assess the similarity of financial reports and demonstrated that auditor style had no impact on the similarity of financial reports. Based on the text vectorization method, Meng et al. (2017) reported that the information content of MD&A would decrease the risk of stock price crash in the future, and only the prospect information exerted a significant impact. In addition, Wang et al. (2018) demonstrated that the similarity of risk information could reflect the amount of risk information increment, and the cosine similarity of risk information in annual reports will decrease audit fees.

To date, some studies have investigated the impact of internal control on the quality of information disclosure, including the qualitative study on the impact of internal control on the level of social responsibility information disclosure; however, limited studies explored the impact of internal control on information disclosure from the quantitative perspective of text information similarity. Currently, the measurement of text information similarity is rarely carried out from the carrier of social responsibility report.

4. Theoretical analysis and Hypothesis

In 2011, the COSO Committee revised the internal control framework and extended the original financial goals to broader goals. The framework integrated financial, environmental, social, and governance information. By constructing and strengthening internal control, enterprises can institutionalize the social responsibility content imposed by external burdens, compelling enterprises to perform social responsibilities. The performance of CSR is mirrored through the disclosure of CSR information, and performing CSR is based on the disclosure of social responsibility information. Based on the theory of impression management, the management executing impression management on social responsibility reports is frequently based on self-interested motives, which affects information users' assessment for corporate images and performances, and even the judgment for managers' reputations. Thus, the authenticity, completeness, and reliability of social responsibility information is influenced by the management's impression management behavior and plays a specific misleading role in decision-making. Of note, effective internal control is a prerequisite for enterprises to exhibit rational behaviors and decisions, as well as is an institutional arrangement to safeguard the interests of related groups. In addition, internal control has become an effective means of modern enterprise management (Yang & Chen, 2013). Thus, a growing number of studies highlighted that to enhance the consciousness of CSR fulfillment, it is essential to vigorously reinforce the construction of internal control system. When the efficiency of internal control of an enterprise is higher, more social responsibilities lie in fulfilling compliance requirements; to spread a good reputation and create a positive social image, the enterprise usually chooses to disclose more responsibility information to the outside world.

If an enterprise aims to develop continually, it is essential to safeguard social interests. Effective internal control enables the enterprise to regulate its internal behavior independently and decrease illegal phenomena. For their own interests, stakeholders have realized that the level of social responsibility information disclosure has garnered increasing public attention. At the moral level, the public supervises and timely attacks the lack of CSR through the media. Positive social responsibility information establishes a good image of the company and evades negative factors on the company's image to some extent. In addition, the readability or complexity of text information directly affects whether the information users outside the relevant organization can better understand the actual performance of the organization's responsibilities. When the readability of the text information disclosed by the relevant organizations is poor, it is challenging for the information users to comprehend and obtain the information they need. In contrast, it demonstrates that the relevant organizations are suspected of covering up the problem to some extent (Ertugrul et al., 2017; Kim et al., 2019). Thus, when relevant organizations disclose information, they should enhance the readability of text information. In addition, enterprises need to disclose detailed information to the maximum extent possible and add pertinent keywords, such as date, place, person, organization name, value, and percentage, so that external stakeholders can understand more information related to enterprises and be interested in reading it—all to better comprehend the actual situation of CSR. Moreover, an effective control mechanism enables managers to credibly commit to a high-disclosure policy (Verrecchia, 2001) and managers might be encouraged to disclose more specific information; lack of strong governance measures make it challenging for managers to do this and quickly return to low-disclosure policies to accomplish short-term objectives. Thus, high-quality internal control tends to further endorse enhancing the quality of social responsibility information disclosure based on improving corporate governance (Li & Zhang, 2017).

The social responsibility report is primarily based on text information. The details of text information disclosure and the selection of words define the subjectivity and tendentiousness of management (Tan et al., 2014). The related research revealed that text information is not only an integral part of information disclosure but also it could influence investors to make relevant decisions (Brown & Knechel, 2016; Muslu et al., 2019; Tan et al., 2014), and the management might "manipulate" the text information to attain the corresponding objective. Regarding the information content of text information, some studies suggested that the less similar the texts of MD&A or 10 K documents were compared with those of the previous issue, the more likely the capital market was to respond to it (Brown & Tucker, 2011; Lang & Stice-Lawrence, 2015). Explicitly, the higher the similarity between the text, the less incremental information provided, and the less information content. Conversely, the less similar the text, the higher its information content. Some other studies highlighted that standard information is the content similar to existing information, whereas unique information is the content unlike the existing information, which can be understood as enterprise's information that differentiates from other enterprises; that is, standard information is too homogeneous and low-quality information, which cannot support investors to enhance their decision-making, and unique information could help investors distinguish good from bad to enhance the accuracy of pricing; thus, high-quality information is more likely to decrease the IPO underpricing level (Hanley & Hoberg, 2010; Hao & Su, 2014). Notably, text similarity assesses the information content by quantifying the text information and evaluating the angle between the text vectors before and after the year; each element is the frequency of each word in the text. We used the TF-IDF weighting function to decrease the weight of standard words in all documents to zero or very close to zero. Of note, TF-IDF heavily weighs words unique to a company,

treating them as very “informative” words. Thus, after the analysis presented above, high-quality internal control mechanism can decrease the text similarity of social responsibility reports and even enhance the information content of social responsibility reports. Accordingly, the following hypothesis is proposed:

Hypothesis 1. *The better the ICQ, the less text similarity of social responsibility reports, and the higher the information content of social responsibility reports.*

Jensen and Meckling (1976) systematically examined the principal-agent relationship between management and shareholders. Per the decentralized structure of property rights, the control of the enterprise is under the manager, who sacrifices the shareholders' interests for his/her personal gain; this is the first type of agency problem. Based on the information asymmetry theory, internal management offers more information advantages than external investors. Of note, management can choose to optimistically disclose social responsibility reports to increase the information transparency, enabling management to whitewash the negative information of the enterprise and build a positive image. Owing to the principal–agent problem, corporate managers face ethical risks and adverse selection. When the agency cost of the enterprise is low, the interests of shareholders and managers are more consistent. The manager considers the enterprise's long-term progress and discloses more information to assist investors to comprehend the enterprise's future development to minimize financing restraints. Conversely, when the enterprise's agency cost is high, the information asymmetry between shareholders and the manager is higher, and managers are more inclined to conceal and decrease voluntary information disclosure or manipulate accounting information, thereby damaging the shareholders' interests to attain self-interest (Luo & Zhu, 2010). The fundamental objective of internal control is to offer a rational and effective system guarantee for the enterprise's economic behavior, and high-quality internal control alleviates the agency problem to some extent, thereby reducing the degree of information asymmetry between managers and shareholders and investors. In addition, internal control can alleviate agency complications by refining the incentive and supervision mechanism of the enterprise. The internal control environment affects the management's decision-making mechanism and restricts the management's behavior. Furthermore, good information and communication channels and strict approval control procedures can hinder the opportunistic behavior of executives, thereby decreasing the enterprise's agency cost.

A social responsibility report has numerous pictures and text information, which serves as a valuable supplement to the digital information, and can provide incremental information beyond the digital information. Nevertheless, the management exercises great discretion in how to disclose it. Augmenting the ICQ can effectively decrease the contradictions between shareholders and agents, alleviate information asymmetry, and enhance the quality of social responsibility information disclosure. Accordingly, the following hypothesis is proposed:

Hypothesis 2. *Internal control improves the information content of social responsibility reports by decreasing agency costs.*

5. Research design

5.1. Sample selection and data source

As China's social responsibility reports started in 2006, this study's research object is listed enterprises in the A-share market that disclosed social responsibility reports during 2006–2017 as the initial research samples. Based on the previous research practices and the research needs, data were performed with following processing: (i) excluding samples from the financial industry; (ii) excluding samples from ST and *ST enterprises; (iii) excluding samples with missing data; (iv) excluding samples with invalid text during social responsibility text processing; (v) to evade the influence of extreme outliers, a Winsorize processing of up and down 1% was performed

Table 1

Sample selection procedure.

Initial sample: firm-year observations of Chinese A-share listed firms from fiscal years 2006–2017	32,128
Less: firm-year observations of the financial industries	(807)
Number of observations remaining	31,321
Less: firm-year observations of ST and *ST companies	(1766)
Number of observations remaining	29,555
Less: firm-year observations of missing data	(4345)
Number of observations remaining	26,361
Less: firm-year observations of invalid text in CSR text processing ^a	(25,210)
Number of observations remaining	1151

^a As not all enterprises have released CSR, we collected 7578 original CSR reports from the Internet. While converting CSR reports from PDF to TXT, 2086 CSR reports failed to convert successfully, rendering the converted part invalid text. In addition, we calculated similarity variables using the vector angle between the company's social responsibility reports in that year and previous year. Thus, 975 samples of the first issue were not included in the calculation category. Then, 126 companies lost their CSR reports samples that had only been released for 1 year, and CSR reports of 252 companies was presented in the picture format, whose results cannot be calculated; thus, these samples were lost. After merging with other variables, 1151 observations were obtained. For discontinuous disclosure, to ensure that the variables were in the calculation dimension of the same year span, we only retained the samples with an interval of 1 year, and the discontinuous part was deleted (e.g., a company has CSR reports for 2007, 2008, and 2010, but not 2009; we only calculated the similarity between the texts of 2007 and 2008, 2010 and 2011, finally obtaining the variable results of 2008 and 2011).

on continuous variables. The final sample size was 1151 enterprise-annual observations. [Table 1](#) reports the specific sample selection procedure. [Fig. 1](#) shows the sample industry distribution.

In May 2008, the Shanghai Stock Exchange issued the guidelines on environmental information disclosure of listed companies of the Shanghai Stock Exchange, stipulating the mandatory disclosure of social responsibility for the first time—if a listed company has a major event related to environmental protection and could exert a great impact on the trading price of its shares and derivatives, it should timely disclose the event situation and its possible impact on the company's operation and stakeholders within 2 days from the date of the event. Companies engaged in thermal power generation, iron and steel, cement, electrolytic aluminum, mineral development, and other industries that markedly affect the environment should disclose or separately disclose relevant environmental information in the company's annual social responsibility report. Companies listed on the list of seriously polluting enterprises by the environmental protection department should disclose relevant environmental information within 2 days after the environmental protection department publishes the list.

In December 2008, the Shanghai Stock Exchange further issued the notice, mandating the sample companies of the “corporate governance sector of the Shanghai Stock Exchange,” the companies issuing overseas listed foreign shares and financial companies to disclose the report on the company's performance of social responsibility. In December of the same year, the Shenzhen Stock Exchange also issued the notice, requiring that “listed companies included in the Shenzhen 100 index should disclose their social responsibility reports in accordance with the provisions of the guidelines on the responsibility of listed companies of the Shenzhen Stock exchange, and encourage other companies to disclose their social responsibility reports.”

We obtained the internal control index from the internal control index database of Dibo. The information content of social responsibility reports was completed using Python web crawler and text extraction technology. In addition, the social responsibility reports were attained from Juchao Information using a web crawler, and all texts were performed word segmentation using Python development source “Jieba” Chinese word segmentation module and word frequency statistics. Based on the vector space model (VSM), we used the weight function “TF-IDF” algorithm to construct the word frequency vector of each text, and then the vector angles between the enterprise's social responsibility reports in the year and the social responsibility reports text of the previous year were calculated to measure the text similarity. Furthermore, the sample data of the remaining variables were obtained from the CSMAR database.

5.2. Model setting and variable definition

The following models were involved in this study: Model (1) corresponded to [Hypothesis 1](#), which examined the correlation between ICQ and information content of social responsibility; Models (2) and (3) corresponded to [Hypothesis 2](#), which examined the impact of internal control on the information content of the social responsibility mechanism.

$$\text{Similarity}_{i,t} = \beta_0 + \beta_1 \text{ICQi}_{i,t} + \gamma \text{Controli}_{i,t} + \sum \text{Industry}_{i,t} + \sum \text{Yeari}_{i,t} + \epsilon_{i,t} \quad (1)$$

$$\text{Mfeei}_{i,t} = \alpha_0 + \alpha_1 \text{ICQi}_{i,t} + \alpha_2 \text{Controli}_{i,t} + \sum \text{Industry}_{i,t} + \sum \text{Yeari}_{i,t} + \epsilon_{i,t} \quad (2)$$

$$\text{Similarity}_{i,t} = \gamma_0 + \gamma_1 \text{ICQi}_{i,t} + \gamma_2 \text{Mfeei}_{i,t} + \gamma_3 \text{Controli}_{i,t} + \sum \text{Industry}_{i,t} + \sum \text{Yeari}_{i,t} + \epsilon_{i,t} \quad (3)$$

In Model (1), explaining variable ICQ denotes the internal control quality, measured by the internal control index provided by Dibo internal control index database.

In addition, the explained variable “Similarity” measures the information content of social responsibility reports. Based on [Meng et al. \(2017\)](#), the overlapping or similar information between this year and last year was defined as content without information, while

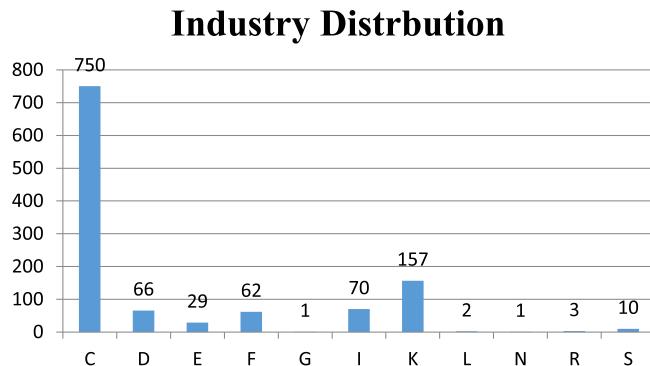


Fig. 1. Sample industry distribution. C: Manufacturing industry D: Electricity, heat, and water industry E: Construction industry F: Wholesale and retail industry. G: Transportation, warehousing, and postal industry I: Information transmission, software, and information technology industry. K: Real estate industry L: Leasing and business services industry N: Water conservancy, environment, and public facilities management industry R: Culture, sports, and entertainment industry S: Comprehensive industry.

different information was defined as content with information, that is, the text with lower similarity had more information content. Of note, we used the social responsibility reports of the same enterprise before and after the year to calculate similarity because the macroeconomic environment, risks, and political policy backgrounds of the same enterprise are similar in a certain period, and the shorter the time interval, the closer the effect of external environmental. Thus, if we compared with the data of the previous year, it would inevitably have a certain similarity with the previous issue to a certain extent and could even directly refer to its previous year's statement. Thus, the influence of this external environment can be well controlled, and the longer the interval of the text disclosure time of social responsibility reports, the higher the difference of the text. Moreover, the difference could be the influence of time trend, and selecting text of similar years for comparison can efficiently evade the influence of this time trend. Hence, when β_1 and δ_1 in the regression model are significantly negative, it supports [Hypothesis 1](#). [Table 2](#) shows the specific calculations process.

The space vector model quantifies the text information by word segmentation, and the similarity is evaluated between texts by calculating the cosine value of the angle between two vectors; this method is intuitive and easy to understand. However, traditional VSM does not consider the syntactic and semantic features of the report. To address this problem, we used the TF-IDF weighting function to decrease the weight of common words in all documents to zero or very close to zero, TF-IDF heavily weights words that are unique to a company, treating them as very "informative" words.

Based on the mediating effect test method ([Wen & Ye, 2014](#)), the agency cost was tested for the existence of a mediating effect between ICQ and text similarity of social responsibility. The specific steps were as follows: (i) the effects of the ICQ on text similarity of social responsibility were tested, and the significance of β_1 in Model (1) was examined; (ii) the effects of the ICQ on agency cost are tested, and the significance of a_1 in Model (2) was examined; (iii) the effects of the ICQ and agency cost on text similarity of social responsibility were analyzed, and the significances of γ_1 and γ_2 in Model (3) were examined. If β_1 , a_1 , γ_1 , and γ_2 in Models (1)–(3) were significant, and the product of $a_1 * \gamma_2$ had the same sign as γ_1 , then agency cost exerted a partial mediating effect, thereby supporting [Hypothesis 2](#).

Based on the relevant studies of [Shen \(2007\)](#), [Li and Zhang \(2017\)](#), [Al-Tuwaijri et al. \(2004\)](#), [Brown and Tucker \(2011\)](#), we selected control variables from three aspects. In this study, Growth, Size, Lev, Roa, Mb were the variables of corporate financial characteristics, while Top1 and Mnghold belonged to corporate governance variables. Firm size is expected to negatively correlate with the similarity, either because large firms are under pressure to modify the CSR to evade political costs of being perceived as being opaque or because small firms could be more likely to cut and paste disclosure from the previous year to save preparation costs. Lev is projected to negatively correlate with the similarity, as companies with high financial risk are more likely to establish a good relationship with creditors and other stakeholders and are more likely to disclose CSR information.

In China, owing to the differences in the historical and cultural environment, many enterprises have relatively high equity concentration. The listed companies with a high concentration usually actively supervise the company's management and operation. Thus, from the standpoint of incentive mechanism, it can be considered that the high concentration will play a positive role in improving and promoting the quality of social responsibility information disclosure.

Table 2
Variable definition.

Variable type	Variable symbol	Variable name and measure
Explained variable	<i>Similarity</i>	According to the research of Brown and Tucker (2011) , Lang and Stice-Lawrence (2015) , based on VSM, the weight function "TF-IDF" was developed. Herein, $TF = \frac{\text{The number of times of a term appearing in the text}}{\text{The total number of terms in the text}}$ (TF: Term Frequency) $IDF = \frac{\text{The total number of texts in the corpus}}{\text{The number of texts including the term}}$ (IDF: Inverse Document Frequency) This method performs word segmentation on text to construct a bag of words, constructs the word frequency vector of each text based on the TF-IDF algorithm, and then uses the following equation to calculate the similarity between two texts (d_1 , d_2), that is, cosine similarity: $(d_1, d_2) = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2} \sqrt{\sum_i y_i^2}}$.
Explaining variable	<i>ICQ</i>	Internal control quality, expressed by Dibo internal control index, plus 1 to take the natural logarithm
Mediating variable	<i>Mfee</i>	Agency cost, Management costs/Operating income
Control variables	<i>Growth</i> <i>Size</i> <i>Lev</i> <i>Roa</i> <i>Mb</i> <i>Top1</i> <i>Mnghold</i> <i>Market</i> <i>Holder</i> <i>Industry</i> <i>Year</i>	The enterprise's growth capacity, the enterprise's annual sales revenue growth rate Enterprise size, total assets at the end of year t take the natural logarithm Asset-liability ratio, the ratio of the book value of liabilities to the book value of assets at the end of year t Return on assets, net profit divided by average total assets Book-to-market ratio, the enterprise's book-to-market ratio in year t The shareholding ratio of the largest shareholder, the number of shares held by the largest shareholder/total number of shares Management shareholding ratio, management shareholding/total number of shares Market index, use Fan Gang index Shareholding ratio of institutional investors <i>Industry</i> virtual variable Virtual variable of the year

Institutional investors are key players in a firm's information environment. Typically, institutional investors demand more corporate disclosures for valuation and monitoring purposes (Bushee and NOE, 2000). The degree of government intervention differs in the regions with different marketization degrees, and the degree of supervision of the media, accounting firms, and other supervision institutions on enterprises differs as well, which exert different effects on enterprise information transparency and then affect the degree of information disclosure. We examined market index's associations with CSR similarity without offering predictions.

6. Empirical analysis

6.1. Descriptive statistics

Table 3 shows the descriptive statistic results of the main variables. The mean of text similarity of the explained variable in the social responsibility report was 0.815 (median: 0.833; standard deviation: 0.109), suggesting that the similarity of social responsibility reports in China between years was higher and the data distribution was reasonable. However, we found some differences between enterprises. The mean of explaining variable ICQ was 6.537 (median: 6.543), the maximum was 6.839, the corresponding internal control score was 946.38, the minimum was 6.082, and the corresponding internal control score was 426.41, suggesting a difference between the ICQ of various enterprises. Moreover, the control variables consist of the relevant characteristic variables inside and outside the enterprise, consistent with previous studies, exhibited a normal distribution and within a reasonable range.

6.2. Correlation analysis

Table 4 presents the correlation coefficients of the main variables. Similarity between ICQ and text similarity of social responsibility was significantly negative at the 1% level, suggesting that without controlling other variables, the higher the ICQ, the lower the text similarity of social responsibility reports, that is, the higher the information content of social responsibility reports, which is consistent with **Hypothesis 1**. In addition, the correlation coefficient indicated that most control variables and similarity variables significantly correlated at the 1% level, reflecting the representativeness and reasonableness of the selection of control variables in this study. Besides, the correlation coefficient between each control variable was basically < 0.5 , suggesting that the empirical model in this study did not have a severe multicollinearity problem. Moreover, the correlation analysis did not control the effects of other variables. Thus, it is also essential to remove the effects of other factors on the text similarity of social responsibility reports through multiple regression analysis to attain more reliable inferences.

6.3. Regression analysis

6.3.1. Information content of ICQ and social responsibility reports: Based on the overall perspective

Table 5 presents the regression results of the impact of ICQ on the information content of social responsibility reports. First, we performed regression on the ICQ and text similarity of social responsibility separately to determine whether the ICQ affected text similarity of social responsibility without considering other factors. The results are shown in column (1) of **Table 5**, that is, the ICQ and text similarity of social responsibility significantly negatively correlated at the 1% level, which initially supported **Hypothesis 1**. Second, we gradually added industry annual fixed effects and enterprise financial characteristics variables based on column (1), columns (2) and (3) in **Table 5**, demonstrate that after considering industry annual fixed effects and enterprise financial characteristics, the ICQ and text similarity of social responsibility significantly negatively correlated at the 1% level. In addition, we added the corporate governance variable based on column (4); the results were similar to column (3), and both significantly negatively correlated at the 1% level. Finally, after controlling the enterprise's financial factors and governance variables, we considered the enterprise's external characteristics variables; the results are shown in column (5) of **Table 5**. The ICQ and text similarity of social responsibility significantly negatively correlated at the 1% level. As shown in **Table 5**, both the ICQ and text similarity of social responsibility significantly negatively correlated at the 1% level, irrespective of the presence of control variables, suggesting that the ICQ can

Table 3

Descriptive statistics of variables.

Variable	Observation	Mean	Standard deviation	Minimum	Median	Maximum
Similarity	1151	0.815	0.109	0.476	0.833	0.983
ICQ	1151	6.537	0.122	6.082	6.543	6.839
Growth	1151	0.180	0.335	-0.429	0.130	1.957
Mfee	1151	0.083	0.076	0.009	0.067	0.951
Size	1151	22.990	1.293	20.330	22.850	26.040
Lev	1151	0.501	0.207	0.063	0.532	0.895
Roa	1151	0.055	0.053	-0.123	0.043	0.231
Mb	1151	0.611	0.269	0.108	0.617	1.131
Top1	1151	0.354	0.155	0.073	0.336	0.743
Mnghold	1151	0.053	0.126	0.000	0.000	0.536
Market	1151	7.978	1.611	2.760	8.180	10.200
Holder	1151	0.059	0.048	0.002	0.049	0.241

Table 4
Correlation analysis of main variables.

Variable	Similarity	ICQ	Growth	Size	Lev	Roa	Top1	Mnghold	Mb	Market	Holder
Similarity	1										
ICQ	-0.165***	1									
Growth	-0.027	0.181***	1								
Size	-0.235***	0.306***	0.032	1							
Lev	-0.157***	0.063**	0.037*	0.508***	1						
Roa	0.031	0.364***	0.199***	-0.076***	-0.457***	1					
Top1	-0.089***	0.161***	0.002	0.235***	0.052**	0.071***	1				
Mnghold	0.190***	-0.079***	0.084***	-0.307***	-0.272***	0.136***	-0.143***	1			
Mb	-0.163***	0.098***	-0.070***	0.608***	0.534***	-0.379***	0.097***	-0.240***	1		
Market	0.075***	-0.074**	-0.035	0.023	-0.077***	0.078***	0.052**	0.101***	-0.083***	1	
Holder	-0.113***	0.152***	0.089***	0.028	-0.031	0.209***	-0.136***	-0.052**	-0.154***	-0.027	1

*, **, and *** signify 10%, 5%, and 1% significance levels, respectively.

Table 5

Internal control quality and information content of social responsibility reports.

Variable	(1) <i>Similarity</i>	(2) <i>Similarity</i>	(3) <i>Similarity</i>	(4) <i>Similarity</i>	(5) <i>Similarity</i>
<i>ICQ</i>	-0.148*** (-5.67)	-0.151*** (-5.83)	-0.098*** (-3.26)	-0.090*** (-3.04)	-0.084*** (-2.82)
<i>Growth</i>			0.007 (0.71)	0.000 (0.03)	0.000 (0.03)
<i>Size</i>			-0.012*** (-3.18)	-0.007* (-1.83)	-0.007* (-1.84)
<i>Lev</i>			-0.060*** (-2.59)	-0.037 (-1.63)	-0.034 (-1.48)
<i>Roa</i>			-0.032 (-0.40)	-0.034 (-0.43)	-0.009 (-0.11)
<i>Mb</i>			-0.024 (-1.26)	-0.030 (-1.61)	-0.031 (-1.64)
<i>Top1</i>				-0.046** (-2.30)	-0.054*** (-2.64)
<i>Mnghold</i>				0.167*** (6.39)	0.158*** (5.99)
<i>Market</i>					0.002 (0.93)
<i>Holder</i>					-0.001** (-2.03)
<i>Industry</i>	No	Yes	Yes	Yes	Yes
<i>Year</i>	No	Yes	Yes	Yes	Yes
<i>Constant</i>	1.785*** (10.43)	1.800*** (10.65)	1.759*** (9.37)	1.595*** (8.59)	1.586*** (8.53)
<i>N</i>	1151	1151	1151	1151	1151
<i>F</i>	32.14	10.09	11.27	13.06	12.10
<i>Adj_R²</i>	0.03	0.09	0.13	0.17	0.17

*, **, and *** imply 10%, 5%, and 1% significance levels, respectively.

Table 6

Quantile regression (1).

Variable	(1)	(2)	(3)	(4)	(5)
<i>ICQ</i>	Q30 0.052 (1.24)	Q40 0.100** (2.25)	Q50 0.115** (2.20)	Q60 0.169*** (2.67)	Q70 0.232*** (3.66)
<i>Growth</i>	-0.002 (-0.12)	0.003 (0.18)	-0.005 (-0.24)	-0.005 (-0.22)	0.0001 (0.00)
<i>Size</i>	0.006 (1.27)	0.005 (0.95)	0.007* (1.66)	0.009 (1.37)	0.005 (0.55)
<i>Lev</i>	0.024 (0.75)	0.071*** (2.98)	0.096*** (4.17)	0.071* (1.89)	0.125*** (2.62)
<i>Roa</i>	0.024 (0.22)	-0.010 (-0.08)	0.002 (0.01)	0.032 (0.20)	0.177 (0.97)
<i>Mb</i>	0.001** (2.05)	0.001** (2.00)	0.001** (2.11)	0.001** (2.23)	0.001 (1.36)
<i>Top1</i>	-0.118*** (-3.65)	-0.157*** (-5.48)	-0.189*** (-5.43)	-0.233*** (-6.32)	-0.236*** (-5.25)
<i>Mnghold</i>	0.049 (1.53)	0.041 (1.28)	0.024 (0.79)	0.050 (1.14)	0.081 (1.34)
<i>Market</i>	0.001 (0.84)	0.001 (0.77)	0.001 (0.87)	0.0008 (0.67)	0.003** (1.97)
<i>Holder</i>	-0.0002 (-0.05)	-0.0003 (-0.07)	0.001 (0.39)	-0.004 (-1.23)	-0.004 (-0.64)
<i>Industry</i>	Yes	Yes	Yes	Yes	Yes
<i>Year</i>	Yes	Yes	Yes	Yes	Yes
<i>Constant</i>	0.604** (1.97)	0.341 (1.15)	0.216 (0.67)	-0.104 (-0.25)	-0.458 (-1.06)
<i>N</i>	1151	1151	1151	1151	1151
<i>Pseudo R²</i>	0.08	0.10	0.10	0.11	0.11

*, **, and *** signify 10%, 5%, and 1% significance levels, respectively.

decrease the text similarity between social responsibility reports, that is, it would increase the information content disclosed by social responsibility reports, which corroborates [Hypothesis 1](#). Besides, this reflects that effective internal control can enhance the enterprise's internal governance, alleviate management's agency problems in information disclosure, and prevent selective disclosure of information. In the regression, we also tested the variance inflation factor (VIF) of each variable; VIFs were all < 5 , suggesting that the model did not have serious multicollinearity problems.

6.3.2. ICQ and information content of social responsibility reports: Quantile regression

To further validate the correlation between ICQ and information content of social responsibility reports, we performed 30%, 40%, 50%, 60%, and 70% quantile regression. Precisely, we took the reciprocal of the similarity variable based on the above. The higher the reciprocal of the similarity, that is, the smaller the original similarity value, the higher the difference between the texts. Considering that a nonlinear relationship could exist between ICQ and difference degree of social responsibility text, it is essential to recalculate the difference degree variables into regression. Based on the similarity of the previous calculation, we also took the inverse value as the reciprocal value; the higher the inverted value was, the higher the difference of the text was. The general regression model focuses on the effect of X on the conditional expectation of Y. Indeed, it is mean regression, which is difficult to reflect the whole picture of conditional distribution. Quantile regression is based on interval regression, which can completely understand conditional distribution. Moreover, quantile regression uses the weighted average of the absolute value of the residual as the objective function of minimization, which is not easily affected by extreme values and is more robust.

[Table 6](#) shows that the coefficient of ICQ increases gradually from 30% to 70%, and the significance also exhibits an increasing trend. However, in the 30% quantile, the regression coefficient between ICQ and difference degree of social responsibility text is not significant. [Table 7](#) shows that with an increase of quantile (30%–40%–50%–60%–70%), the quantile regression coefficients of ICQ displayed an upward trend and were significantly positive. The conditional distribution of ICQ on the similarity of social responsibility texts was higher in the parts with lower similarity than in the parts with higher similarity, suggesting that the higher the quality of internal control, the lower the similarity of social responsibility texts, further supporting the conclusion of [Hypothesis 1](#).

6.3.3. The information content test of social responsibility text similarity

To test the similarity of social responsibility texts, based on the similarity evaluated above, we took the reciprocal of the similarity variable. The higher the reciprocal, the higher the difference of the text. For reference ([Yin et al., 2019](#)), we obtained data related to stock bar posts of listed companies from China Research Data Service Platform (CNRDS) and used the difference between the number of stock bar posts in 3 days after the release of the social responsibility report [0,2] and the number of stock bar posts in the last 3 days before the release of the social responsibility report [−3, −1] to measure the change DIF of investors' attention to the company. The higher the information content of social responsibility, the more it can catch the investors' attention. As shown in [Table 8](#), a significant

Table 7
Quantile regression (2).

	(1)	(2)	(3)	(4)	(5)
Variable	Q30 <i>Similarity</i>	Q40 <i>Similarity</i>	Q50 <i>Similarity</i>	Q60 <i>Similarity</i>	Q70 <i>Similarity</i>
<i>ICQ</i>	0.025** (2.46)	0.045** (2.56)	0.054** (2.06)	0.077** (2.39)	0.100*** (2.84)
<i>Growth</i>	−0.001 (−0.15)	0.001 (0.18)	−0.002 (−0.30)	−0.002 (−0.19)	−0.000 (−0.00)
<i>Size</i>	0.003 (1.06)	0.002 (0.54)	0.003 (0.82)	0.004 (1.04)	0.002 (0.57)
<i>Lev</i>	0.012 (1.01)	0.033*** (2.98)	0.045*** (3.15)	0.035 (1.44)	0.052** (2.35)
<i>Roa</i>	0.015 (0.40)	0.002 (0.05)	0.001 (0.02)	0.015 (0.21)	0.079 (0.93)
<i>Mb</i>	0.0003** (2.16)	0.0003* (1.84)	0.0004** (2.37)	0.0004** (2.29)	0.0004** (2.26)
<i>Top1</i>	−0.055*** (−4.23)	−0.074*** (−5.20)	−0.090*** (−6.04)	−0.109*** (−4.79)	−0.112*** (−4.61)
<i>Mnghold</i>	0.023* (1.83)	0.020 (1.34)	0.012 (1.18)	0.020 (1.19)	0.036* (1.76)
<i>Market</i>	0.001 (0.92)	0.001 (0.68)	0.001 (0.68)	0.0004 (0.49)	0.001* (1.90)
<i>Holder</i>	−0.0001 (−0.10)	−0.0003 (−0.26)	0.001 (0.41)	−0.002 (−1.59)	−0.002 (−0.90)
<i>Industry</i>	Yes	Yes	Yes	Yes	Yes
<i>Year</i>	Yes	Yes	Yes	Yes	Yes
<i>Constant</i>	0.816*** (9.03)	0.711*** (5.65)	0.644*** (4.20)	0.502*** (2.99)	0.382** (2.00)
<i>N</i>	1151	1151	1151	1151	1151
<i>Pseudo R</i> ²	0.09	0.10	0.11	0.11	0.12

*, **, and *** signify 10%, 5%, and 1% significance levels, respectively.

Table 8

The information content test of social responsibility text similarity.

Variable	(1) Dif	(2) Dif	(3) Dif
<i>Similarity</i>	0.523*** (6.05)	0.140* (1.78)	0.154** (2.12)
<i>Growth</i>		-0.101** (-2.26)	-0.074* (-1.81)
<i>Size</i>		0.399*** (22.02)	0.459*** (26.49)
<i>Lev</i>		-0.352*** (-3.40)	-0.463*** (-4.85)
<i>Roa</i>		-2.963*** (-8.68)	-2.354*** (-7.46)
<i>Mb</i>		-0.836*** (-8.91)	-1.081*** (-12.35)
<i>Top1</i>			-0.016*** (-17.52)
<i>Mnghold</i>			-1.119*** (-8.80)
<i>Holder</i>			-0.028*** (-9.28)
<i>Market</i>			-0.015* (-1.73)
<i>Industry</i>			
<i>Year</i>			
<i>Constant</i>	6.331*** (24.90)	-0.974** (-2.49)	-1.095*** (-2.95)
<i>N</i>	2134	2134	2134
<i>F</i>	44.63	67.95	83.71
<i>Adj_R²</i>	0.32	0.48	0.55

*, **, and *** signify 10%, 5%, and 1% significance levels, respectively.

Table 9

Mechanism test based on the agency cost.

Variable	(1) <i>Similarity</i>	(2) <i>Mfee</i>	(3) <i>Similarity</i>
<i>ICQ</i>	-0.084*** (-2.82)	-0.032* (-1.73)	-0.068** (-2.10)
<i>Mfee</i>			0.083* (1.83)
<i>Growth</i>	0.000 (0.03)	-0.024*** (-4.25)	0.003 (0.28)
<i>Size</i>	-0.007* (-1.84)	0.004 (1.61)	-0.010** (-2.36)
<i>Lev</i>	-0.034 (-1.48)	-0.068*** (-4.80)	-0.010 (-0.42)
<i>Roa</i>	-0.009 (-0.11)	-0.287*** (-5.87)	0.022 (0.25)
<i>Mb</i>	-0.031 (-1.64)	-0.091*** (-7.83)	-0.025 (-1.09)
<i>Top1</i>	-0.054*** (-2.64)	-0.010 (-0.76)	-0.046** (-2.24)
<i>Mnghold</i>	0.158*** (5.99)	-0.015 (-0.95)	0.154*** (5.79)
<i>Market</i>	0.002 (0.93)	0.000 (0.02)	0.000 (0.16)
<i>Holder</i>	0.001** (-2.03)	0.016 (0.40)	-0.145** (-2.20)
<i>Industry</i>	Yes	Yes	Yes
<i>Year</i>	Yes	Yes	Yes
<i>Constant</i>	1.586*** (8.53)	0.313*** (2.73)	1.427*** (7.07)
<i>N</i>	1151	1151	1151
<i>F</i>	12.10	31.55	10.14
<i>Adj_R²</i>	0.17	0.36	0.14

*, **, and *** signify 10%, 5%, and 1% significance levels, respectively.

positive correlation exists between the similarity of social responsibility texts and the number of stock bar posts, regardless of whether the control variables were added. Hence, when the similarity of social responsibility text is lower, the information content is higher, and investors focus more on the social responsibility information.

6.3.4. ICQ and information content of social responsibility reports: Mechanism test

Table 9 shows the mediating effect test results of agency cost. Column (1) demonstrates that the ICQ exerts a negative inhibitory effect on text similarity of social responsibility, which is -0.084 , and is significant at the 1% level. Column (2) establishes that the regression coefficient of internal control ICQ and agency cost Mfee is -0.032 , and it is significantly negative at the level of 10%, indicating that the high level of ICQ reduces the enterprise's agency cost; similarity between internal control ICQ, agency Cost Mfee and text similarity of social responsibility in column (3) are both significant, regression coefficient (γ_1) is -0.068 , and is significant at the 5% level. Based on the conventional practice (Wen & Ye, 2014), we only compared symbol of $-0.032*0.083$ with that of -0.068 . Same symbols indicated that agency cost played a partial mediating effect in the influence of internal control on information content of social responsibility, thereby validating Hypothesis 2.

6.4. Stability test

6.4.1. Measurement of replacing text similarity of social responsibility

The higher the differences in the number of pages, words, and paragraphs between social responsibility reports, and the more new or decreased disclosures, and the less similar the social responsibility information disclosure of the current year to the previous year. Columns (1), (2), and (3) of Table 10 use the number of pages of social responsibility reports in the current year minus that in the previous year, which are expressed as $\Delta Page$, the number of words of social responsibility reports in the current year minus that in the previous year, which is expressed as $\Delta Word$, and the paragraph of social responsibility reports in the current year minus that in the previous year, which is expressed as $\Delta Line$, respectively; they are taken as the replacement variables of information content of social responsibility. All results revealed that the regression coefficients of ICQ were significantly positive, suggesting that the better ICQ, the more incremental information of the social responsibility information disclosure, implying that the information content of social responsibility reports is higher, and the assumptions made in this study are valid.

In addition, by comparing the horizontal text similarity with other companies in the same industry, we captured the difference between the disclosure content of this company and other companies. Specific practices will be in the same industry besides the i company with years of all other enterprise social responsibility reports, using the vector space model, the TF-IDF weight function, and synthesis of word frequency vector. As a benchmark, we calculated the company's social responsibility text and synthesis of word frequency vector angle to measure the similarity of a text; the higher the similarity, the higher the degree of coincidence between the

Table 10

Measurement methods of replacement variables.

Variable	(1) $\Delta Page$	(2) $\Delta Word$	(3) $\Delta Line$	(4) Similarity
<i>ICQ</i>	0.005* (1.93)	0.032** (2.55)	0.002*** (2.71)	-0.021* (-1.77)
<i>Growth</i>	0.062 (0.10)	-0.796 (-0.29)	-0.165 (-0.98)	-0.010*** (-2.95)
<i>Size</i>	0.052 (0.21)	1.184 (1.08)	0.015 (0.22)	0.009*** (6.21)
<i>Lev</i>	-0.118 (-0.08)	-5.361 (-0.80)	-0.110 (-0.27)	-0.028** (-2.57)
<i>Roa</i>	7.026 (1.30)	1.198 (0.05)	0.524 (0.36)	-0.029 (-0.66)
<i>Mb</i>	0.116 (0.09)	-1.636 (-0.30)	-0.052 (-0.16)	0.015* (1.79)
<i>Top1</i>	0.540 (0.40)	-8.66 (-1.48)	-0.157 (-0.44)	-0.017* (-1.79)
<i>Mnghold</i>	-1.031 (-0.59)	-4.997 (-0.65)	-0.315 (-0.68)	0.004 (0.19)
<i>Market</i>	-0.096 (-0.73)	-0.245 (-0.43)	-0.017 (-0.47)	0.003*** (3.76)
<i>Holder</i>	0.014 (0.32)	-0.000 (-0.00)	-0.003 (-0.23)	0.053* (1.79)
<i>Industry</i>	Yes	Yes	Yes	Yes
<i>Year</i>	Yes	Yes	Yes	Yes
<i>Constant</i>	-3.790 (-0.79)	-35.85* (-1.71)	-1.267 (-0.99)	0.057 (0.72)
<i>N</i>	1084	1073	1073	1854
<i>F</i>	1.72	2.11	1.45	15.67
<i>Adj_R²</i>	0.01	0.02	0.02	0.073

*, **, and *** signify 10%, 5%, and 1% significance levels, respectively.

Table 11
Endogenous treatment.

Fixed-effect Model (1)		Tool variable method (2)				First-order differential model (3)		Information disclosure policy (4)		Missing variable (5)	
Explaining variable	Similarity	First stage	ICQ	Second stage	Similarity	Explaining variable	ΔSimilarity	Explaining variable	Similarity	Explaining variable	Similarity
14	ICQ	-0.072** (-2.22)	MICQ (14.96)	1.071*** (-2.33)	ICQ (-0.31)	-0.168** (-2.75)	ΔICQ (-0.57)	-0.072*** (-2.75)	ICQ (-0.12)	-0.085*** (-2.72)	ICQ (-3.09)
	Growth	-0.001 (-0.07)	Growth (3.20)	0.027*** (0.31)	Growth (-1.11)	0.003 (-1.41)	ΔGrowth (0.33)	-0.004 (0.14)	Regulation (0.03)	-0.001 (-0.12)	Growth (0.06)
	Size	-0.008** (-1.87)	Size (12.50)	0.044*** (10.47)	Size (0.73)	-0.005 (0.73)	Δsize (0.52)	0.002 (0.52)	Growth (-1.48)	0.000 (-0.034)	Size (-0.03)
	Lev	-0.029 (-1.25)	Lev (-0.63)	-0.013 (-0.63)	Lev (-1.41)	-0.032 (-1.41)	Δlev (0.33)	0.016 (-1.55)	Size (-1.70)	-0.007 (-0.039*)	Lev (-1.70)
	Roa	-0.006 (-0.07)	Roa (10.47)	0.734*** (0.73)	Roa (0.73)	0.075 (0.73)	Δroa (0.52)	0.052 (0.52)	Lev (-1.48)	-0.034 (-0.03)	Roa (-0.04)
	Mb	-0.029 (-1.32)	Mb (-1.09)	-0.020 (-1.29)	Mb (-1.29)	-0.025 (-1.29)	ΔMb (-1.88)	-0.047* (-1.88)	Roa (-0.14)	-0.011 (-0.14)	Mb (-2.42)
	Top1	-0.056*** (-2.69)	Top1 (2.09)	0.039** (-2.30)	Top1 (-2.30)	-0.048** (-2.30)	ΔTop1 (-0.31)	-0.026 (-0.31)	Mb (-1.50)	-0.032 (-0.37)	Top1 (5.37)
	Mnghold	0.155*** (5.85)	Mnghold (0.88)	0.021 (0.88)	Mnghold (6.05)	0.159*** (6.05)	ΔMnghold (-1.07)	-0.089 (-1.07)	Top1 (-2.63)	-0.054*** (-2.63)	Mnghold (-1.27)
	Market	0.002 (0.77)	Market (1.07)	0.002 (0.73)	Market (0.73)	0.001 (0.73)	ΔMarket (0.74)	0.006 (0.74)	Mnghold (5.99)	0.158*** (5.99)	Market (0.70)
	Holder	0.001* (-1.90)	Holder (0.99)	0.060 (0.99)	Holder (-1.84)	-0.001* (-1.84)	ΔHolder (-0.21)	-0.000 (-0.21)	Market (0.93)	0.002 (0.93)	Holder (-1.04)

information in the company's social responsibility report and that of other companies in the same industry, but the less the information content. On the contrary, the more information content. As shown in column (4) of Table 10, the quality of internal control significantly negatively correlates with the similarity of social responsibility report text at the 10% level; thus, the hypothesis in this paper is still valid.

6.4.2. Impact of using fixed-effects model to control individual level

To regulate the enterprise's individual level impact and test the robustness of the research conclusions, we used the panel fixed-effects model to perform regression on the correlation between ICQ and social responsibility text information similarity. Column (1) of Table 11 shows the panel fixed-effects model regression results. The regression results revealed that the ICQ and social responsibility reports text similarity significantly negatively correlated at the level of 5%, and that the regression conclusions assumed in this study are still valid.

6.4.3. Use tool variable to eliminate the endogenous problem

A crucial factor affecting this study's conclusions was obtained from endogenous issues. The better the social responsibility information disclosure level of an enterprise, the better likelihood of its social responsibility performance, which, in turn, contributes to the enterprise's internal control construction and improvement of the level of internal control. To control its impact, based on previous studies (Faccio et al., 2011; Yang & Chen, 2013; Yang et al., 2019), the annual mean ICQ (MICQ) of the enterprise's industry was taken as the tool variable because enterprises in the same year and industry had some elements in common in the operating environment and enterprise characteristics, and a correlation existed between their internal control levels.

In addition, column (2) of Table 11 establishes that the symbol of the MICQ is significantly positive at the 1% level; hence, it satisfies the principle of relevance. However, there is no evidence that the industry average internal control level affects the enterprise's social responsibility text information disclosure; hence, it satisfies the principle of exogenousness. Column (2) of Table 11 shows the regression results of the second stage. The regression coefficient of ICQ remained significantly negative at the 5% level. Meanwhile, the F statistic of Cragg–Donald Wald was used to determine whether the tool variable was weak. The F value obtained was 223.682, which is > 10 , suggesting the validity of the tool variable selection and that of the regression conclusion assumed in this study.

6.4.4. Use the first-order differential model to solve the problem of missing variables

The first-order differential model was used for endogeneity testing to eliminate the potential impact of missing variables that do not change over time on the results of the text regression. After the first-order differential was performed on all continuous variables, the

Table 12
Cross-section analysis.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Large-scale corporations (Big = 1) Similarity	Small- and medium-sized enterprises (Big = 0) Similarity	State-owned enterprises (SOE = 1) Similarity	Non-state enterprises (SOE = 0) Similarity	High media attention (Media = 1) Similarity	Low media attention (Media = 0) Similarity
ICQ	-0.084*** (-2.63)	-0.068 (-0.81)	-0.057 (-1.21)	-0.111*** (-2.77)	-0.088*** (-2.61)	-0.010 (-0.15)
Growth	-0.011 (-1.09)	0.054** (2.58)	-0.014 (-1.12)	0.020 (1.44)	-0.001 (-0.02)	0.005 (0.21)
Size	-0.014*** (-3.41)	0.008 (0.75)	-0.006 (-1.03)	-0.005 (-0.98)	-0.008* (-1.78)	0.001 (0.01)
Lev	-0.056** (-2.13)	0.002 (0.03)	-0.049 (-1.39)	-0.040 (-1.32)	-0.034 (-1.28)	-0.030 (-0.65)
Roa	-0.083 (-0.87)	0.085 (0.54)	0.191 (1.63)	-0.206* (-1.87)	-0.025 (-0.27)	0.092 (0.53)
Mb	-0.026 (-1.23)	-0.086** (-2.04)	-0.005 (-0.18)	-0.067** (-2.52)	-0.028 (-1.25)	-0.055 (-1.36)
Top1	-0.040* (-1.81)	-0.152*** (-3.23)	-0.040 (-1.42)	-0.094*** (-3.16)	-0.037 (-1.58)	-0.114** (-2.52)
Mnghold	0.160*** (4.79)	0.126*** (3.01)	0.137*** (3.78)	0.177*** (4.62)	0.193*** (5.93)	0.082* (1.85)
Market	-0.001 (-0.28)	0.012** (2.24)	0.014** (2.31)	-0.007** (-2.36)	0.002 (0.91)	-0.001 (-0.20)
Holder	0.034 (0.46)	-0.737*** (-5.48)	-0.209** (-2.19)	-0.087 (-0.96)	-0.049 (-0.67)	-0.578*** (-3.68)
Industry	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Constant	1.751*** (8.80)	1.106* (1.94)	1.241*** (4.21)	1.783*** (7.00)	1.594*** (7.60)	0.995** (2.04)
N	936	215	560	591	911	240
F	11.00	6.31	6.37	8.29	9.47	3.57
Adj_R ²	0.18	0.32	0.16	0.21	0.16	0.16

*, **, and *** signify 10%, 5%, and 1% significance levels, respectively.

regression results demonstrated that first-order differential explaining variable ΔICQ and first-order differential explained variable $\Delta Similarity$ in column (3) of Table 11 significantly negatively correlated at the 1% level, suggesting the validity of the regression conclusions assumed in this study.

6.4.5. Exclusions of effects of external policies

On April 3, 2014, the China Securities Regulatory Commission released *Memorandum No. 21 of information disclosure: issues related to periodic reports*, which mandated that CSR reports should provide complete details about the current status, issues observed and plans of social responsibility based on the industrial characteristics, while selective disclosures of good news should be avoided; it could affect the content of social responsibility information disclosed. To exclude the effects of external policies, the dummy variable "Regulation" was set, took 1 in 2014, 2015, 2016, and 2017, and 0 in other years. Column (4) of Table 11 shows the regression results considering the impact of the information disclosure policy. The regression results revealed that the ICQ was negative and significant at the level of 1%, suggesting that even after controlling the information disclosure policy factors, ICQ exerted a significant negative impact on text similarity of social responsibility and that the regression conclusions assumed in this study were still valid.

6.4.6. Omitting variables problem

Dyer et al. (2017) reported that 10-K disclosures over time showed trends with increments in length, boilerplate, stickiness, and redundancy. If this is also true with CSR reports, the lower text similarity can be caused by the general increase in length but not necessarily with useful information so the information content does not necessarily improve. In addition, we controlled the length of disclosure in the tests to mitigate the issue. Column (5) of Table 11 shows the regression results of adding the social responsibility report length (page). The coefficient of internal control quality ICQ is still significantly negative at the 1% level, suggesting the validity of the regression conclusions assumed in this study.

7. Further analysis

7.1. Cross-section analysis

7.1.1. Effects of enterprise size

In 2008, China promulgated the "Basic Standards for Enterprise Internal Control" and supporting guidelines, necessitating that it is executed on the listed enterprises in the main board, and inspiring non-listed large- and medium-sized enterprises to implement it. Then, it is progressively implemented on the listed enterprises in small and medium board and growth enterprises market. Iliev (2010) claimed that the impact of executing SOX in large-scale corporations is better than that of small- and medium-sized enterprises. The equity, personnel, and organizational structure of large-scale corporations are relatively complex, and the business types of enterprises are diverse, rendering implementing an internal control scheme in large-scale corporations easier than that in small- and medium-sized enterprises. Thus, whether the effects of ICQ on text similarity of social responsibility differed between different enterprise sizes was tested empirically in this study. As shown in columns (1) and (2) of Table 12, a big virtual variable was set and grouped per the business income of the industry in the year; if the business income exceeded the industry income in the year, Big was 1, otherwise it was 0. The empirical results revealed that the ICQ in column (1) was significantly negative at the 1% level, and the ICQ in column (2) was negative, but not significant, suggesting that compared with small- and medium-sized enterprises, the enactment of internal control in large-scale corporations could exert its effect better, that is, high-level internal control could decrease the text similarity of social responsibility and enhance the information content of social responsibility reports.

7.1.2. Effects of property right nature

In the context of different property rights, internal control plays different roles. Based on the principal–agent theory, enterprises have complex principal–agent relationships; however, enterprises with different property right natures have multiple separation degrees between corporate management and operating rights, causing the principal–agent problem to vary in severity. As the separation degree between the claim right and the final control right of the remaining capital in the SOEs is higher than that of non-state enterprises, the agency cost of SOEs would be higher than that of non-state enterprises. In addition, the internal control of non-state enterprises can restrain the unproductive investment of enterprises (Wang et al., 2015b) and improve the quality of accounting information (Wang et al., 2015a). The information disclosure of non-state enterprises is the result of internal control and incentives. The information disclosure of SOEs is more of political pressure. Internal control in non-state enterprises exerts a higher impact on improving social responsibility quality of information disclosure (Li & Zhang, 2017). Hence, whether the impact of ICQ on text similarity of social responsibility is different in the context of different property rights was tested empirically in this study. As shown in columns (3) and (4) of Table 12, the SOE virtual variable was set and grouped according to the type of the ultimate controller of the enterprise. If the ultimate controller were an SOE, it would be 1; otherwise it would be 0. The empirical results revealed that the ICQ in column (4) significantly negatively correlated at the 1% level, and the ICQ in column (3) was negative, but not significant, suggesting that compared with state-controlled enterprises, the internal control of non-state-controlled enterprises exerted a more significant effect on the social responsibility text information, ICQ can decrease the text similarity of social responsibility more, that is, it could increase the information content of social responsibility reports.

7.1.3. Effects of media attention

As an external monitoring method, media attention could alleviate the information asymmetry and principal–agent difficulties

inside and outside the enterprise, and could affect the reputation of the government, company, or manager to alter the manager's behavior to affect internal governance (Dyck et al., 2008; Fang & Peress, 2009; Yu et al., 2011), which could effectively decrease agency cost, and the level of corporate information disclosure also increases with the increase of media attention. As a robust external supervision force, the media can prompt reputation-conscious professional managers to enhance their management, regulate their behavior, and establish a good image of a citizen to appear in front of the public (Bloomfield & Wilks, 2000), which can achieve governance functions, such as augmenting operational efficiency and correcting violations, so that under the restraints of reinforcing internal management, enterprises could increase their willingness and quality to disclose their social responsibility implementation to improve their image and reputation and shift their focus. Thus, media attention exerts a positive impact on the CSR level. The augmentation of social responsibility information quality is attained under the promotion of media attention. Hence, whether the impact of ICQ on text similarity of social responsibility differs under different media attention situations was tested empirically in this study. As shown in columns (5) and (6) of Table 12, the "Media virtual" variable was set and grouped according to the median of the number of media reports in the industry in the year. If the number of media reports exceeded the median of the number of media reports in the industry in the year, the Media was 1; otherwise it was 0. The empirical results revealed that the ICQ in column (5) significantly negatively correlated at the 1% level, and the ICQ in column (6) was negative, but not significant, suggesting that compared with enterprises with low media attention, the higher media attention can promote the corporate governance effect of internal control, that is, the higher the media attention is, the more ICQ can decrease the text similarity of social responsibility, that is, it can enhance information content of social responsibility reports.

7.2. Heterogeneity analysis of mechanism

To elucidate the difference of agency cost mechanism in different situations, we performed heterogeneity analysis on the agency cost mechanism. As in small- and medium-sized enterprises, SOEs, and enterprises with intensive public attention, its ICQ exerted no significant impact on text similarity of social responsibility, making it necessary to stop the mediating effect test (Wen & Ye, 2014). Hence, only the heterogeneity of agency cost in large-scale corporations, non-state enterprises, and enterprises with intensive public attention was considered in this study. Table 13 shows the empirical results. In large-scale corporations, column (1) demonstrates that the ICQ exerts a negative inhibitory effect on text similarity of social responsibility (-0.084), and it is significant at the 1% level. Column (2) shows that the internal control ICQ and agency cost Mfee are significantly negative at the 1% level, suggesting that the high-level ICQ in large-scale enterprises decreases the company's agency cost. Column (3) shows that "Similarity" between the internal control ICQ, agency cost Mfee, and text similarity of social responsibility are both significant, and the regression coefficient of the internal control ICQ is -0.077 , which is significant at the 5% level. Based on previous studies (Wen & Ye, 2014), we compared the symbol of $-0.063*0.116$ with that of -0.077 ; their similarity indicated that agency cost played a partial mediating effect in the impact of internal control on the information content of social responsibility.

In non-state enterprises, column (4) demonstrates that the internal control ICQ and text similarity of social responsibility are significantly negatively correlated at the 1% level (-0.111); however, the internal control ICQ and agency cost Mfee in column (5) are not significantly correlated. In addition, the agency cost Mfee and text similarity of social responsibility in column (6) are not significant, as the agency cost of non-state enterprises is relatively small, and the role of internal control in agency problem is feebler than that of SOEs, necessitating the Bootstrap method to test whether the indirect effect was significant. The Bootstrap method in the last row of column (5) tests the Z-value of the indirect effect of the product of the two coefficients. The result is -1.72 and is significant at the 10% level, suggesting that the indirect effect of the product of the two coefficients is significant. Furthermore, the regression coefficient between the internal control ICQ and text similarity of social responsibility in column (6) is not significant. All the abovementioned tests supported that the agency costs in non-state enterprises are full mediating. The abovementioned test showed that if the quality of internal control intends to affect the text similarity of social responsibility, it must be completed through an intermediary of agency cost.

In enterprises with intensive public attention, column (7) describes that the ICQ and text similarity of social responsibility are significantly negatively correlated at the 1% level (-0.088) and are significant at the 1% level. Column (8) demonstrates that the agency cost and internal control of the company are significantly negative at the level of 10%, suggesting that when the media attention of the enterprise is high, its ICQ decreases the company's agency cost. In column (9), the ICQ, agency cost, and text similarity of social responsibility are significant, and the regression coefficient of internal control is -0.065 , which is significant at the level of 10%. Based on previous studies (Wen & Ye, 2014), we compared the symbol of $-0.037*0.093$ with that of -0.065 ; if they were similar, agency cost played a partial mediating effect in the influence of internal control on the information content of social responsibility.

8. Conclusions and suggestions

8.1. Conclusions

In this study, the listed enterprises in the A-share market in 2006–2017 were taken as the research objects, and from the standpoint of text similarity, the impact of ICQ on the information content of social responsibility reports was tested empirically. The empirical results suggest that high-quality internal control can improve the information content of social responsibility reports. The conclusions remain valid even after adopting quantile regression. After replacing the explained variables and using the fixed-effects model, tool variable, first-order differential model, and stability test that excludes policy influence, the main conclusions remain valid. In addition,

Table 13

Heterogeneity analysis of mechanism.

Variable	Large-scale corporations			Non-state enterprises			High media attention		
	(1) Similarity	(2) Mfee	(3) Similarity	(4) Similarity	(5) Mfee	(6) Similarity	(7) Similarity	(8) Mfee	(9) Similarity
<i>ICQ</i>	-0.084*** (-2.63)	-0.063*** (-3.54)	-0.077** (-2.39)	-0.111*** (-2.77)	0.029 (0.80)	-0.023 (-0.58)	-0.088*** (-2.61)	-0.037* (-1.67)	-0.065* (-1.83)
<i>Mfee</i>			0.116* (1.94)			-0.055 (-1.08)			0.093* (1.76)
<i>Growth</i>	-0.011 (-1.09)	-0.024*** (-4.16)	-0.008 (-0.81)	0.020 (1.44)	-0.039*** (-3.71)	-0.003 (-0.23)	-0.001 (-0.02)	-0.027*** (-4.03)	0.006 (0.54)
<i>Size</i>	-0.014*** (-3.41)	0.007*** (2.83)	-0.015*** (-3.58)	-0.005 (-0.98)	0.018*** (3.71)	-0.004 (-0.70)	-0.008* (-1.78)	0.005* (1.68)	-0.010** (-2.12)
<i>Lev</i>	-0.056** (-2.13)	-0.015 (-1.01)	-0.055** (-2.07)	-0.040 (-1.32)	-0.100*** (-3.60)	-0.071** (-2.31)	-0.034 (-1.28)	-0.068*** (-3.80)	-0.020 (-0.73)
<i>Roa</i>	-0.083 (-0.87)	-0.065 (-1.22)	-0.075 (-0.79)	-0.206* (-1.87)	-0.459*** (-5.11)	-0.195* (-1.94)	-0.025 (-0.27)	-0.301*** (-4.81)	-0.054 (-0.52)
<i>Mb</i>	-0.026 (-1.23)	-0.064** (-5.43)	-0.019 (-0.87)	-0.067** (-2.52)	-0.129*** (-5.61)	-0.107*** (-4.13)	-0.028 (-1.25)	-0.101*** (-6.85)	-0.033 (-1.35)
<i>Top1</i>	-0.040* (-1.81)	-0.002 (-0.16)	-0.040* (-1.80)	-0.094*** (-3.16)	0.015 (0.61)	0.038 (1.40)	-0.037 (-1.58)	-0.009 (-0.54)	-0.044* (-1.73)
<i>Mnghold</i>	0.160*** (4.79)	0.004 (0.23)	0.159*** (4.78)	0.177*** (4.62)	-0.029 (-1.24)	0.048* (1.87)	0.193*** (5.93)	-0.018 (-0.81)	0.185*** (5.05)
<i>Market</i>	-0.001 (-0.28)	0.003** (2.30)	-0.001 (-0.43)	-0.007** (-2.36)	0.001 (0.69)	0.001 (0.53)	0.002 (0.91)	0.001 (0.89)	0.001 (0.37)
<i>Holder</i>	0.034 (0.46)	0.000 (0.76)	0.000 (0.41)	-0.087 (-0.96)	0.001 (1.13)	-0.002** (-2.05)	-0.049 (-0.67)	0.000 (0.30)	-0.000 (-0.54)
<i>Industry</i>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>Year</i>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>Constant</i>	1.751*** (8.80)	0.364*** (3.29)	1.709*** (8.55)	1.783*** (7.00)	-0.385 (-1.54)	1.175*** (4.30)	1.594*** (7.60)	0.319** (2.33)	1.512*** (6.78)
<i>N</i>	936	936	936	591	591	591	911	911	911
<i>F</i>	11.00	23.74	10.69	8.29	15.53	7.23	9.47	23.65	7.83
<i>Adj R²</i>	0.18	0.33	0.18	0.21	0.38	0.22	0.16	0.37	0.16
<i>Bootstrap</i>				-1.72*					

*, **, and *** signify 10%, 5%, and 1% significance levels, respectively.

the mechanism test reveals that internal control can enhance the information content of social responsibility reports by alleviating the company's agency problems. The cross-section analysis reveals that after distinguishing enterprise size, property right nature, and media attention, only in large-scale corporations, enterprises with high media attention, and non-state enterprises, ICQ can significantly enhance the information content of social responsibility reports. Furthermore, the heterogeneity test of the mechanism reveals that agency cost exerts a partial mediating effect in large-scale corporations and enterprises with intensive public attention, and a full mediating effect in non-state enterprises.

8.2. Suggestions

This study offers various suggestions as follows:

- (1) *Listed enterprises should focus on internal control.* Listed enterprises, especially SOEs and small- and medium-sized enterprises, must focus on constructing internal control and use the internal control system to enhance the information disclosure environment. Regarding the five elements of internal control, the company must create a good internal environment, build a suitable information communication mechanism in and outside the company, upgrade the internal supervision system, efficiently play the role of internal control to augment the quality of information disclosure, and improve the information content of the social responsibility information report, especially for SOEs.
- (2) *Establish an internal and external collaborative governance mechanism.* Owing to the limitations of internal control itself, to play its crucial mechanism to alleviate the principal-agent problem, listed enterprises can introduce media attention to create an effective collaborative supervision mechanism, and jointly enhance the information content of social responsibility reports.
- (3) *The scheme of social responsibility information disclosure should be standardized.* Text information is highly flexible. Thus, regulatory authorities should issue regulations related to social responsibility information disclosure. In addition, specific requirements should be enacted regarding the content and format of the information disclosure in social responsibility reports, as well as to what extent and how much to disclose. Furthermore, to prevent false statements and major omissions, selective disclosures of good news should be avoided to optimize the information content of text.

This study has following limitations: (i) owing to data limitation, it failed to present a further detailed classification on types of social responsibility text information to elucidate the effects of internal control scheme on different types of information; (ii) although

TF-IDF decreases the weight of commonly used words in text, it cannot completely solve the problem of text semantic features; (iii) other possible mechanisms between ICQ and information content of social responsibility were not explored. Thus, our future research will focus on the limitations mentioned above and further extend the logic and mechanism of the influence of ICQ on social responsibility text information, as well as provide a new basis for investors to ascertain the information content of social responsibility reports.

Funding

National Natural Science Foundation of China [72072183]; the Humanities and Social Sciences Youth Fund project of the Ministry of Education [20YJC630063]; National Social Science Fund General Project [18BGL069].

References

- Al-Tuwaijri, S.A., Christense, T.E., Hughes, K.E., 2004. The relations among environmental disclosure, environmental performance, and economic performance: a simultaneous equations approach. *Account. Organ. Soc.* 29 (5–6), 447–471.
- Bills, K.L., Cobabe, M., Pittman, J., Stein, S.E., 2018. To share or not to share: The importance of peer firm similarity to auditor choice. Working paper.
- Bloomfield, R.J., Wilks, T.J., 2000. Disclosure effects in the laboratory: Liquidity, depth, and the cost of capital. *Account. Rev.* 75 (1), 13–41.
- Boritz, J.E., Hayes, L., Timoshenko, L., 2016. Determinants of the readability of SOX 404 reports. *J. Emerg. Technol. Account.* 13 (2), 145–168.
- Brown, S., Knechel, W., 2016. Auditor-client compatibility and audit firm selection. *J. Account. Res.* 54 (3), 725–775.
- Brown, S., Tucker, J., 2011. Large-sample evidence on firms' year-over-year MD&A modifications. *J. Account. Res.* 49 (2), 309–346.
- Bushee, B.J., Noe, C.F., 2000. Corporate disclosure practices, institutional investors, and stock return volatility. *J. Account. Res.* 38, 171–202.
- Bushman, R.M., Smith, A.J., 2001. Financial accounting information and corporate governance. *J. Account. Econ.* 31 (1–3), 237–333.
- Campopiano, G., De Massis, A., 2015. Corporate social responsibility reporting: A content analysis in family and non-family firms. *J. Bus. Ethics* 129 (3), 511–534.
- Chen, L., 2016. Managerial power, internal control and environment information disclosure. *Ecol. Econ.* 32 (8), 90–93 in Chinese.
- De Franco, G., Fogel-Yaari, H., Li, H., 2016. MD&A textual similarity and auditors. Working paper.
- Dhaliwal, D.S., Radhakrishnan, S., Tsang, A., Yang, G., 2012. Nonfinancial disclosure and analyst forecast accuracy: International evidence on corporate social responsibility disclosure. *Account. Rev.* 87 (3), 723–759.
- Dong, W., Chen, J., Chen, H., 2017. Does internal control quality affect analyst behaviors? Empirical evidence from China's stock market. *J. Financ. Res.* 12, 191–206 in Chinese.
- Duan, Z., He, Y., Zhong, Y., 2017. Is CSR information disclosure objective or not: An empirical research of Chinese listed enterprises based on text mining CSR information. *Nankai Bus. Rev.* 4, 62–72 in Chinese.
- Dyck, A., Volchkova, N., Zingales, L., 2008. The corporate governance role of the media: evidence from Russia. *J. Finance* 63 (3), 1093–1135.
- Ertugrul, M., Lei, J., Qiu, J., Wan, C., 2017. Annual report readability, tone ambiguity, and the cost of borrowing. *J. Finan. Quant. Anal.* 52 (2), 811–836.
- Faccio, M., Marchica, M.T., Mura, R., 2011. Large shareholder diversification and corporate risk-taking. *Rev. Financ. Stud.* 24 (11), 3601–3641.
- Fang, L., Peress, J., 2009. Media coverage and the cross-section of stock return. *J. Finance* 64 (5), 2023–2052.
- Hanley, K.W., Hoberg, G., 2010. The information content of IPO prospectuses. *Rev. Financ. Stud.* 23 (7), 2821–2864.
- Hao, X., Su, Z., 2014. Can Primary risk disclosure mitigate IPO underpricing? Evidence from text analysis. *J. Financ. Econ.* 5, 42–53 in Chinese.
- Hoberg, G., Phillips, G., 2010. Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Rev. Financ. Stud.* 23 (10), 3773–3811.
- Hoberg, G., Phillips, G., 2016. Text-based network industries and endogenous product differentiation. *J. Polit. Econ.* 124 (5), 1423–1465.
- Huang, P., Li, S., 2020. Social responsibility reports tone and stock price crash risks. *Audit Econ. Res.* 1, 69–78 in Chinese.
- Iliev, P., 2010. The effect of SOX section 404: costs, earnings quality, and stock prices. *J. Finance* 65 (3), 1163–1196.
- Jensen, M.C., Meckling, W.H., 1976. Theory of the firm: managerial behavior, agency costs and ownership structure. *J. Financ. Econ.* 3 (4), 305–360.
- Ji, L., Zhang, L., Tian, J., 2016. An analysis on the readability of the CSR information disclosed by the Chinese listed enterprises: A perspective from managerial power and restraint mechanism. *Account. Econ. Res.* 1, 23–35 in Chinese.
- Johnston, J., Zhang, J., 2018. Auditor style and financial reporting similarity. Working Paper.
- Kim, C., Wang, K., Zhang, L., 2019. Readability of 10-K reports and stock price crash risk. *Contemp. Account. Res.* 36 (2), 1184–1216.
- Lang, M.H., Stice-Lawrence, L., 2015. Textual analysis and international financial reporting: Large sample evidence. *J. Account. Econ.* 60 (9), 110–135.
- Lee, J.E., 2018. Internal control deficiencies and audit pricing: evidence from initial public offerings. *Account. Financ.* 58 (4), 1201–1229.
- Li, X., 2013. Perplexity and thinking of internal control research. *Account. Res.* 6, 54–61 in Chinese.
- Li, Z., 2014. Internal control and environment information disclosure: Empirical evidence from China's listed enterprises in the manufacturing industry. *China Population Resour. Environ.* 24 (6), 77–83 in Chinese.
- Li, Z., You, W., 2013. The basic effect and spillover effect of internal control: A study from the perspective of connotation mining. *Modern Manag. Sci.* 5, 48–50 in Chinese.
- Li, Z., Zhang, T., 2017. Internal control, property right nature and social responsibility information disclosure: Empirical evidence from China's listed enterprises. *Account. Res.* 10, 86–92 in Chinese.
- Lo, K., Ramos, F., Rogo, R., 2017. Earnings management and annual report readability. *J. Account. Econ.* 63 (1), 1–25.
- Luo, W., Zhu, C., 2010. Agency cost and corporate voluntary disclosure. *Econ. Res. J.* 10, 145–157 in Chinese.
- Meng, Q., Yang, J., Lu, B., 2017. The informative content of management discussion and analysis and stock price crash risks: a study based on text vectorization method management. *China Ind. Econ.* 12, 134–152 in Chinese.
- Muslu, V., Muthu, S., Radhakrishnan, S., Tsang, A., 2019. Corporate social responsibility report narratives and analyst forecast accuracy. *J. Bus. Ethics* 154 (4), 1119–1142.
- Price, M.K., Doran, J.S., Peterson, D.R., 2012. Earnings conference calls and stock returns: The incremental informativeness of textual tone. *J. Bank. Financ.* 36 (4), 992–1011.
- Qiao, Y., You, X., 2015. Empirical Study on the relationship between the effectiveness of internal control and the quality of environmental information disclosure internal control. *Stat. Decision* 23, 166–169 in Chinese.
- Shen, H., 2007. Corporate characteristics and social responsibility information disclosure: Empirical evidence from China's listed enterprises. *Account. Res.* 3, 9–16 in Chinese.
- Sun, Y., 2012. The Effect of corporate responsibility report's disclosure clarity and attestation on investment decisions: An experimental research. *Audit Res.* 4, 97–104 in Chinese.
- Tan, H.T., Wang, E.Y., Zhou, B.O., 2014. When the use of positive language backfires: The joint effect of tone, readability, and investor sophistication on earnings judgments. *J. Account. Res.* 52 (1), 273–302.
- Verrecchia, R.E., 2001. Essays on disclosure. *J. Account. Econ.* 32, 97–180.
- Wang, X., Gao, X., He, J., 2018. Risk information disclosure and audit fees in annual report: based on the angle of cosine similarity of text. *Audit Res.* 5, 100–106 in Chinese.
- Wang, Y., Li, M., Tang, M., 2015a. Property right nature, internal control and accounting information quality: empirical evidence from main board listed enterprises in China. *Financ. Econ.* 4, 97–106 in Chinese.

- Wang, Z., Zhang, J., Zheng, Q., 2015b. Quality of internal control, property right and corporate inefficiency investment: Evidence from Chinese listed enterprises. *Manag. Rev.* 9, 95–107 in Chinese.
- Wen, Z., Ye, B., 2014. Analyses of mediating effects: the development of methods and models. *Adv. Psychol. Sci.* 5, 731–745 in Chinese.
- Yang, D., Chen, H., 2013. Internal control, institutional environments and stock liquidity. *Econ. Res. J.* 1, 132–143 in Chinese.
- Yang, D., Wang, J., Chen, L., 2019. “Overcorrect” or “reasonable control”? - the role of internal control in enterprise innovation. *Bus. Manag. J.* 8, 113–129 in Chinese.
- Yin, Z., Sheng, Y., Chen, Q., 2019. Can analysts reduce stock price synchronicity – An empirical study based on textual analysis of research reports. *China Ind. Econ.* 1, 156–173 in Chinese.
- Yu, Z., Tian, G., Qi, B., Zhang, H., 2011. Corporate governance mechanism of media attention: An investigation from the perspective of earnings management. *Manag. World* 9, 127–140 in Chinese.
- Zhang, J., Cai, Y., Ni, G., 2019. Effects of disclosure tone of corporate social responsibility and disclosure integrity of financial information on investors' perceived fulfillment of corporate social responsibility: An experimental research. *Nankai Business Rev.* 1, 206–212 in Chinese.
- Zhang, X., Wang, J., Xue, Y., 2016. Application of semantic analysis in studies of enterprise environmental information disclosure. *Account. Res.* 1, 87–94 in Chinese.