

Decoding Corporate Disclosure: Evaluating Information Richness Versus Redundancy

Xutong Chen^a, Ling Liu^{*a}

*^aSchool of Management Science and Engineering
Southwestern University of Finance and Economics, China*

Abstract

The complexity of financial disclosures plays a crucial role in financial markets. Traditional metrics for measuring information complexity, such as file size, have focused primarily on the length of the information, implying lengthy disclosures due to managers' intention to conceal unfavorable information. However, research indicates that adapting to economic advancements or the inherent complexity of corporate operations also results in excessively lengthy disclosures, making it challenging to keep information disclosures succinct. To address this gap, this paper introduces the compression ratio as an innovative metric for evaluating disclosure complexity. It effectively filters out the influence of information richness to concentrate on the impact of information redundancy. The results further reveal that complex annual reports are associated with weaker earnings persistence in profitable companies, whereas unprofitable companies exhibit stronger persistence.

Keywords: Information Disclosure Complexity; Compression Ratio; Earnings Persistence

1. INTRODUCTION

Corporate information disclosure has long been a focal point of research in the finance and accounting area. Effective information disclosure mitigates information asymmetry between market participants, thereby enhancing the efficiency of resource allocation and facilitating the overall effectiveness of capital markets (Healy & Palepu, 2001). A crucial aspect of information disclosure lies in its complexity. When information disclosure attains a high level of clarity and simplicity, it becomes a powerful tool for mitigating information asymmetry (Loughran & McDonald, 2014; Guay et al., 2016; De Franco et al., 2014). Furthermore, the intricacy of financial disclosures has been shown to markedly affect investors' understanding of the information presented. This, in turn, can lead to diverse market phenomena. Higher complexity and ambiguity in textual information escalate investors' costs in processing company-specific information, diminishing the efficiency of information dissemination and resource allocation and consequently reducing stock market effectiveness (You & Zhang, 2009; Lee, 2012; Li, 2008; Lo et al., 2017). Conversely, when textual information surpasses the comprehension levels of the majority of its target audience, communication barriers arise. Notably, research indicates that more complex information disclosure is often associated with reduced earnings information quality, highlighting the impact of disclosure complexity on market perceptions (Li, 2008; Lee, 2012; Li & Zhang, 2015; Lo et al., 2017; Lehavy et al., 2011).

While the importance of information complexity has been widely acknowledged, controversy surrounds the measurement methods for information disclosure complexity. Current measurements typically focus on the "length" of the information, such as the number of completed words or file size. This focus is understandable, as higher complexity is often attributed to corporate managers aiming to obscure negative information through intricate and lengthy disclosures (Li, 2008; Loughran & McDonald, 2014). Furthermore, studies also suggest that managers tend to disclose repetitive and irrelevant content to hide information, thus enhancing the company's valuation (Bloomfield, 2002). Repetitive data then increase the length of the disclosure while simultaneously

*Corresponding author. Tel: +86-28-87081931
E-mail: lingliu@swufe.edu.cn

submerging genuinely significant information beneath a sea of meaningless words, exacerbating information asymmetry (Du & Yu, 2021; Wendai et al., 2022).

However, some studies argue that although redundant information may contribute to the length of an annual report, many other factors may also account for its extended length. For example, the advancement of globalization and international trade has led to the growth of corporations, which in turn has resulted in longer annual reports (Blankespoor et al., 2019). Simultaneously, to align with economic advancements and regulatory imperatives, accounting standards and disclosure criteria have expanded, amplifying corporate information disclosure content (Wendai et al., 2022). This kind of rich information can comprehensively reflect the true state of a company, aiding investors in better understanding the company's operations and prospects (Basu & Palazzo, 2008). Thus, it is necessary to establish a measurement that can distinguish between rich and redundant information.

To avoid the conflation of rich and redundant information, we propose a new measurement of the complexity of information disclosure—the compression ratio (compressed file size/original file size). Compression algorithms are designed to reduce data size by eliminating redundancy, unnecessary information, or patterns. During the compression process, redundant information is eliminated (Wade, 1994). This process helps to clarify the distinction between rich and redundant information. Specifically, the compression ratio is smaller with higher repetition, indicating a stronger presence of redundant information in the information disclosure and a larger compression ratio signifies fewer repetitive and redundant details, indicating a stronger presence of relevant information (Kolmogorov, 1963; Lelewer & Hirschberg, 1987; Poynton, 2012). Thus, the compression ratio effectively filters out the influence of information richness, concentrating on the impact of information redundancy.

The proposed measurement is further validated by examining the impact of information disclosure complexity on a company's future earnings. Existing research shows a close correlation between company performance and information disclosure quality (Lang & Lundholm, 1996; Francis et al., 2005). The motivation behind corporate managers crafting complex disclosures is to mask possible adverse future company performance (Li, 2008). When the current performance is unfavorable, increasing the complexity of information disclosure does not improve the company's actual performance but can be used to mislead investors, temporarily concealing poor future performance. In contrast, complex disclosure can be used to confuse investors' judgments, thereby concealing a company's future poor performance (Li, 2008; Xu et al., 2019).

Furthermore, this paper also examines the moderating effect of information disclosure complexity on corporate earnings persistence. After controlling for file size, the impact of the compression ratio on future corporate performance remains statistically and economically significant, emphasizing its relevance. Our research findings reveal that companies with more intricate disclosures exhibit lower earnings persistence when they are profitable, whereas companies' disclosures with higher complexity demonstrate higher earnings persistence in unprofitable times.

This paper has two main contributions. Firstly, a new measure of information disclosure complexity—the compression ratio is introduced. Employing the compression ratio as a fresh indicator for studying information disclosure complexity opens ways for an in-depth exploration of the complexity of disclosure content. The compression ratio enables a detailed analysis of redundant and repetitive information within firms' disclosures, separating it from valuable content and, thus, shedding light on the core factors contributing to information disclosure complexity and offering a deeper understanding of the true quality of information contained in disclosures. Secondly, this paper examines the influence of information disclosure complexity on a company's future performance, revealing its moderating effect on earnings persistence. These findings provide a new perspective for understanding the underlying reasons for the complexity of corporate information disclosure, offering valuable insights for investors, scholars, and regulatory bodies alike.

In addition to these contributions, the compression ratio elucidates the composition of complex information disclosures from dual perspectives, highlighting that their complexity predominantly stems from the inclusion of redundant and repetitive information. This finding suggests a managerial tendency towards the obfuscation of information through repetitive content. By enhancing our grasp of information complexity, this approach simultaneously deepens our understanding of the relationship between corporate information disclosure quality and financial performance. By demonstrating the link between information disclosure quality and earnings persistence, we significantly extend the scope of the literature in the field.

The remainder of this paper is arranged as follows. Section 2 discusses the relevant literature. Section 3 explores the empirical measures of annual complexity. Section 4 presents the empirical findings of the impact of compression ratio on corporate performance and Section 5 concludes the study.

2. LITERATURE REVIEW

The complexity of information disclosure is a crucial metric for evaluating the information disclosure quality of listed companies (Plumlee, 2003). It has a significant impact on the information environment (Bloomfield, 2002). A higher level of complexity triggers adverse market reactions (Miller, 2010; Loughran & McDonald, 2014; Lawrence, 2013; Kim & Verrecchia, 1991). For analysts, when companies provide effective and more information, more analysts pay attention, leading to more accurate earnings forecasts and reduced forecast dispersion and revision volatility (Lang & Lundholm, 1996). Furthermore, greater disclosure complexity may cause investors unable to comprehend the information promptly and accurately, leading to insufficient immediate reactions and sustained responses over an extended period (You & Zhang, 2009; Lee, 2012; Loughran & McDonald, 2014).

2.1 Former measurement of disclosure complexity

Initially, various readability formulas were primarily used to measure textual complexity (Flesch, 1948; Farr et al., 1951; Anderson, 1983). The Fog Index is the most commonly utilized one and is determined by analyzing both sentence length (measured in words per sentence) and the proportion of complex words (evaluated by syllable count). A higher Fog Index value indicates poorer readability (Biddle et al., 2009; Miller, 2010; Lawrence, 2013; Bushee et al., 2018); however, the Fog Index limits English text. To bridge language differences, cross-lingual measures have been proposed. For instance, Li (2008) demonstrated that the natural logarithm of word count in annual reports yields results similar to those of the Fog Index. Similarly, Loughran & McDonald (2014) introduced a complexity metric based on file size, measured in bytes.

Current metrics for assessing text complexity, such as analyzing word characters or stroke counts, are often language-specific and may not accurately reflect the true complexity of the text. For instance, although certain accounting terminologies might be challenging for novice investors and retail traders, they are unlikely to increase comprehension costs for seasoned investors and analysts familiar with the content (Loughran & McDonald, 2014). Therefore, specialized vocabulary may not be an appropriate measure of text complexity. Additionally, cross-lingual readability indicators such as text length also have limitations. A longer text might simply indicate that a company has disclosed more information in its annual report. Research shows that an increase in text length can be attributed to various motivations and does not necessarily signify poor readability or high complexity (Eng & Mak, 2003; Lo et al., 2017; Dhaliwal et al., 2012; Wang et al., 2024). This finding suggests that using text length as a sole measure of complexity might be limited in its effectiveness.

2.2 Complexity from richness and redundancy views

Both language-specific and cross-language metrics for complexity are often measured from the perspective of judging information length. However, no consensus exists in academia on whether longer information disclosure necessarily implies greater complexity. Current research linking information disclosure length with complexity often stems from the perspective of managerial motivations. The management obfuscation hypothesis suggests that when a company performs poorly, managers have the incentive to obscure information (Bloomfield, 2002). This argument is rooted in the "Incomplete Revelation Hypothesis", which posits that information incurring high processing costs might not be fully reflected in market prices (Grossman & Stiglitz, 1980; Bloomfield, 2002). Managers may seek to strategically conceal unfavorable information by deliberately providing lengthy and less transparent disclosures, thus intentionally conveying more complex information to confuse investors' perceptions (Clatworthy & Jones, 2001; Rutherford, 2005; Long & Jia, 2023).

Some studies argue that deliberately creating redundant information is one of the most common methods used to confuse investors' understanding of entire information content (Loughran & McDonald, 2014). First, managers might repeatedly disclose redundant introductory and explanatory information, such as unnecessary explanations of common sense concepts, commonly used legal terms, and financial jargon, to ensure compliance with mandatory information disclosure reporting requirements (Navissi & Sridharan, 2017). Additionally, annual reports are found to have homogenization and standardization problems. Shaw & Pecorari (2013) examined 251 annual reports published by 36 UK-listed companies from 2000 to 2012. They found a high degree of repetition of vocabulary and phrases in the reports. The chapters in the annual reports were also highly similar, indicating excessive standardization and fragmented repetition. Finally, many redundant generic pieces of information might

be repeatedly disclosed by managers due to motives such as protecting trade secrets, maintaining competitiveness, or avoiding litigation. This repetition serves the purpose of reducing disclosure content related to the company's actual operations and information closely related to its future development trends (Abraham & Shrikes, 2014). However, this perspective only proves that managers are motivated to use longer texts and even more words to increase investors' processing information costs, which does not necessarily imply that any lengthy annual report is complex.

The increase in the length of annual reports can be attributed to various motivations of managers, and a lengthy report does not indicate excessive complexity (Eng & Mak, 2003; Lo et al., 2017; Dhaliwal et al., 2012). An expansion in the length of information disclosure, including annual reports, could stem from factors such as the presence of multiple business segments within the company that require detailed explanations, managers voluntarily adding valuable disclosure content, or providing more detailed explanations for certain aspects of the report (Bloomfield, 2008; Eng & Mak, 2003). These practices do not increase the complexity of annual reports but rather reduce investors' information processing costs, effectively improving information disclosure efficiency (Xu et al., 2019).

The complexity of corporate annual reports, in addition to being determined by the length of the document, may well be influenced by the inclusion of richer information or the use of repetitive redundant information. In other words, the complexity of an annual report is not solely due to its length. Specifically, when discussing the complexity of corporate annual reports, in addition to the quantity of information, considering information quality is essential. Overemphasizing complexity based only on the length of the annual report blends rich and redundant information. This approach is inadequate for directly measuring the complexity of annual reports because it fails to distinguish between factors contributing to high complexity under managerial manipulation. Therefore, more in-depth research and discussion are needed. In response to these limitations, a text compression method for annual reports is proposed. This method leverages the unique properties of compression ratios, which effectively condense redundant information, to distinguish between disclosures that are unnecessarily complex and misleading and those that are genuinely rich in information.

2.3 Disclosure complexity and earnings persistence

Managerial opportunism theory suggests that when earnings persistence declines, managers are more likely to intentionally heighten the complexity of annual reports (Li, 2008). Earnings persistence, which reflects the consistency of earnings over time (Nelson & Skinner, 2013), is commonly used as an indicator of earnings quality (Dechow et al., 2010). When current earnings are not sustainable or losses are consistent, managers might deliberately produce more intricate annual reports. Doing so increases investors' information processing costs, thus obscuring the fleeting nature of positive outcomes or the enduring nature of negative results.

Consequently, the complexity of annual reports plays a moderating role in the relationship with earnings persistence is plausible (refer to Fig 1). In essence, when companies report poor performance or unfavorable news for the current period, managers are motivated to amplify annual report complexity. This action is aimed at clouding investors' expectations about future earnings. Therefore, the complexity of annual reports not only affects earnings persistence but also varies depending on the changing conditions of earnings.



Fig. 1. The moderating relationship of annual report complexity on earnings persistence

Although there is a wealth of research on earnings persistence, many studies do not distinguish between profitable and unprofitable companies in their analysis. This oversight is notable because, in practice, companies in these differing states of profitability demonstrate distinct patterns in the creation of information disclosure complexity (Bloomfield, 2002, 2008; Li, 2008). The behavior and strategies regarding information disclosure can significantly vary depending on whether a company is experiencing profitability or loss, influencing the complexity and interpretation of its financial reports.

For profitable companies, during periods of short-term profitability, unfavorable news might accompany their financial reports. When unfavorable news arises, company managers might use more extensive explanations in their reports (Bloomfield, 2008). Additionally, managers could increase investors' information processing costs by crafting overly complex annual reports, thereby mitigating or delaying the market's reaction to adverse news (Bloomfield, 2002). More complex disclosures can lead to higher current earnings and lower future earnings, signifying a lack of earnings persistence, as negative news does not quickly affect stock prices (Li, 2008). In contrast, the complexity of information disclosure has minimal (or even opposite) effects on companies with high earnings persistence. These companies lack negative news to conceal (Bloomfield, 2008), and firms with sustained good performance may prefer transparent disclosures to reduce information processing costs (Li, 2008). Therefore, when a company discloses unsustainable profits, the motivation is to make annual reports more complex to reduce investors' expectations of future profit declines.

For unprofitable companies, encountering losses triggers negative responses, adversely impacting earnings and stock prices. These adverse effects, in turn, reduce management's interests. Managers of such companies may engage in extensive explanations and discussions to meet investor demands (Bloomfield, 2008). They also have incentives to use tactics, such as earnings and expectation management, to disclose more complex information with the aim of avoiding investors' negative reactions and providing optimistic profit forecasts (Ahmadi et al., 2020). In essence, managers guide investors' beliefs and expectations in favorable directions. Therefore, opportunistic managers of consistently unprofitable companies may employ complex information disclosure in an attempt to obfuscate the market's true reactions to the company's sustained negative earnings.

This research delves into the intricacies of corporate annual report complexity by examining the relationship between annual report compression ratios and corporate earnings persistence. Through this exploration, we reveal the underlying motivations driving the complexity of these reports, thereby uncovering the strategic intentions behind information disclosure and the behaviors of corporate management. This approach offers valuable insights into how and why companies craft their financial disclosures, providing a deeper understanding of the tactics used in corporate communication.

3. DATA, COMPLEXITY MEASUREMENT, AND METHODOLOGY

3.1 Data collection

The sample was selected from 3705 Chinese A-share listed companies on the Shanghai and Shenzhen Stock Exchanges from 2009 to 2019. The sample selection involved the following steps: (1) exclusion of annual report data from companies designated as ST/*ST due to their special characteristics and (2) removal of samples with missing data. Following these criteria, the final sample consisted of 3,414 companies, forming an annual balanced panel dataset. To minimize the influence of outliers on the regression results, the continuous variables were winsorized at the 1% levels of the sample distribution. The annual report data of all listed companies were sourced from the China Securities Regulatory Commission's designated website for listed company information disclosure—JuChao InfoNet. Other corporate features and financial indicators were obtained from the CSMAR database.

3.2 Compression ratio as complexity measurement

Corporate information disclosure serves as the primary means to reduce information asymmetry and minimize agency costs; however, existing complexity metrics conflate rich and redundant information, making it unsuitable for directly assessing the quality of information disclosure. Therefore, a nuanced distinction between those aspects is essential and warrants further investigation. Based on information theory, data compression involves encoding information, using fewer bits than the original representation, to reduce the file size (Wade, 1994). The process of reducing the data file size is commonly known as data compression, where redundant and repetitive parts are compressed, unnecessary or less important information (such as spaces and tabs) is removed, and useless metadata (such as file headers and tags) are deleted (Pujar & Kadlaskar, 2010). Compression methods mainly include lossy and lossless compression (Mahdi et al., 2012). For text files, lossless compression methods are typically employed.

Lossless compression does not cause any information to be lost, allowing the compressed file to be restored to the original file through decompression (Pujar & Kadlaskar, 2010). During this process, the compressed file retains the same information as the original file but uses fewer characters for transmission, reducing data redundancy.

According to information theory, high redundancy in a document indicates significant repeated content, which in turn suggests relatively low actual information content. Therefore, proposing that the compression ratio, defined as the ratio of the compressed file size to the original file size (EQ:1), can be a viable indicator of the extent of redundant information in annual reports is reasonable. This ratio can also reflect the amount of meaningful information content. A lower compression ratio, indicative of less of a reduction in file size, may suggest greater redundancy, whereas a higher ratio could imply more substantial, meaningful content in the report. This dual functionality makes the compression ratio a potentially effective tool for evaluating the quality and utility of the information presented in corporate annual reports

$$CompressionRatio = \frac{CompressedFileSize}{OriginalFileSize} \quad (1)$$

where *OriginalFileSize* refers to the original size of the annual report PDF file, while *CompressedFileSize* indicates the size of the annual report after compression.

The Deflate algorithm (Deutsch, 1996), a well-established and traditional compression method, is based on the principles of LZ77 and Huffman coding (Oswal et al., 2016). It is extensively used for compressing PDF files of annual reports. Deflate is frequently associated with zip-compressed files and is known for its efficiency and effectiveness in data compression. Files compressed using the Deflate algorithm typically have the “.zip” file extension, indicating the method of compression used. This algorithm’s widespread adoption and reliability make it a suitable choice for compressing and analyzing the content of corporate annual reports.

3.3 Assessment of compression ratio

The length of the information does not necessarily indicate greater complexity, as it could result from increased content demands. To demonstrate the difference between traditional complexity measurement, file size, and our proposed measurement, compression ratio, we compare the trends of these two measurements across two vastly different industry sectors, capital market services and automobile manufacturing (Fig 2). The capital market services industry focuses on financial services facilitating capital flow, which is required to provide detailed explanations of their business models, financial products, and services, as well as potential market risks in their disclosure. The automobile manufacturing industry is centered on the design, production, and sale of vehicles; thus, factors such as environmental impact, safety, and compliance may play a significant role in the information disclosed.

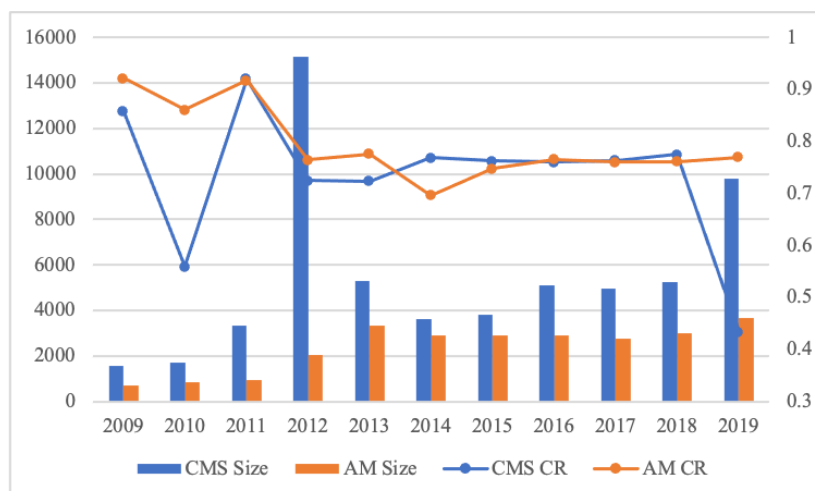


Fig. 2. This figure shows the comparison of the trends in the annual report PDF file sizes (in kilobytes, KB) and compression ratios (calculated using the Deflate algorithm) between the vastly different sectors of capital market services (CMS) and automobile manufacturing (AM).

As shown in Fig 2 that from 2009 to 2019, the sizes of the annual reports of capital market services significantly exceeded those of automobile manufacturing, particularly in 2012, where the gap reached 13091 KB (annual

reports for capital market services were 15126 KB, whereas those for automobile manufacturing were 2035 KB). However, the differences in the compression ratios between these two sectors were not substantial. In 2011, 2016, and 2017, for example, the annual report file sizes of capital market services surpassed those of automobile manufacturing; however, their compression ratios were nearly identical. Remarkably, in 2015 and 2016, the compression ratios of capital market services' annual reports were lower than those of automobile manufacturing. These observations highlight the point that relying solely on the textual length of annual reports is insufficient for accurately gauging their complexity. The compression ratio metric previously proposed mitigates this issue effectively.

3.4 Disclosure complexity in profitable and unprofitable firms

Managers of both profitable and unprofitable companies employ different strategies in using complexity in their disclosures. In profitable firms, managers often increase complexity to obscure unfavorable news, raising investors' information processing costs and delaying market reactions, which reduces earnings persistence (Bloomfield, 2008; Li, 2008). For unprofitable companies, complexity is frequently used to mask ongoing losses and present optimistic forecasts, influencing investor expectations and postponing negative market responses (Ahmadi et al., 2020).

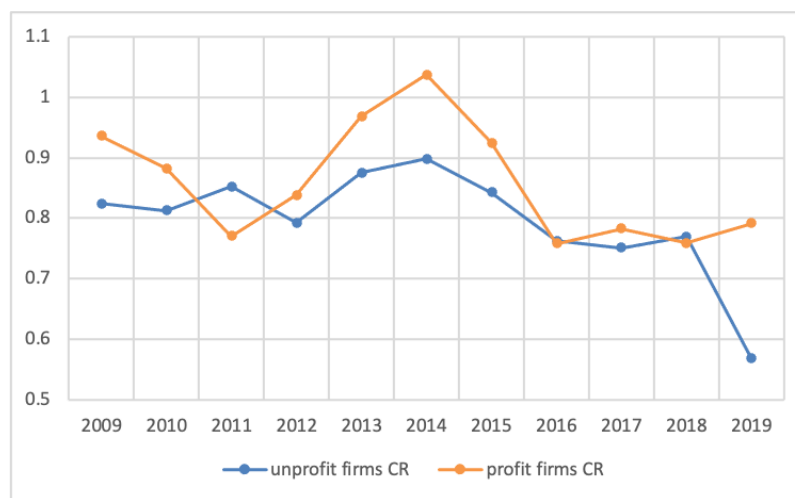


Fig. 3. This figure shows the comparison of the trends in the compression ratios (calculated using the Deflate algorithm) of unprofitable and profitable firms over time from 2009 to 2019.

Fig 3 reveals the trends in the compression ratios of annual reports for profitable and unprofitable firms between 2009 and 2019. For most of the period, profitable companies consistently exhibited higher compression ratios, particularly between 2009 and 2014. The compression ratio for profitable firms peaked in 2014. However, after 2016, their compression ratios steadily declined, reaching a lower level by 2019. In contrast, the compression ratios for unprofitable companies remained relatively stable initially, but after 2017, they experienced a sharp decline, dropping significantly to around 0.6 by 2019. We observe notable differences in the compression ratios between profitable and unprofitable firms in fig 3. An independent sample t-test conducted on the compression ratios of profitable and unprofitable companies revealed a statistically significant difference ($t\text{-value} = -229.793$), indicating clear disparities in the complexity of their disclosures. This statistically significant difference indicates distinct levels of complexity in the reports of profitable versus unprofitable companies, as reflected in their compression ratios.

3.5 Earnings persistence and information disclosure complexity

Research indicates that as annual reports grow in complexity, positive news tends to fade quickly, whereas negative news may linger for longer (Kothari et al., 2009). Managers may tactically use the intricacy of annual reports to conceal or postpone possible negative future developments (Li, 2008). Therefore, it can be supposed that the more complex the annual report is, the poorer the earnings persistence for profitable companies, whereas unprofitable companies exhibit better earnings persistence. To investigate this phenomenon and delve into the causes of heightened complexity in annual reports, a methodology similar to that of Li (2008) is utilized, which involves applying a first-order autoregressive model to assess the persistence of company earnings.

$$Earnings_{t+1} = \beta_0 + \beta_1 Earnings_t + \beta_2 CompressionRatio + \beta_3 Earnings_t \times CompressionRatio + Controls + \varepsilon \quad (2)$$

where $Earnings_t$ represents the company's earnings for period T , and $Earnings_{t+1}$ signifies the company's earnings for period $T + 1$. The earnings calculation follows the method outlined by Li (2008), involving operating profit divided by total assets. $CompressionRatio$ stands for the annual report complexity measure, denoting the compression ratio (compressed file size / original file size). $Earnings_t \times CompressionRatio$ represents the interaction term between the annual report compression ratio and the company's current period earnings, elucidating how changes in annual report complexity (compression ratio) affect the persistence of the company's earnings, holding other conditions constant.

The control variables closely adhere to previous research carried out by Li (2008) and Xu et al. (2019). The analysis begins by considering firm size (Scale) and growth prospects (Growth). Firm size is indicative of a company's operational scope and business environment, with larger firms typically producing more comprehensive and complex annual reports. Similarly, companies with growth-oriented strategies, which often entail intricate and ambiguous business models, tend to have more elaborate reports.

Operational (O_cplx) and financial (F_cplx) complexity are also included as control variables. Companies with complex business operations and financial structures are more likely to generate detailed annual reports. Additionally, firms operating in unstable business environments may communicate more intricately with investors, thereby increasing report complexity. This necessitates the inclusion of earnings volatility (Sd_e) and stock return volatility (Ret_vol) as controls.

We recognize the impact of significant corporate events on reporting and incorporate seasoned equity offerings (SEOs) and mergers and acquisitions (MAs) as control variables. Furthermore, financial leverage (Leverage) and years since initial listing (Age) are considered. Firms with greater debt financing capabilities and a longer listing history typically exhibit reduced information asymmetry and uncertainty.

Finally, the model includes the natural logarithm of the annual report file size (LnSize), calculated from the original PDF file size, to address the potential concurrent influence of this size on both the compression ratio and earnings persistence. Table 1 provides detailed definitions and methodologies for these control variables.

Table 1. Data definitions summary of key variables in the study of annual report complexity and earning persistence.

Variables	Definition
Measurement of complexity	
<i>Com_deflate</i>	Compression ratio calculated after applying the Deflate algorithm to file.
<i>Com_lzma</i>	Compression ratio calculated after applying the LZMA algorithm to file.
<i>Com_bzip2</i>	Compression ratio calculated after applying the Bzip2 algorithm to file.
Dependent variable	
<i>Earnings</i>	Operating profit divided by total assets.
Control variables	
<i>Scale</i>	Natural logarithm of the stock market value at the end of the previous period.
<i>Growth</i>	Growth rate of sales revenue.
<i>Leverage</i>	Total liabilities of the company divided by total assets at the end of the previous period.
<i>O_cplex</i>	Operational complexity: Natural logarithm of the number of company segments.
<i>F_cplex</i>	Financial complexity: Natural logarithm of the number of accounting entries involved in the company in the balance sheet, income statement, and cash flow statement.
<i>Se_d</i>	Earnings volatility: Standard deviation of earnings over the last five years.
<i>Ret_vol</i>	Stock return volatility: Standard deviation of monthly stock returns over the past year.
<i>Age</i>	Number of years the company has been listed.
<i>SEO</i>	"1" if the company has issued additional shares in the current accounting period, and "0" otherwise.

<i>MA</i>	"1" if the company underwent a merger and acquisition (M&A) reorganization in the current accounting period, and "0" otherwise.
<i>LnSize</i>	Logarithm of annual report PDF file size (in kilobytes, KB).

4. RESULTS AND DISCUSSION

4.1 Summary statistics

Table 2 presents the descriptive statistics, offering insights into the characteristics of annual reports of the companies examined. On average, the annual reports of these listed companies exhibit substantial file sizes, as indicated by the mean and median of the logarithmically transformed file sizes, which are 7.818 and 7.909, respectively, with a standard deviation of 0.696. According to the conventional interpretation of logarithmic scales, these figures imply a significant volume of content within the annual reports.

Table 2. Summary statistics for all variables

Variables	N	Mean	Std.Dev	25th	50th	75th
<i>Com_deflate</i>	22961	0.761	0.326	0.712	0.767	0.852
<i>Com_lzma</i>	22961	0.737	0.368	0.702	0.737	0.831
<i>Com_bzip2</i>	22961	0.760	0.370	0.716	0.765	0.853
<i>Earnings</i>	22961	0.033	0.137	0.012	0.034	0.065
<i>Scale</i>	22961	22.158	1.489	21.168	21.931	22.881
<i>Growth</i>	22961	7.062	893.870	-0.029	0.103	0.269
<i>Leverage</i>	22961	0.478	0.646	0.294	0.458	0.623
<i>O_cplex</i>	22961	2.444	1.046	1.792	2.485	3.091
<i>F_cplex</i>	22961	5.002	0.389	4.625	5.118	5.425
<i>Se_d</i>	22961	0.039	0.362	0.007	0.016	0.035
<i>Ret_vol</i>	22961	0.130	0.103	0.088	0.115	0.151
<i>Age</i>	22961	10.966	6.653	5.000	10.000	16.000
<i>LnSize</i>	22961	7.818	0.696	7.486	7.909	8.162

Regarding the compression ratios calculated using the Deflate algorithm (LZMA algorithm, Bzip2 algorithm), the average ratios are 0.761 (mean) and 0.737 (median). The corresponding standard deviations are 0.326 (mean) and 0.368 (median). Notably, minimal discrepancies exist in the compression ratio values derived from the three algorithms, suggesting a consistent pattern across methodologies. This finding highlights the consistently high compression ratios of the annual reports, which approach 1. In essence, the disclosed information in these reports is remarkably comprehensive, marked by relatively limited redundant and repetitive content. The following section delves into a detailed comparative analysis of the compression ratio data for the textual information in the annual reports and the original PDF file sizes.

4.2 Earnings persistence and complexity

This section presents an investigation of the influence of annual report complexity on earnings persistence and aims to test whether more intricate annual reports lead to reduced earnings persistence for profitable firms while enhancing earnings persistence for unprofitable firms. The sample is stratified into two subsets: profitable firms, characterized by positive current earnings (Earning > 0), and unprofitable firms, characterized by zero or negative current earnings (Earning ≤ 0).

In the case of profitable firms that report positive earnings, both industry and year effects are controlled. The findings presented in Table 3 (1) showcase the regression outcomes among next-year earnings, current earnings, the compression ratio, and their interactions. The analysis encompasses all annual data from profitable firms and employs the Deflate compression algorithm. The coefficient for the compression ratio is -0.0237 (standard deviation = 0.0021), whereas the interaction term coefficient is 0.877 (standard deviation = 0.0377). Both coefficients are statistically significant at the 1% level, confirming what is expected. The findings indicate that a decrease in the overall annual report compression ratio (indicating an increase in redundant and repetitive information) correlates with diminished earnings persistence for profitable firms. From an economic perspective, under constant conditions, a reduction in the lower quartile of the compression ratio (from 0.852 to 0.712) substantially decreases the earnings persistence of profitable firms in the subsequent year by 0.12 (calculated as $0.877 * (0.852 - 0.712)$, where 0.877 represents the interaction term coefficient). This finding underscores the significant economic implications associated with fluctuations in the compression ratio within the interquartile range, consistent with prior research findings (Li, 2008).

Similarly, as shown in Table 3 (2) for the annual sample of unprofitable (negative earnings) companies, it is evident that the compression ratio significantly negatively influences the persistence of losses. By utilizing the Deflate compression algorithm to compute the annual report compression ratio to explain negative earnings persistence, the coefficient for the compression ratio (*CompressionRatio*) is -0.00884 (with a standard deviation of 0.0524) and is not statistically significant. However, the interaction term coefficient is -1.620 (with a standard deviation of 0.2429), indicating that the interaction between the annual report compression ratio and earnings is significantly negative at the 1% level, aligning with theoretical expectations. This result implies that as the overall compression ratio of the annual report decreases (indicating more repeated and redundant information in the report), the earnings persistence of unprofitable companies improves. From an economic perspective, assuming that all other conditions remain constant, a decrease in the quartile range of the compression ratio for unprofitable companies (from 0.852 to 0.712) results in a 0.22 increase in earnings persistence for the second year (calculated using the formula $-1.620 * (0.852 - 0.712)$, where -1.620 represents the interaction term coefficient). This further underscores the substantial economic significance of the compression ratio.

Table 3. Panel regression for the annual autoregressive model of company earnings.

Earnings_{t+1} Variables	(1) Profitable firm-year	(2) Unprofitable firm-year
<i>Com_deflate</i>	-0.0237*** (0.0021)	-0.0884 (0.0524)
<i>Earnings_t × Com_deflate</i>	0.877*** (0.0377)	-1.620*** (0.2429)
<i>Earnings_t</i>	-0.317*** (0.0265)	0.00595 (0.0264)
Controls	Yes	Yes
Industry	Yes	Yes
Year	Yes	Yes
N	20886	2646
Adj.R-sq	0.2785	0.1989

The empirical results further support the existence of a negative relationship between the complexity of annual reports and earnings persistence in profitable companies. Conversely, a positive correlation is found for companies incurring losses. This means that using the compression ratio as a measure of the complexity of annual reports enables the differentiation between rich and meaningful information, indicated by a high compression ratio, and redundant or repetitive information, indicated by a low compression ratio.

The findings reveal that corporate managers often increase the complexity of annual reports by adding redundant and repetitive content, complicating the task of distinguishing short-term, positive news from long-term, negative information. Specifically, companies strategically incorporate a substantial amount of redundant information in their annual reports, such as repetitive content or reiterations of information already known to investors. This intentional inclusion of unnecessary and irrelevant information seems to be aimed at masking unfavorable news for the future and alleviating investors' concerns. Such a strategic approach to information disclosure leads to a progressive increase in the complexity of annual reports, making it more difficult for users to interpret the information provided. This, in turn, obscures the true condition of the company, leading to potentially negative economic consequences.

For investors, failing to recognize patterns of strategic obfuscation in complex disclosures can result in significant economic consequences. Misinterpreting such reports as thorough and transparent may lead to overvaluation of a company's stock, prompting investment decisions based on incomplete or misleading information. This can have long-term effects, particularly when the company's poor performance eventually becomes apparent, leading to sharp declines in stock prices and substantial financial losses. Conversely, a higher compression ratio—indicating less redundancy and greater conciseness—typically signals greater transparency, more accurate reflection of a company's financial health, and better corporate governance, enabling investors to make more informed decisions.

Compression ratios also serve as a predictive tool. A sudden decline in the compression ratio, reflecting increased redundancy, may indicate impending unfavorable news. Investors who identify this pattern can adjust their expectations and reduce their financial risk before the company's poor performance is fully reflected in earnings reports. Overall, compression ratios provide a quantitative method for distinguishing between reports that are complex due to content richness and those that are strategically complex to obscure a company's true financial state. This makes compression ratios a valuable tool for investors, analysts, and regulators to mitigate the risks of information asymmetry and avoid being misled by overly complex disclosures.

According to the management obfuscation hypothesis, managers have the incentive to obscure information when a company performs poorly (Bloomfield, 2002). Consequently, a negative correlation might exist between a company's current performance and the complexity of its annual report. To demonstrate that managers engage in strategic information disclosure to hide future performance rather than current performance, a more thorough analysis is warranted.

4.3 Current earnings and complexity

This section undertakes a detailed analysis to substantiate that managers complicate annual reports to obfuscate future performance. The relationship between company performance and the complexity of annual reports is thoroughly examined. The regression results, displayed in Table 4, highlight the influence of current earnings on the complexity of annual reports. In all regression models shown in Table 4, the determinants of annual report complexity, as outlined in Table 1, are factored into the control variables. Additionally, to accommodate potential variations across different periods and industries, fixed effects for both year and industry are incorporated into the analysis.

The positive coefficients of current earnings indicate that companies with higher earnings typically exhibit larger compression ratios in their annual reports. The results suggest that these companies' reports contain richer information (represented by larger compression ratios) and fewer redundant details (indicated by smaller compression ratios). In contrast, companies with lower earnings tend to include more redundant information in their annual reports, making the reports more complex. In column (1) of Table 4, when using the Deflate compression algorithm, the coefficient for current earnings is 0.00227 (with a standard deviation of 0.0602), which is significant at the 1% level.

Similar results are obtained when the company's earnings level is replaced by a binary dummy variable indicating whether the current earnings are greater than 0 (1 if yes, 0 otherwise) to determine the impact of dummy variables on the firm's performance, distinguishing between profit and loss levels. This approach aids in elucidating the disparity in the means of the two groups. This transformation allows for a greater focus on the positive or negative levels of earnings, highlighting the importance of a company's profitability. In column (2) of Table 4, the coefficient for the profit/loss dummy variable is 0.00891 (with a standard deviation of 0.0043) which is significant at the 1% level. These findings suggest that the annual reports of companies experiencing losses tend to contain more redundant information than those of profitable companies, thereby increasing their complexity.

Notably, in the regression results of company earnings on compression ratios, the R-squared values used to explain the relationship between the two variables are too small (0.0224 and 0.0264). Furthermore, the coefficients for current earnings are exceedingly weak (0.0227 and 0.00891). This finding indicates that the influence of current earnings on the complexity of annual reports is minimal, with very limited explanatory power. Therefore, despite the statistical significance of the impact of current earnings on the complexity of annual reports, its economic effect is quite weak. Thus, it is evident that managers who write complex annual reports may well focus more on a company's future performance.

Table 4. Panel regression for the company's current earnings on annual report complexity.

Variables	(1) Compression ratio	(2) Compression ratio
<i>Earnings_t</i>	0.00227*** (0.0602)	
<i>Profit/loss dummy</i>		0.00891*** (0.0043)
Controls	Yes	Yes
Industry	Yes	Yes
Year	Yes	Yes
N	23532	23532
Adj.R-sq	0.0224	0.0264

The interpretation of the results can be approached from two perspectives. First, annual reports encompass vast information regarding the company's current and historical performance. Hence, the inference is that managers do not solely make annual reports more complex to conceal negative aspects related to current performance. Second, if the reported positive earnings in the current annual report result from strategic manipulation, managers might deliberately complicate the report. In this scenario, the complexity of the annual report could be intended to obscure information rather than provide more authentic details.

The findings further validate the phenomenon that managers writing complex annual reports are not merely attempting to conceal their current poor performance. In contrast, they might be trying to mask or postpone potential negative future developments, showing a tendency to incorporate redundant information to enhance the complexity of the annual report.

4.4 Robustness check

To ensure the robustness of the model, two key tests were carried out in this section. First, the VIFs test and correlation tests were conducted on the relevant variables. These tests revealed no significant issues of multicollinearity between the explanatory and dependent variables. The VIF values were predominantly below 10 (for more details, see Table 5), and the correlation coefficients did not exceed 0.5, indicating acceptable levels of multicollinearity.

Table 5. VIF test for variables

Variables	VIF	1/VIF
<i>Com_deflate</i>	1.58	0.6326
<i>Lnpdf</i>	2.02	0.4948
<i>Sd_e</i>	1.05	0.9545
<i>Leverage</i>	1.76	0.5694
<i>Age</i>	1.32	0.7557
<i>SEO</i>	1.08	0.9297
<i>MA</i>	1.14	0.8743
<i>F_cplex</i>	11.81	0.0847
<i>B_cplex</i>	1.90	0.5264
<i>Growth</i>	1.01	0.9920
<i>Com_deflate</i>	2.60	0.3848

Second, in the calculation of the annual report complexity measure, specifically through the compression ratio, the Deflate algorithm was employed as a widely accepted standard in data compression due to its well-established balance between compression efficiency and computational speed. However, to enhance the robustness of the analysis and mitigate potential biases or inaccuracies associated with any single algorithm, two additional algorithms—LZMA and Bzip2—were utilized. LZMA, which is the default algorithm for the 7z compression format, was selected for its high compression ratio, adaptability in terms of dictionary size, and efficient performance in both compression and decompression processes, alongside its low memory usage during decompression, making it particularly suitable for large-scale textual data such as annual reports (Alakuijala et al., 2015). Bzip2, a lossless compression algorithm based on the Burrows-Wheeler transform, was chosen due to its ability to achieve superior data compression, typically reducing file sizes by 10%-15% through advanced techniques. Furthermore, Bzip2 provides high-quality compression while maintaining efficient decompression speeds and compatibility with widely used compression formats (Alakuijala et al., 2015).

The selection of these algorithms was based on their distinct but complementary characteristics, which ensure a comprehensive evaluation of the compression ratio as a proxy for information disclosure complexity. By employing these alternative algorithms, our study reduces the risk of algorithm-specific biases while simultaneously enhancing the generalizability of the findings. The compression ratios derived from LZMA and Bzip2 served as alternative complexity measures and were incorporated into robustness tests, further validating the empirical results.

Fig 4 illustrates that the differences in the annual report compression ratios obtained using the three algorithms are minimal. Notably, the compression ratios computed via the Deflate and Bzip2 algorithms are almost identical. Moreover, the largest discrepancy in compression ratio values, which is observed between LZMA algorithm and the other two algorithms, remains within a narrow margin of 0.05.

The results of comparing the results of the three compression algorithms (Deflate, LZMA, Bzip2) in terms of the impact on earnings persistence are presented in Table 6. Regardless of whether the company is profitable or generate losses, the outcomes yielded by the three algorithms (Deflate, LZMA, Bzip2) exhibit striking similarities. In profitable firms, the interaction terms of the compression ratios obtained from all three algorithms (Deflate, LZMA, Bzip2) showed significant positive correlations, with coefficients of 0.877 (t-statistic: 0.0377), 0.842 (t-statistic: 0.0354), and 0.904 (t-statistic: 0.0379), respectively. Similarly, for unprofitable companies, the interaction terms of the compression ratios obtained from all three algorithms were significantly negatively correlated, with coefficients of -1.620 (t-statistic: 0.2429), -1.427 (t-statistic: 0.1974), and -1.347 (t-statistic: 0.1992), respectively. These results underscore the robustness of the findings presented before.

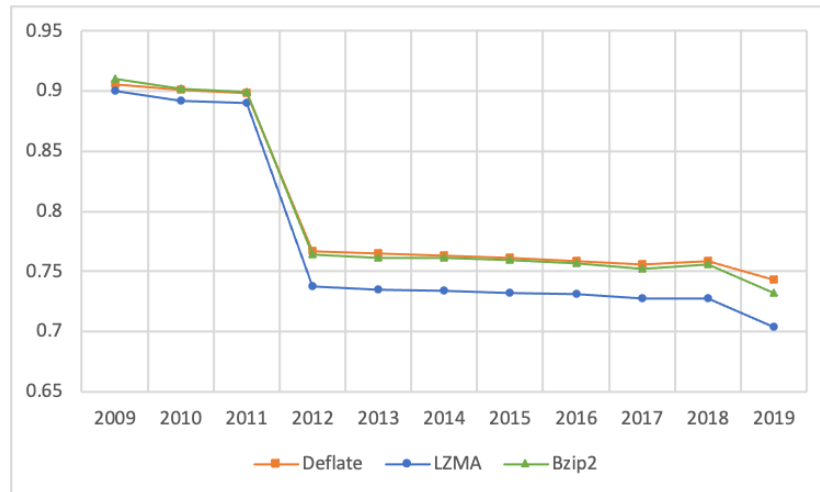


Fig. 4. The figure shows the comparison of compression ratio results obtained through three different compression algorithms (Deflate, LZMA, and Bzip2) over time from 2009 to 2019.

5. CONCLUSION

The complexity of financial information disclosures has a significant impact on the information environment (Bloomfield, 2002). Most current measurements of information complexity concentrate on the number of longer words (Biddle et al., 2009; Miller, 2010; Lawrence, 2013; Bushee et al., 2018) used in the file or the length of the entire document (Li, 2008; Loughran & McDonald, 2014). However, the increase in the length of the information disclosure can be attributed to the firm's environment (Blankespoor et al., 2019) or various motivations of managers (Bloomfield, 2002), and there is no consensus on whether longer information disclosure necessarily implies greater complexity (Eng & Mak, 2003; Lo et al., 2017; Dhaliwal et al., 2012).

Table 6. Panel regression for the annual report complexity on earnings persistence by 3 different compression ratios using unprofitable firm-year.

Earnings _{t+1} Variables	Profitable firm-year		
	(1) Deflate	(2) LZMA	(3) Bzip2
<i>Com_deflate</i>	-0.0237*** (0.0021)		
<i>Earnings_t × Com_deflate</i>	0.877*** (0.0377)		
<i>Com_lzma</i>		-0.0211*** (0.0020)	
<i>Earnings_t × Com_lzma</i>		0.842*** (0.0354)	
<i>Com_bzip2</i>			-0.0226*** (0.0021)
<i>Earnings_t × Com_bzip2</i>			0.904*** (0.0379)
<i>Earnings_t</i>	-0.317*** (0.0265)	-0.273*** (0.0243)	-0.332*** (0.0265)
Controls	Yes	Yes	Yes
Industry	Yes	Yes	Yes
Year	Yes	Yes	Yes
N	20886	20886	20886
Adj.R-sq	0.2785	0.2793	0.2795

In this study, the “compression ratio” is presented as an innovative metric designed to evaluate the complexity of corporate information disclosures. By effectively filtering out valuable information from redundant content, this measure reveals the strategic motivations underlying financial disclosure. Our research identifies that managers may obscure a firm's financial status by repeatedly incorporating redundant information in complex disclosures, which suggests a deliberate approach by managers to create information disclosures that may delay or hide negative information from investors, questioning the integrity of corporate disclosures

Additionally, we find that the dynamics of corporate information disclosure reveal a significant relationship between report complexity and corporate performance. In profitable companies, complex reports are associated with less consistent earnings but indicate more consistent earnings in unprofitable companies. This suggests that report complexity is used strategically in financial disclosures, influencing investors' perceptions of a company. Importantly, after controlling for file size, the impact of the compression ratio on future corporate performance remains statistically and economically significant, further emphasizing the relevance and utility of this ratio in assessing report complexity.

In essence, Our findings contribute to a deeper understanding of corporate disclosure strategies by highlighting the value of nuanced evaluation metrics like the compression ratio. This metric provides critical insights into the transparency and complexity of a company's financial disclosures, making it a valuable tool for investors, analysts, and regulators. Companies with low compression ratios may be using redundant content to obscure poor performance, signaling potential obfuscation. Recognizing these patterns is essential for investors to avoid overvaluation and financial losses. Thus, the compression ratio serves not only as a measure of report complexity but also as an indicator of managerial intent, helping to identify companies that may be deliberately masking negative results. These insights emphasize the need for greater transparency and accountability in corporate reporting.

This study also has several limitations. First, the complexity measurement focuses only on information richness and ignores differences in language features, grammar, and textual logic. These factors could render a firm's disclosures more obscure and intricate (Crossley, 2020). Second, annual reports represent just one channel through which managers communicate with investors. Compared to other media, such as conference calls and financial statements, communication in annual reports is relatively constrained. Hence, alternative communication channels might provide more robust models for exploring the relationship between complexity and firm performance. These limitations suggest opportunities for future research.

ACKNOWLEDGMENTS :

This research is supported by the National Natural Science Foundation of China [71401139] and Sichuan Science and Technology Program[2019YJ0251].

REFERENCES

- Abraham, S. and Shrivs, P. J. (2014). Improving the relevance of risk factor disclosure in corporate annual reports. *The British accounting review*, 46(1):91– 107.
- Ahmadi, Y., Banimahd, B., Talebnia, G., and Pourzamani, Z. (2020). The effect of earnings management on the relationship between earnings forecast error and earning persistence: Test of management overconfidence theory. *International Journal of Finance & Managerial Accounting*, 5(19):99–115.
- Alakuijala, J., Kliuchnikov, E., Szabadka, Z., and Vandevenne, L. (2015). Comparison of brotli, deflate, zopfli, lzma, lzham and bzip2 compression algorithms. *Google Inc, pages 1–6*.
- Anderson, J. (1983). Lix and rix: Variations on a little-known readability index. *Journal of Reading*, 26(6):490–496.
- Basu, K. and Palazzo, G. (2008). Corporate social responsibility: A process model of sensemaking. *Academy of management review*, 33(1):122–136.
- Biddle, G. C., Hilary, G., and Verdi, R. S. (2009). How does financial reporting quality relate to investment efficiency? *Journal of accounting and economics*, 48(2-3):112–131.
- Blankespoor, E., Dehaan, E., Wertz, J., and Zhu, C. (2019). Why do individual investors disregard accounting information? the roles of information awareness and acquisition costs. *Journal of Accounting Research*, 57(1):53–84.
- Bloomfield, R. (2008). Discussion of “annual report readability, current earnings, and earnings persistence”. *Journal of Accounting and Economics*, 45(2-3):248– 252.
- Bloomfield, R. J. (2002). The “incomplete revelation hypothesis” and financial reporting. *Accounting Horizons*, 16(3):233–243.
- Bushee, B. J., Gow, I. D., and Taylor, D. J. (2018). Linguistic complexity in firm disclosures: Obfuscation or information? *Journal of Accounting Research*, 56(1):85–121.
- Clatworthy, M. and Jones, M. J. (2001). The effect of thematic structure on the variability of annual report readability. *Accounting, Auditing & Accountability Journal*, 14(3):311–326.
- Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, 11(3):415–443.
- De Franco, G., Vasvari, F. P., Vyas, D., and Wittenberg-Moerman, R. (2014). Debt analysts' views of debt-equity conflicts of interest. *The Accounting Review*, 89(2):571–604.
- Dechow, P., Ge, W., and Schrand, C. (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of accounting and economics*, 50(2-3):344–401.
- Deutsch, P. (1996). Deflate compressed data format specification version 1.3. Technical report.
- Dhaliwal, D. S., Radhakrishnan, S., Tsang, A., and Yang, Y. G. (2012). Nonfinancial disclosure and analyst forecast accuracy: International evidence on corporate social responsibility disclosure. *The accounting review*, 87(3):723–759.
- Du, S. and Yu, K. (2021). Do corporate social responsibility reports convey value relevant information? evidence from report readability and tone. *Journal of business ethics*, 172:253–274.
- Eng, L. L. and Mak, Y. T. (2003). Corporate governance and voluntary disclosure. *Journal of accounting and public policy*, 22(4):325–345.
- Farr, J. N., Jenkins, J. J., and Paterson, D. G. (1951). Simplification of flesch reading ease formula. *Journal of applied psychology*, 35(5):333.
- Flesch, R. (1948). *A new readability yardstick journal of applied psychology* 32: 221–233.

- Francis, J., LaFond, R., Olsson, P., and Schipper, K. (2005). The market pricing of accruals quality. *Journal of accounting and economics*, 39(2):295–327.
- Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American economic review*, 70(3):393–408.
- Guay, W., Samuels, D., and Taylor, D. (2016). Guiding through the fog: Financial statement complexity and voluntary disclosure. *Journal of Accounting and Economics*, 62(2-3):234–269.
- Healy, P. M. and Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of accounting and economics*, 31(1-3):405–440.
- Kim, O. and Verrecchia, R. E. (1991). Market reaction to anticipated announcements. *Journal of Financial Economics*, 30(2):273–309.
- Kolmogorov, A. N. (1963). On tables of random numbers. *Sankhy`a: The Indian Journal of Statistics, Series A*, pages 369–376.
- Kothari, S. P., Shu, S., and Wysocki, P. D. (2009). Do managers withhold bad news? *Journal of Accounting research*, 47(1):241–276.
- Lang, M. H. and Lundholm, R. J. (1996). Corporate disclosure policy and analyst behavior. *Accounting review*, pages 467–492.
- Lawrence, A. (2013). Individual investors and financial disclosure. *Journal of accounting and economics*, 56(1):130–147.
- Lee, Y.-J. (2012). The effect of quarterly report readability on information efficiency of stock prices. *Contemporary Accounting Research*, 29(4):1137–1170.
- Lehavy, R., Li, F., and Merkley, K. (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review*, 86(3):1087–1115.
- Lelewer, D. A. and Hirschberg, D. S. (1987). Data compression. *ACM Computing Surveys (CSUR)*, 19(3):261–296.
- Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and economics*, 45(2-3):221–247.
- Li, Y. and Zhang, L. (2015). Short selling pressure, stock price behavior, and management forecast precision: Evidence from a natural experiment. *Journal of Accounting Research*, 53(1):79–117.
- Lo, K., Ramos, F., and Rogo, R. (2017). Earnings management and annual report readability. *Journal of accounting and Economics*, 63(1):1–25.
- Loughran, T. and McDonald, B. (2014). Measuring readability in financial disclosures. *the Journal of Finance*, 69(4):1643–1671.
- Mahdi, O. A., Mohammed, M. A., and Mohamed, A. J. (2012). Implementing a novel approach an convert audio compression to text coding via hybrid technique. *International Journal of Computer Science Issues (IJCSI)*, 9(6):53.
- Miller, B. P. (2010). The effects of reporting complexity on small and large investor trading. *The Accounting Review*, 85(6):2107–2143.
- Navissi, F. and Sridharan, V. (2017). Determinants of target costing adoption: a research note. *Journal of Management Accounting Research*, 29(1):67–77.
- Nelson, M. W. and Skinner, D. J. (2013). How should we think about earnings quality? a discussion of “earnings quality: Evidence from the field”. *Journal of Accounting and Economics*, 56(2-3):34–41.
- Oswal, S., Singh, A., and Kumari, K. (2016). Deflate compression algorithm. *International Journal of Engineering Research and General Science*, 4(1):430–436.
- Plumlee, M. A. (2003). The effect of information complexity on analysts’ use of that information. *The Accounting Review*, 78(1):275–296.
- Poynton, C. (2012). Digital video and HD: Algorithms and Interfaces. *Elsevier*.
- Pujar, J. H. and Kadlaskar, L. M. (2010). A new lossless method of image compression and decompression using huffman coding techniques. *Journal of Theoretical & Applied Information Technology*, 15.
- Rutherford, B. A. (2005). Genre analysis of corporate annual report narratives: A corpus linguistics–based approach. *The Journal of Business Communication (1973)*, 42(4):349–378.
- Shaw, P. and Pecorari, D. (2013). Types of intertextuality in chairman’s statements. *Nordic Journal of English Studies*, 12(1):37–64.
- Wade, G. (1994). Signal coding and processing. Cambridge university press.
- Wendai, L., Jing, F., and Bin, L. (2022). Does more voluntary environmental information disclosure cut down the cost of equity: heavy pollution industries in china. *Environmental Science and Pollution Research*, 29(42):62913–62940.
- Xu, W., Yao, Z., and Chen, D. (2019). Chinese annual report readability: measurement and test. *China Journal of Accounting Studies*, 7(3):407–437.
- You, H. and Zhang, X.-j. (2009). Financial reporting complexity and investor underreaction to 10-k information. *Review of Accounting studies*, 14:559–586.