Title	A text-mining analysis on non-financial reporting practices and the quality of disclosure
Sub Title	
Author	陸, 詩瓊(Lu, Shiqiong) 林, 高樹(Hayashi, Takaki)
Publisher	慶應義塾大学大学院経営管理研究科
Publication year	2021
Jtitle	
JaLC DOI	
Abstract	
Notes	修士学位論文. 2021年度経営学 第3888号
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40003001-00002021-3888

慶應義塾大学大学院経営管理研究科修士課程

学位論文（　　2021　年度）

論文題名

A Text-Mining Analysis on Non-Financial Reporting Practices and the Quality of Disclosure

主　査	林高樹
副　査	河野 宏和
副　査	村上 裕太郎
副　査	岩本 隆

氏　名	陸詩瓊

論 文 要 旨

所属ゼミ	林高樹　研究室	氏名	陸詩瓊

（論文題名）

A Text-Mining Analysis on Non-Financial Reporting Practices and the Quality of Disclosure

（内容の要旨）

As the trend of sustainability accelerates, stakeholders like investors, community, and employees have a need of non-financial information in their decision-making. However, the quality concern toward voluntary non-financial disclosure occurs.

Thus, this study was carried out to evaluate the quality of non-financial reporting practices and aimed to assist stakeholders such as investors and employees to understand non-financial discourses through the extent of non-financial disclosures' quality.

This study examined a total of 113 non-financial standard-alone disclosures of 26 companies in the automotive and parts sector. Specially, we have applied text-mining techniques to explore the quality measurement and conduct a cross-area comparison, including firms in East Asia, Europe, and North America.

This study showed that firms from different areas share the same tendency of disclosing less modified reports. Moreover, in the aspect of providing specific information, specificity performance is significantly associated with the length of disclosure. Some firms in this sector use their non-financial disclosures to provide more and much specific information, indicating that they pursue the high quality of voluntary non-financial disclosure and consider these disclosures as a great communication tool.

Additionally, results indicate that the company's size, profitability level, change of financial performance were found to be insignificantly associated with the extent of non-financial disclosure's quality.

CONTENTS

1. Introduction

1.1 Background and Motivation

In 2015, the Sustainable Development Goals (Known as SDGs) were set up by the United Nations General Assembly and were agreed by world leaders that they will be achieved by the year 2030. Furthermore, with the increasing attention on sustainability, organizations like PRI (Principles for Responsible Investment) work to promote sustainable investment through the incorporation of environmental, social, and governance (ESG) factors in investment decisions and active ownership. The concept of sustainable/responsible investment forces business leaders and investors to take action to help realize these goals.

Besides, many studies confirmed an increasing focus on non-financial information (Arvidsson 2011, Tarquinio et al., 2020). This increase appears to be both regulatory and demand-driven. Except for the information related to the sustainability issues, there is also significant demand for non-financial information related to other aspects, for example, intellectual capital information.

Therefore, there is a variety of users are paying interest in non-financial disclosure published by companies. At first, investors, one of the most important users, are calling for additional disclosure for their decision-making. Investors want to know how a company is considering the impact of these sustainability issues on its business model, risk strategy, and also the effect on its financial statements. Investors need to understand the future challenges that the company faces, and what the company’s plans are to deal with these challenges.

In addition to regulators and investors, other stakeholders like customers and employees also increasingly demand the communication of sustainability information about firms which may affect their purchasing and occupational behavior. For example, consumers who are sensitive to social and environmental issues may call for information about sustainable sourcing, manufacturing, and products. In the case of companies running the business between companies or business is under a call for sustainable acting in the supply chain by their business customers.

As the demand for additional disclosure by investors and other stakeholders is increasing, many companies have started voluntarily disclosing their Companies’ objectives, actions, and performances on sustainability issues, especially implications of some hot issues faced like climate-related challenges.

However, even though the non-financial disclosure is under increasing focus and practice, the variable quality is under concern as it is voluntary but not compulsory. Research has shown us that CSR reporting practices are symbolic rather than substantive (Michelon et al., 2015). Under the force of the demand, it seems that companies tend to enlarge the length of their annual reports or increase their disclosure frequency and reporting items. When it comes to the problem of diluted information, Michelon had shown us evidence that information disclosed in stand-alone reports and that in longer reports tend to be more diluted.

The fact is that non-financial disclosure, which is difficult to quantify even if it is necessary information for grasping the long-term strategy and future movements of a company, is under variable performance. To figure out the factors relative to the disclosure performance, the study also seeks to establish if differences exist in disclosure practice between the firms with different characteristics. Since the EU is known as a global leader in the movement of sustainability, firms in the EU seem to be heavily affected by regulation. In addition, conducting a comparison of different types of firms may help give evidence to identify those factors.

Towards the quality concern, we are wondering about the non-financial reporting practices and the quality difference of these disclosures between firms. In addition, we are also greatly interested in whether text mining techniques can be applied to assess the quality of non-financial disclosures.

1.2 Purpose of this study

The purpose of this research is twofold.
On the one hand, this study aims to conduct text analysis to evaluate the quality of non-financial discourses to assist stakeholders such as investors and employees to understand non-financial discourses through the extent of non-financial disclosures' quality. In this study, we apply the existing disclosure quality evaluation methods which have been used to evaluate textual information disclosed in traditional financial reporting to the voluntarily non-financial reporting.

On the other hand, we will explore factors that drive varying quality practices. To figure out the factors relative to the disclosure performance, we will test if differences exist in disclosure practice between the firms with different characteristics.

In this paper, 113 public disclosures (standalone reports) published by 26 firms listed in the Forbes ranking of the global top 2000 companies in 2021 have been examined.

Using text analysis techniques, we evaluate the quality of those non-financial reporting items and conduct a time-series period comparison as well as a comparison of disclosure practices among Continental European, East Asia, and North American companies. Moreover, the relationship between financial characteristics and the quality of non-financial disclosure was investigated additionally.

1.3 Originality and Contributions

In this study, we contribute to the non-financial disclosure literature by expanding the use of text mining techniques to evaluate the quality of non-financial reports, which was usually applied in the literature of traditional financial statements.

This study can assist stakeholders especially investors and employees to identify global automotive and parts companies through the extent of non-financial disclosures' quality which contributes to the understanding of determinants of non-financial disclosure to improve the communicative value and implementation of quality disclosure. In particular, the results of this study about the cross-area comparison of quality performance should quite matter to regulators and standard setters worldwide. It contributes to the future improvement of reporting requirements in each country/area.

Moreover, we propose a study-specific data-processing method and use the latest text-mining technique, BERT, to deal with complicated and large-size documents like sustainability reports even with hundreds of pages.

This study also contributes to providing new evidence to the skepticism about the use of longer non-financial reports as tools used to dilute information. Besides, we provide new evidence to the debate about the relationship between corporates' financial performance and voluntary disclosure which has even not established a consistent conclusion.

2. Literature Review

2.1 non-financial disclosure

2.1.1 What is non-financial disclosure?

Non-financial disclosure is regarded as an important tool for corporate communication. However, there is no generally accepted definition of non-financial disclosure, but this term has often referred to the process of communicating the social and environmental effects of organizations to the stakeholders (Tarquinio et al., 2020). With this in mind, we have chosen to examine the non-financial reporting covering the social and environmental issues beyond the traditional financial statements in this study. Hence, the initial study items include Annual reports, CSR reports; integrated reports; Environment reports; Sustainability reports; and other named reports with similar contents.

2.1.2 Stand-alone reporting

CSR reporting has evolved from information on the corporate environmental and social policies included in annual reports to stand-alone combined reports that include social, environmental, and economic/financial information (Buhr, 2002, Cho et al., 2015, Milne and Gray, 2010). Recent trends in environmental disclosure and reporting practices suggest a largely increasing number of stand-alone reports, which include social, environmental and economic/financial information (Cho et al., 2011).

In preparing the non-financial disclosure, a company's management chooses the issues that they consider sufficiently important or problematic to report publicly. According to some authors, what particularly makes these stand-alone reports remarkable is that they represent a clear engagement of corporations with the increasingly critical issues of environmental and social responsibility, as well as businesses sustainability (Gray & Herremans, 2012).

Additionally, as the stand-alone reports are always regularly annual produced, it allows an inter-firm comparison periodically.

Thus, even the non-financial disclosure can be issued on the companies' home-page descriptions or shown as a part of annual reports. In this paper, we focus on the Stand-alone non-financial reports and use them as our initial study items.

2.2 Disclosure Quality Evaluation Using Text-Mining Techniques

Firms publish non-financial reporting such as corporate social responsibility (CSR) reports in an increasing amount. However, although these reports are easily accessed, their quality is hard to evaluate because manual analyses are difficult due to the nature of text data. With the development of computer programs, an increasing amount of researchers use NLP (Natural language processing) to extract language features from these text data.

However, while there is substantial academic literature on quality evaluation using text-mining techniques towards textual files which were regulatory required and with strict guidelines like 10-K textual disclosure, attributes of voluntarily textual non-financial disclosure like sustainability report have received less attention. One of the reasons is that it is much more challenging to assess the content of voluntary non-financial disclosure due to the lengthy and complex documents.

2.2.1 Readability

Firstly, in many previous studies, readability is one of the most well-known tools for assessing the annual reports, as one of the quality indicators of text documents in reports(Li 2008, Loughran, T., & McDonald, B. 2013). The FOG index as well as the Gunning Fog index is the most commonly adopted text readability index. The index is defined as a weighted average of the number of words per sentence, plus the percentage of long words (containing more than two syllables). Regarding the formula, it means that the words and sentences should be exactly divided and counted by the machine. However, since there is always some special characters or punctuations wrongly converted from PDF documents into text files because of the Unicode or some other reasons, we consider the readability index is inapplicable for this study, because of the punctuation removal (including the sentences divided marks) process in this study was conducted in the data cleaning process. Besides, as the nature of voluntary disclosure, each firm has its style to convey the story that they want to show. Visual elements such as graphs, charts, tables, photographs, diagrams are commonly used. The PDF conversion process will extract all contents in these visual elements and contents listing with bullet points without exact sentences divided marks. Thus, as it is difficult to measure the sampling documents' sentence length and word complexity, the readability index, FOG index as well as the Gunning Fog index is not put into consideration in this paper to evaluate the quality of non-financial disclosure.

2.2.2 Boilerplate Index

Besides, the use of the Boilerplate index is also a well-established approach to assess the reporting disclosures. Researchers use the n-gram technique in text mining to identify boilerplate phrases (Lang and Stice-Lawrence, 2015). The percentage of sentences containing boilerplate phrases is measured as an index to evaluate the disclosure quality. For the same reason as readability, the boilerplate is also inapplicable for this study as the sentences splitting problem.

2.2.3 Specificity

The growth in textual research makes measuring specificity an important tool in assessing financial documents. Specificity, known as an index assessing the number of entities scaled by the total number of words. Entities will be identified by the Named Entity Recognizer (NER) tool, and always be categorized into seven types, including locations, people, organizations, dollar amounts, percentages, dates, or times (Hope et al., 2016). The specificity is higher, should be more relevant for management teams when they design their disclosure strategies and the more specific the disclosures are. In other words, using specific named entities tends to be more context-specific.

The prior study (Hope et al., 2016) showed us that more specific risk-factor disclosures benefit analysts, one significantly important kind of user of financial statements, to be better able to assess fundamental risk. Similarly, Konno's study (2010) also suggested that investors respond positively to financial reports with the high specificity of textual information when they are published.

The latest study (Paananen et al., 2021) also suggested that specific environmental liabilities disclosure reduces information asymmetry referring to less forecast error and dispersion conducted by analysts.

Seebeck and Kaya (2021) used methods from computational linguistics to quantify the level of specificity of auditor risk disclosures and found that a more specific description of KAM is significantly and positively associated with capital market reactions, suggesting that specificity performance is a desirable feature for investors.

Though it is unclear whether investors value specific non-financial disclosures such as sustainability reports as incrementally informative, given the evidence above in related areas of firm risk disclosures (Hope et al., 2016), environmental liabilities disclosures (Paananen et al., 2021), and auditor risk disclosures (Seebeck and Kaya, 2021), we assume that disclosures not only financial statements; but also the voluntary non-financial disclosures should not be so generic that they can be applied to any firm. The more precise the disclosures are, the more communicative value they provide to users. Hence, analysts or investors who use the non-financial disclosure to conduct the assessment or decision-making are affected by firms' disclosure practice on content-specific performance.

Following Hope, Hu, & Lu (2016), Paananen, Runesson & Samani (2021), and Seebeck & Kaya(2021), in our study, we calculate specificity as to how often the text refers to specific entities to capture whether the report specifically offers a snapshot related to corporate rather than general information.

2.2.4 Modification

The modification score has been used in the past to investigate the level of modification conducted by companies in the narrative information in the financial reporting such as MD&A.

Brown and Tucker (2011) used stickiness to assess the level of similarity of two documents and found that MD&A that are updated less over time (“sticky” disclosures) have muted stock price responses.

The algorithm used to calculate the similarity of any two documents is measured by the angle between the two vectors representing the documents: a smaller angle indicates more similar documents. This score is bounded between 0 and 1 with a higher score indicating more similarity as well as higher stickiness index ($\theta=0$, $\cos\theta=1$), and the modification score is 1 minus the square of similarity score.

Computation of similarity of specific objects in natural language processing is already applied to evaluate the non-financial information disclosed in the financial statement but not even in the voluntary disclosures such as sustainability reports.

In this research project, we aim to use the text mining technique above that has already been applied to evaluate the non-financial information in financial reports and explore the quality features from large-size non-financial disclosure documents.

2.3 Factors that affect the quality practice on voluntary non-financial disclosure

Based on the prior review (Zamil et al.,2021), drivers of corporate voluntary reporting such as company size, age, leverage, liquidity, profitability, corporate governance, and ownership structure were widely investigated. These firm-specific determinants were the most examined in the previous studies, however, the result is still inconclusive while work on the country-related factors was limited.

Firstly, a great amount of previous empirical studies revealed the company’s size is positively related to the extent of CSR disclosure (Giannarakis, 2014; Cowen, 1987; Rahman et al., 2011; Gamerschlag et al., 2011). Cowen (1987) suggested that in general, large companies tend to be more visible and receive attention from external stakeholder groups; thus, they publish more CSR information to legitimize their initiatives (Cowen, 1987). Morover, Giannarakis (2014) indicated that the company’s size, as measured by total assets, has a positive relationship with

CSR disclosure. Rahman (2011) pointed out several reasons for this reaction, such as that large companies are more visible to investors, absorb extra costs for CSR disclosure, attend the maintenance of their good corporate image and retain the customers' loyalty and talented employees.

Based on these previous reviews, we will use the turnover, the total asset, the total profits, and the market value of equity as the measure of company size.

As far as the empirical studies are concerned, the relationship between corporates' profitability and voluntary disclosure has not reached a consistent conclusion. For example, Gamerschlag et al. (2011) found a positive relationship between profitability and higher environmental disclosures while Ho & Taylor (2007) indicated that less profitable companies tend to disclose more information regarding social and environmental disclosures to demonstrate their contribution to society.

When it comes to country-specific determinants, previous studies also gave evidence that country-level factors affect the disclosure of non-financial reports. For example, Adnan et al. (2018) conducted a cross-country analysis and found that CSR reporting is more prevalent in companies in countries in which the society is individualistic and also in societies where there is low power distance. Ali et al. (2018) suggested the critical role of CSR-promoting institutions for CSR disclosure in low-income nations, which were thought to be lacking in developing countries. Moreover, De Villiers & Ana Marques (2016) discovered that countries with greater investor protection, higher levels of democracy, more effective government services, higher quality regulations, more press freedom, and a lower commitment to environmental policies are more likely to disclose more CSR information. Therefore, we can assume that the performance of non-financial disclosures varies in cross-countries companies.

2.4 Global Automotive and Parts Industry

Modapothala et al.(2010) used text mining and multi-discriminatory analysis to appraise corporate sustainability reports and found that disclosures made by the companies differ across industrial sectors. Thus, since non-financial disclosures among sectors are quite different, in this paper, we examine non-financial reporting disclosed by companies in one sector and conduct an inter-sector comparison on disclosure performance. Also, as discussed before, we focus more specifically on standalone Non-financial reports.

In this paper, we aim to examine non-financial reporting and evaluate the disclosure quality of them between groups categorized by geography characteristics. Hence, focusing on one sector facilitates comparison as we assume that companies in the same sector face similar challenges and also report to users with similar profiles. With the consideration discussed below, we have chosen the automotive and parts sector as our study sampling sector.

First, the automotive and parts sector represents a major component of the global economy and large companies in the business can be identified. Companies whose main business is selling automotive, as well as companies manufacturing parts are running their business all over the world in a great size. Thus, large companies in a different part of the supply chain can be identified if one wishes to figure out whether the corporate size and profitability characteristic affect the disclosure performance in the same sector.

The second reason to focus on the automotive and parts industry is that the operations of this sector are spread all over the world. For the purpose to explore whether the non-financial disclosures vary in firms in different continent groups, the worldwide operators in the automotive and parts industry is one of the ideal sampling choices. Putting eyes on this sector, we can compare the disclosure performance between companies in different countries.

Moreover, the automotive and parts sector consists of a complex supply chain spread all over the world. Companies in the automotive and parts sector are almost global investors with a wide range of businesses and employ a large number of workers globally. Thus, this sector is subject to social exposures across a variety of issues including labor relations, community impacts, and supply chain concerns which indicate that they may face a great demand for information about these issues to communicate. Besides, as the non-financial disclosures cover multiple topic concerns about the social and environmental aspects of the organizations, the profile of reporting users should be including a variety of stakeholders. Thus, not only the outsider users like investors and community; but also the insider users such as employees will be interested in the reporting disclosed. For our study, we consider that disclosure reports can be easily obtainable in the automotive and parts sector.

In sum, we believe that the automotive and parts sector is a suitable target for assessing the quality of non-financial disclosures in this study.

3. Hypotheses Development

As discussed above, we considered that some indices used in assessing the textual disclosure in financial reports like MD&A or regulatory-required files like 10K are also applicable to the voluntarily non-financial disclosure towards the quality concern.

First, the specificity helps to evaluate whether the non-financial disclosure is specific to help obtain desirable information from corporates even though the topics or content in these reporting varies widely among firms. A higher specificity score indicates that the disclosure should be more detailed and relevant.

Second, concerns towards modification occurs: are disclosures sticky over time? It can be assumed that companies in the same sector face similar business environments and operating conditions, so the modification scores should be lower if companies modify the reports less.

Given the costs of preparing long narrative documents, managers may simply use last year's disclosure as a template and make minor changes even they face great changes where changes can refer to economic changes like the change in operation and the change in the business environments.

Thus, the Vector Space Model (VSM)-based modification scores can capture changes in narrative disclosure, and show us the modification behaviors of corporates on the non-financial disclosures over time. A document that is very similar to that from the previous year does not reveal much new information.

In addition, we assume that there are some factors assumed to affect the disclosure performance among firms in the same sector.

As discussed in the introduction part and literature review, requirements are variable from the country. For example, EU law requires certain large companies to disclose information on the way they operate and manage social and environmental challenges so that companies in Europe should be under greater pressure to disclose specific and updated disclosure. Besides, in previous studies, some country-specific determinants were also suggested to affect the disclosure of non-financial reports. Thus, we can assume that there is a variety of quality performance in non-financial disclosure in different continents or countries.

However, regarding the Forbes ranking of the global top 2000 companies in 2021, the Forbs analyst suggested that the economic impact from largest companies all over the world has been greatly changed and tends to be a significant global imbalance(Eliza Haverstock, 2021). She indicated that roughly three-fourths of companies on the Global 2000 are based in just 10 countries. After the U.S. and China, Japan, the U.K., and South Korea are home to the most list-makers.

Besides, even located in the same East Asia continent, culture and economic development are greatly different among China, Japan, and South Korea. Hence, in this study, when we want to conduct a continental comparison on disclosure practice, we choose not to simply continentally group the sampling firms but categorize them into 5 groups, including the EU, North America, China, Japan, and South Korea. Using this standard to group sampling firms, we develop our first hypothesis below.

Hypothesis 1: The level of <u>quality measurements</u> of non-financial disclosure differ among firms in five area groups.
H1-1. The level of <u>specificity</u> of non-financial disclosure differs among firms in five area groups.
H1-2. The level of <u>modification</u> of non-financial disclosure differs among firms in five area groups.

Second, regarding concerns towards sticky disclosure, the following hypotheses will be also tested.

Hypothesis 2: There are tendencies changes in the quality measurements.
H2-1. The level of <u>specificity</u> of non-financial disclosure decrease over time.
H2-2. The level of <u>modification</u> of non-financial disclosure decrease over time.

Next, Michelon 's research in 2015 showed us that Based on the previous study, information disclosed in stand-alone reports and that in longer reports tend to be more diluted.
To test whether longer reports have poor quality, hypothesis 3 was established. Consistent with the previous studies, we supposed that the quality measurements of non-financial disclosure are negatively associated with the length of disclosures.

Hypothesis 3: The quality measurements of non-financial disclosure are associated with the length of disclosure.
H3-1. The level of <u>specificity</u> of non-financial disclosure is negatively associated with the length of disclosure.
H3-2. The level of <u>modification</u> of non-financial disclosure is negatively associated with the length of disclosure.

Finally, even if the previous empirical results are contradictory, it is hypothesized, in this study, that financial performance (including companies' size, profitability, and the level of financial performance's change) positively affects the quality of voluntary disclosure:

Hypothesis 4: There is a positive relationship between financial performance and the quality measurements of non-financial disclosure.
H4-1. There is a positive relationship between financial performance and the specificity level of non-financial disclosure.
H4-2. There is a positive relationship between financial performance and the modification level of non-financial disclosure.

To test the four hypotheses developed above, we conduct the main analysis methods called LMM (Linear Mixed Model) and ANOVA (analysis of variance).

4. Data and Research Design

4.1 Sample Selection and Data Collection

4.1.1 Sampling companies and Text Data Collection

There are two types of data used in this research. The first is text data from non-financial reports. The second is the financial performance data of companies.

To collect data for our study, we started with picking up sampling companies and then downloaded reports disclosed by these companies. Non-financial disclosures of major companies in the Global Automotive and Parts sector are obtained and used as the basis for text mining analysis.

The companies chosen are the largest companies listed in the Forbes ranking of the global top 2000 companies in 2021. The Forbes ranking takes into account sales, profits, assets, and market values of the companies and provides a useful indicator of the biggest public companies in the world. Thus, these prominent companies should provide meaningful and representative insights on the practice of non-financial disclosures in the automotive and parts industries.

All stand-alone nonfinancial disclosures issued in the 5 years (2016-2020) by companies listed in **Table1** are downloaded either from the corporate website or Corporate Register. The 5-year period covers a long period that provides the opportunity to evaluate the normal quality of voluntary disclosure over time.

Table 1: List of sampling companies

No.	Name of the companies	Name used in coding	Countries/ Territories	Area Group
1	Toyota Motor Corporation	Toyota	Japan	Japan
2	Volkswagen AG	Volkswagen	Germany	EU
3	Daimler AG	Daimler	Germany	EU
4	General Motors Company	GM	USA	North America
5	BMW AG	BMW	Germany	EU
6	Honda Motor Company Ltd	Honda	Japan	Japan
7	Hyundai Motor Company	Hyundai	South Korea	South Korea
8	AB Volvo	Volvo	Sweden	EU
9	Tesla Inc	Tesla	USA	North America

10	Kia Motors Corporation	KIA	South Korea	South Korea
11	Ford Motor Company	Ford	USA	North America
12	Suzuki Motor Corporation	Suzuki	Japan	Japan
13	BYD Company Limited	BYD	China	China
14	Subaru Corporation	Subaru	Japan	Japan
15	Nissan Motor Co Ltd	Nissan	Japan	Japan
16	Great Wall Motor Company Limited	GreatWall	China	China
17	Michelin Group	Michelin	France	EU
18	Magna International	Magna	Canada	North America
19	Denso	Denso	Japan	Japan
20	Aptiv	Aptiv	Ireland	EU
21	Continental	Continental	Germany	EU
22	Bridgestone	Bridgestone	Japan	Japan
23	Aisin Seiki	Aisin	Japan	Japan
24	LKQ	LKQ	USA	North America
25	BorgWarner	BorgWarner	USA	North America
26	Knorr-Bremse	KnorrBremse	Germany	EU

Selection criteria on reporting are stated following.
First, we examine only English-language reporting which can be under the same text mining process.

Second, we examine only text-convertible reporting when it’s in the PDF version. In the case of BYD in 2020 and Tesla in 2019, the PDF version of the reports were created from a scanned document so the text in the disclosures could not be extracted. Hence, these two reports are excluded from the sampling list.

Third, only one stand-alone report was added to the sample for each company each year. A firm may publish periodic non-financial reports in numerous forms, including sustainability reports, annual reports, impact reports, integrated reports, corporate social responsibility reports, social and environmental reports, and annual & sustainable development reports. The author examined the reports and found that specific reports are always specialized publications excerpted from the integrated one. For example, the “Environmental Report 2016. Toward the

Toyota Environmental Challenge 2050" issued by Toyota is covering only environmental initiatives excerpted from the "Sustainability Data Book 2016". Hence, only the longest stand-alone non-financial reports were selected.

Moreover, we noticed that the reporting period varies from company to company (i.e., January to December, April to March, etc.) and the reporting year can be something like 2019-20. We examined the content disclosed in the reports and figured out the reporting period based on the contents. Then we reclassified the reports by this substantive reporting period to a specific calendar year. (i.e., report for the year ended December 31, 2020, and report for the year ended March 31, 2021 are classified into the group Year 2020).

Table 2: Sample composition

Firm		Continent		Year	
Variable	No. of Obs	Variable	No. of Obs	Variable	No. of Obs
Toyota	5	Asia	57	2016	21
Volkswagen	5	EU	38	2017	21
Daimler	5	North America	18	2018	24
GM	5			2019	24
BMW	5			2020	23
Honda	5				
Hyundai	5				
Volvo	5				
Tesla	2				
KIA	5				
Ford	5				
Suzuki	4				
BYD	4				
Subaru	5				
Nissan	5				
GreatWall	5				
Michelin	5				
Magna	2				
Denso	4				
Aptiv	3				
Continental	5				
Bridgestone	5				
Aisin	5				
LKQ	1				

BorgWarner	3				
KnorrBremse	5				
Total	**113**	**Total**	**113**	**Total**	**113**

After these adjustments to the data, we finally collected a sample of 113 report observations issued by 26 companies based in 9 different countries that referred to the period 2016–2020. Details related to the sample composition categorized by firms, continents, years, and business models are provided in **Table 2**.

4.1.2 Financial Data Collection

Financial Data were calculated based on the data (sales, profits, assets, and market values in dollars) collected from the Forbes ranking of the global top 2000 companies list, which is also taken into account for the Forbes ranking.

Table 3: Financial Data Type

data	**Description**
lgMV	the natural logarithm of the market value of the sampling firm
lgSales	the natural logarithm of sales amount of the sampling firm
lgProfit	the natural logarithm of profits amount of the sampling firm
lgAssets	the natural logarithm of assets amount of the sampling firm
ROA	Return On Asset Ratio
PER	Price Earnings Ratio
PMR	Profit Margin Ratio
change_Sales	Percentage change in Sales value over a year-period
change_Profits	Percentage change in Profits value over a year-period
change_Assets	Percentage change in Assets value over a year-period
change_MV	Percentage change in MV value over a year-period

4.2 Text data processing

When we processed the text data before analysis, we began from files conversation. The **Pdfminer** package in python was used to convert the PDF version of the non-financial reporting to text files. The output of this step was text files containing all text contents extracted from the PDF documents, including word, punctuation, and numbers. Then we performed some basic preprocessing following to prepare the text data for analysis.

The first step comes with removing the noises in the data; here in the text domain, noise is referred to as something like special non-alphabetic characters, URLs, and use of parentheses, square brackets, white spaces, line break, and punctuations.

Secondly, we transformed all the uppercase words into lowercase words to ensure that the same strings in different cases will be equivalent to each other. Then, the texts were split into smaller units (tokens to words) and runs normalization (stemming, lemmatization) to get the base forms of words. After the above process, the entire corpus is consists of a list of words tokenized and noise removed of all 113 reporting observations. Finally, text may contain stop words such as "is, am, are, this, a, an, the, etc.", so we also remove them with the filter function.

Preprocesses mentioned above are conducted in the same python environment using Colaboratory, which allows to write and execute Python code through a browser from Google Research.

4.3 Quality Measurements

4.3.1 Quality Measurements: Specificity Score

Specificity, as the number of specific words or phrases conveying specific information relevant to the disclosing firm (e.g. persons or products), divided by the number of total words in the non-financial disclosures.

To implement this construct on a large-scale length of non-financial disclosure documents, we use the Named Entity Recognition (NER) technique to identify and extract specific names belonging to eighteen entity categories listed below. NER refers to a natural-language-processing task that seeks to locate and classify atomic elements in text into predefined categories. In this study, we use the free and open-source library, spaCy, to conduct the work of featuring NER in Python.
SpaCy's named entity recognition has been trained on the OntoNotes 5 corpus and it currently supports the following 18 different entity types, including person, organization, and location. **Table 4** presents entity types and detailed descriptions of these eighteen categories.

Table 4: Named Entity types

Source: https://catalog.ldc.upenn.edu/docs/LDC2013T19/OntoNotes-Release-5.0.pdf

Type	Description
PERSON	People, including fictional.
NORP	Nationalities or religious or political groups.

FAC (FACILITY)	Buildings, airports, highways, bridges, etc.
ORG (ORGANIZATION)	Companies, agencies, institutions, etc.
GPE	Countries, cities, states.
LOC (LOCATION)	Non-GPE locations, mountain ranges, bodies of water.
PRODUCT	Vehicles, weapons, foods, etc. (Not services)
EVENT	Named hurricanes, battles, wars, sports events, etc.
WORK_OF_ART	Titles of books, songs, etc.
LAW	Named documents made into laws.
LANGUAGE	Any named language.
DATE	Absolute or relative dates or periods.
TIME	Times smaller than a day.
PERCENT	Percentage, including “%”.
MONEY	Monetary values, including unit.
QUANTITY	Measurements, as of weight or distance.
ORDINAL	“first”, “second”, etc.
CARDINAL	Numerals that do not fall under another type.

Table 5: Checking of entity types in Colaboratory

```
nlp = spacy.load("en_core_web_sm")
ner_lst = nlp.pipe_labels['ner']
print(len(ner_lst))
print(ner_lst)
```

```
18
['CARDINAL', 'DATE', 'EVENT', 'FAC', 'GPE', 'LANGUAGE', 'LAW', 'LOC', 'MONEY', 'NORP',

'ORDINAL', 'ORG', 'PERCENT', 'PERSON', 'PRODUCT', 'QUANTITY', 'TIME', 'WORK_OF_ART']
```

For words in the disclosure documents after preprocessing, in a document published by firm *i,* in year *t,* the amount of named entities recognized by the NER technique is defined as $ent_{i,t}$ while all words were tokenized in the same document is counted as $total_{i,t}$. We use the following formula to calculate the $specificity_{i,t}$ of the document published by firm *i,* in year *t*.

$$specificity_{i,t} = \frac{ent_{i,t}}{total_{i,t}} \qquad (1)$$

In addition, we assumed that non-financial reporting disclosed by companies shares similar vocabulary and grammar with general management magazines like Harvard Business Review.

Thus, the article "Sustainable Business Went Mainstream in 2021", which was downloaded from the Harvard Business Review's official website, was used to check the appropriateness of using spaCy to conduct the work of featuring and counting NER in Python.

The message is that managing climate and other ESG issues is core to business value. Many banks agreed: JPMorgan Chase, Citi, Morgan Stanley, and Bank of America (to name a few) committed from $1 trillion to $2.5 trillion to invest in climate action (clean technologies) and sustainable development (e.g., affordable housing and efforts to improve racial equity). For context, I worked with Bank of America in 2008 on the first commitment of this kind — it was for $25 *billion*. Trillions is serious, mainstream money. And at the COP26 meeting in November, a new group representing $130 trillion in assets (that's a lot — well above global

annual GDP) formed the Glasgow Financial Alliance for Net Zero, co-chaired by Michael Bloomberg and former Bank of England head Mark Carney.

Figure 1: Excerpt from <Sustainable Business Went Mainstream in 2021 >

First, we used the same tool **Pdfminer** package in python in the study to convert the PDF article to text files. Then the output of text contents extracted from the PDF document was used to perform NER in Python with spaCy. The visualization output of name entities extracted is shown below using the "displacy" package of spaCy.

message is that managing climate and other ESG issues iscore to business value. Many banks agreed: JPMorgan Chase PERSON , Citi GPE , Morgan Stanley ORG , and Bank of America ORG (to name a few) committedfrom $1 trillion MONEY to $2.5 trillion MONEY to invest in climate action (cleantechnologies) and sustainable development (e.g., affordablehousing and efforts to improve racial equity). For context, Iworked with Bank of America ORG in 2008 DATE on the first ORDINAL commitment ofthis kind – it was for $25 billion MONEY . Trillions PERSON is serious, mainstreammoney. And at the COP26 meeting in November DATE , a new grouprepresenting $130 trillion MONEY in assets (that's a lot – well above global annual DATE GDP) formed the Glasgow Financial Alliance ORG for Net Zero PRODUCT ,co-chaired by Michael Bloomberg PERSON and former Bank of Englandhead Mark Carney ORG .Supply chains:

Figure 2: Excerpt from <Sustainable Business Went Mainstream in 2021 > NER output

The named entities in the article highlighted by the author (**Fiture2**) and that extracted using Python with spaCy come to the same results of 18 entities.

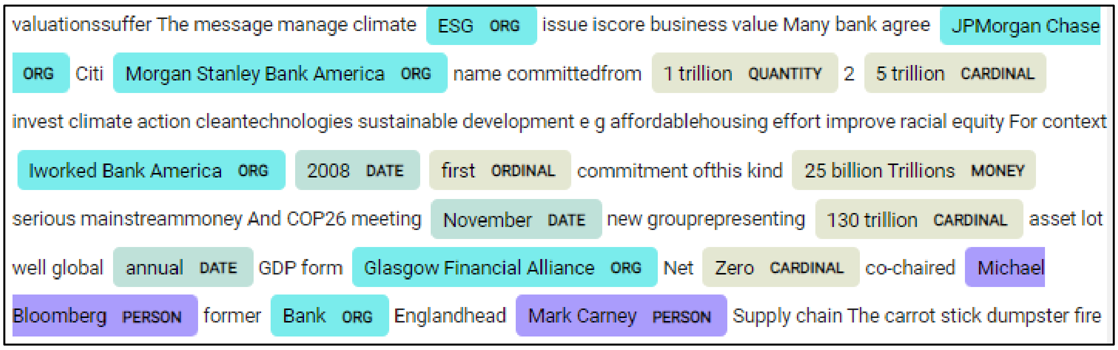


Figure 3: Excerpt from <Sustainable Business Went Mainstream in 2021 > NER output after cleaning preprocess

Besides, to check the effect on NER from data processing, we also perform the NER using the article after pre-processing including non-alphabetic characters and stop words removal, lowercase transformation. We found that some entities like "Citi" was failed to be featured after the punctuation removal processing, the "Bank of America" was also missed as the word "of" was removed in the stop words removal processing.

Table 6: Specificity result for two different preprocessing

	subject	content	word_count	ents_count	specificity
1	HBR_NER	GovernmentSustainable Business WentMainstream ...	2104	139.0	0.066065
2	HBR_Count	GovernmentSustainable Business WentMainstream ...	1395	131.0	0.093907

Thus, the specificity calculated for these two different preprocessing were quite different. In this study, we use the common-used NLTK library to remove stop words in documents. The stop words list in this library is composed of 179 words currently, some of them was listed below.

Examples of stop words list:

> *'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'during', 'before', 'after', 'above', 'below'*

As the goal of calculating the specificity is to figure out the detailed information provided by management from non-financial disclosures, we consider that the stop words were meaningless but important in the grammar. Besides, the amount of stop words and punctuation is great which will lead to a diluted specificity result. For these considerations, we use the different pre-processing corpus to calculate the specificity index. In other words, we count the output of NER performed on the corpus keeping stop words and punctuation left as the named entities amount; while considering the total number of words in the text after stop words and punctuation removal processing, measured as the number of total words.

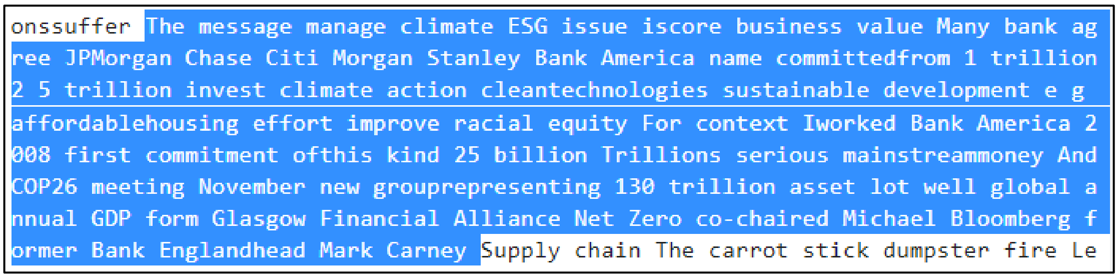

onssuffer The message manage climate ESG issue iscore business value Many bank agree JPMorgan Chase Citi Morgan Stanley Bank America name committedfrom 1 trillion 2 5 trillion invest climate action cleantechnologies sustainable development e g affordablehousing effort improve racial equity For context Iworked Bank America 2008 first commitment ofthis kind 25 billion Trillions serious mainstreammoney And COP26 meeting November new grouprepresenting 130 trillion asset lot well global annual GDP form Glasgow Financial Alliance Net Zero co-chaired Michael Bloomberg former Bank Englandhead Mark Carney Supply chain The carrot stick dumpster fire Le

Figure 4: Excerpt from <Sustainable Business Went Mainstream in 2021 > cleaning result of documents for counting words

The recalculation of specificity of the article **<Sustainable Business Went Mainstream in 2021 >** should be ***139/1395=0.09964***.

When it comes to the sampling document in this study, we randomly choose the sustainability report of 2020 disclosed by Ford Motor Company in 2021 to check the performance of NER in Python with Spacy.


Recognition for the F-150 and Mustang Mach-E

In the first few weeks of 2021, two of our vehicles received major accolades, including the 2021 Mustang Mach-E being named as North American Utility Vehicle of the Year™ and the 2021 F-150 being named the North American Truck of the Year™. The first time since 2014 that one brand has won multiple North American vehicle of the year awards in one year, these accolades highlight our constant dedication to innovation. It sets these vehicles apart as benchmarks for design, safety, handling, driver satisfaction and value for money.

The Mustang Mach-E has received further recognition for the features and driving experience it offers, winning Best Car to Buy in 2021 by The Car Connection and Green Car Reports.

We are honored to have received such recognition for two vehicles that we have worked on diligently to combine sustainability with function and user experience.

Figure 5: Excerpt from < Ford_2020_Integrated Sustainability and Financial Report > Page 44

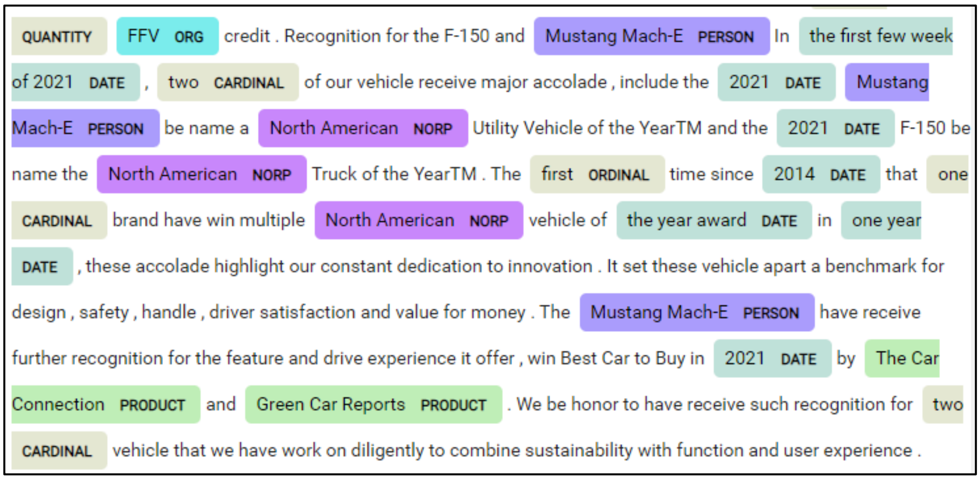


Figure 6: Example of NER using spaCy, Excerpt from < Ford_2020_Integrated Sustainability and Financial Report >

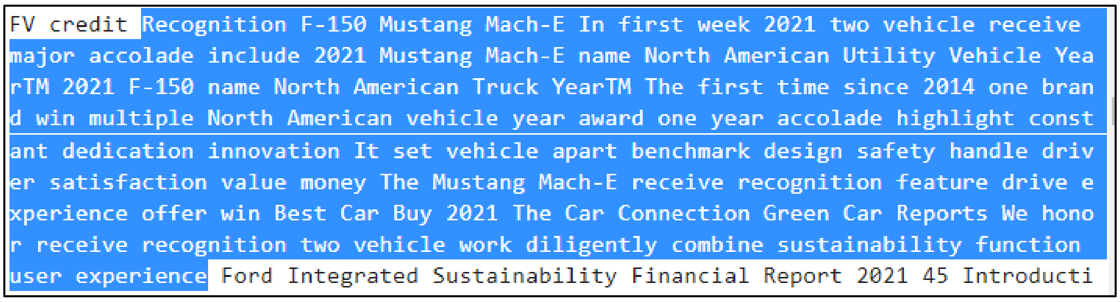


Figure 7: Cleaning Result For Total Word Counting, Excerpt from < Ford_2020_Integrated Sustainability and Financial Report >

To sum, our sample-specific index of named entities is better suited to our specific context compared to the index calculated by the number of named entities scaled by the total number of raw documents.

4.3.2 Quality Measurements: Modification Score

As same as most prior studies, we use the definition of cosine similarity to calculate the level of modification in a firm's current year non-financial reporting to that from the previous year. Two documents expressing the same meaning with dissimilar wording will return a lower similarity score than two documents that happen to contain the same words while expressing different meanings. It means that a document with a high similarity score is very similar to that from the previous year does not reveal much new information. Therefore, the modification score which is calculated from 1 minus the square of similarity score, owns an opposite meaning. It measures the extent to which two documents are different and indicate the level of modification in disclosures between years.

The similarity is determined by comparing word vectors or word embeddings, multi-dimensional meaning representations of a word. One of the most common and effective ways of calculating similarities is; Cosine Distance/Similarity. It is the cosine of the angle between two vectors, which gives us the angular distance between the vectors.

We use the Vector Space Model (VSM) to calculate the similarity score as Varun (2020) summarized, which is also called cosine similarity. The VSM represents a document as a vector in an n-dimensional Euclidean space. The similarity of any two documents is measured by the angle between the two vectors representing the documents: a smaller angle indicates more similar documents.

Thus, the VSM-based similarity score is defined as follows:

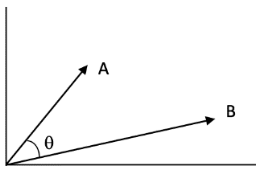
$$Sim = \cos(\theta) = \frac{v_1}{||v_1||} \cdot \frac{v_2}{||v_2||} = \frac{v_1 \cdot v_2}{||v_1||\ ||v_2||},$$

Formula to calculate cosine similarity between two 768 dimension vectors A and B is,

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

where θ is the angle between v1 and v2, ‖v1‖ is the vector length of v1, and ‖v2‖ is the vector length of v2.

In a two-dimensional space, it will look like this,



The similarity score is bounded between 0 and 1 with a higher score indicating more similarity (cosθ= 1). For straight understanding towards modification level between documents, the modification score is defined as 1 minus the square of similarity score. The formula of modification score is shown as follows: (2)

$$\boldsymbol{Modification\ \ score\ \ = 1 - cos\theta^2} \qquad (2)$$

4.3.3 BERT

In the study, we used the pre-trained BERT (Bidirectional Encoder Representations from Transformers) (Devlin, Jacob, et al., 2018) model from Huggingface to embed our corpus.

The BERT base model has 12 layers (transformer blocks), 12 attention heads, 110 million parameters, and a hidden size of 768. BERT is trained on unlabelled text including Wikipedia and Book corpus. BERT uses transformer architecture, an attention model to learn embeddings for words. BERT consists of two pre-training steps Masked Language Modelling (MLM) and Next Sentence Prediction (NSP). In BERT training text is represented using three embeddings, Token Embeddings + Segment Embeddings + Position Embeddings.

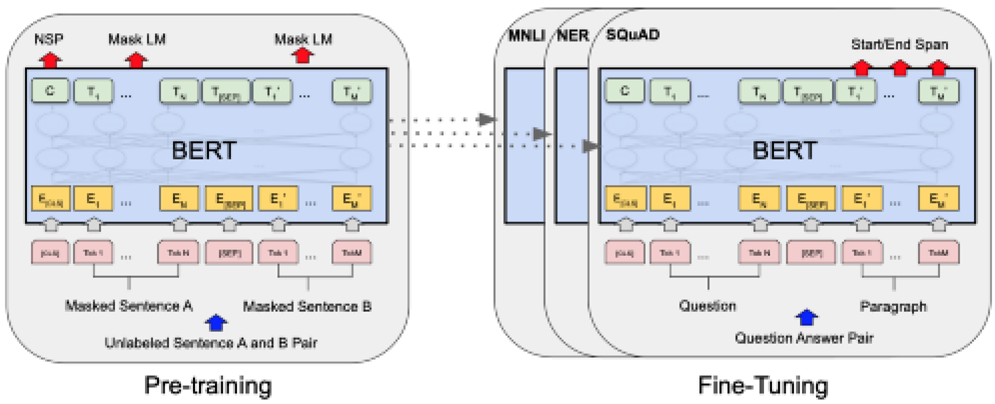


Figure 8 BERT training architecture (Image from https://arxiv.org/pdf/1810.04805.pdf)

In our study, we used this pre-trained BERT model to get the vectors with 768 dimensions and then calculated the cosine similarity between every two non-zero vectors referring to specific documents.

```
array([[-0.3383414 ,  0.9402209 ,  0.86816096, ..., -0.33399346,
        -0.02846767, -0.15315923],
       [-0.35999316,  1.0729098 ,  0.94970113, ..., -0.44663033,
         0.30602345, -0.05951089],
       [-0.31903636,  1.0129235 ,  0.8655949 , ..., -0.36769807,
         0.1857047 , -0.20660567],
       ...,
       [-0.5272247 ,  0.94025165,  1.4199097 , ..., -0.5037987 ,
         0.06029107,  0.24659222],
       [-0.32831478,  0.45648575,  1.350754  , ..., -0.94171363,
        -0.42582616, -0.40759784],
       [-0.61402035,  0.45634538,  1.2624216 , ..., -0.68097365,
        -0.46326974, -0.5893667 ]], dtype=float32)
```

Figure 9: Vectors generated by BERT

Due to the pre-process problem as discussed on the part of specificity, we use the most cleaning corpus to generate the vectors. First, as the stop words, punctuation, and the corporates' names can be assumed to be meaningless, the change of those words should not be considered to be a substantial modification, we conduct the removal process. Besides, only updating the numeric digits also cannot be taken into account as a substantial modification, numeric digits should be also removed before generating the vector using BERT.

4.3.4 Sum of different corpuses

As mentioned above, in computing the two metrics, we perform a differentiated text preprocessing pipeline based on research-specific requirements. Therefore, we use Table 6 to summarize the preprocessing steps and corresponding purposes of three corpuses.

Table 7: Sum of different corpuses

Corpus	Preprocessing steps	purpose
Corpus 1	• URL removal • Normalization • Firms' name removal	• Featuring and counting named entities using spaCy • For calculating the specificity score
Corpus 2	• URL removal • Normalization • Punctuation(including special characters) removal • Lemmatization • Firms' name removal and stop words removal	• Counting meaningful total words in the text • For calculating the specificity score
Corpus 3	• URL removal • Normalization • Punctuation(including special characters) removal • Numeric digits removal • lowercase • Lemmatization • Firms' name removal and stop words removal	• Generating vectors using BERT • For calculating the modification score

4.4 Sum of Variables

To investigate the relationship between the quality of disclosures and the characteristics of firms, we use a variety of variables in our research.

First, the quality measurements including specificity score and modification score were taken as quality variables, calculated using text-mining techniques like NER and BERT by the author. Second, other quantitative variables introduced above were also used in our study, including length of sampling reports, company size (market value, turnover, profit, and asset), profitability (ROA, PER, profit margin ratio), and change variables.

Table 8: Variables List

Variables	**Description**
specificity	the number of named entities scaled by the number of total words in the sampling reports while total words are defined as total counting words of Corpus2.
modification	1 minus the square of cosine similarity score between two 768 dimension vectors representing every two annual reports from the same sampling company
period	Numeric agency for years. Year 2016: Period = 1; Year 2017: Period = 2; Year 2018: Period = 3 Year 2019: Period = 4; Year 2020: Period = 5
lg_word	length of sampling reports, measured as the natural logarithm of total counting word of Corpus2
lgMV	the natural logarithm of the market value of the sampling firm
lgSales	the natural logarithm of sales amount of the sampling firm
lgProfit	the natural logarithm of profits amount of the sampling firm
lgAssets	the natural logarithm of assets amount of the sampling firm
ROA	Return On Asset Ratio
PER	Price Earnings Ratio
PMR	Profit Margin Ratio
change_Sales	Percentage change in Sales value over a year-period
change_Profits	Percentage change in Profits value over a year-period
change_Assets	Percentage change in Assets value over a year-period
change_MV	Percentage change in MV value over a year-period

4.5 Data Observation

4.5.1 Summary Statistics and Cross Matrix of Variables

The collected and calculated data is composed of two main types of data, one is quantitative indices of sampling reporting documents, and the other is financial data including market value, sales, profits, assets, return on total asset ratio, price-earnings ratio, profit margin ratio, and change rate of market value, sales, profits, assets.

Before the empirical research, we observe the summary Statistics and Cross Matrix of these data shown in **Table9** and **Table 10**.

From **Table9**, we found that even both specificity and modification are within the range of zero to one, specificity is in a smaller range between 0.063 and 0.172 while the modification is from 0.003 to 0.624. Compared with the range in previous studies, scores calculated in this study are in a different range. For example, in Hope's study (2016), the mean for specificity of risk-factor disclosures in 10-K files is 0.054 while that for the rest of the 10-K (excluding risk-factor disclosures) and MD&A section is 0.19 and 0.23. In the case of modification, modification scores of MD&A documents in Brown's study (2010) ranged from 0 to 0.97, and the mean value is 0.155.

We considered the following reasons for the differences in these distribution ranges. First, the object of analysis, that is, the type of text files is different. In Hope's analysis(2016) we can see that even within the same 10-K file, text descriptions in different parts had different levels of specificity. However, in our research, the observations are non-financial reports, which differ from the 10-K files in terms of length, wording, and content. Second, in the pre-processing part, which is an important part of the text analysis process, different authors may use different tools and steps, thus affecting the final analysis results. Third, in terms of specificity, there are some commonly used tools for named entity recognition, such as Stanford NER and SpaCy NER. In Hope's study(2016), 7 types of specific names were identified using the Stanford NER tool ((1) names of persons, (2) names of locations, (3) names of organizations, (4) quantitative values in percentages, (5) money values in dollars, (6) times, and (7) dates), while in our analysis using SpaCy NER tool, 18 types discussed above were identified. The standard and definition of the distinction between types will be different, which is also considered to be one of the reasons for the final specificity scores' range difference. Similarly, in the calculation of the modification scores, we used BERT to generate 768-dimensional vectors.

Moreover, financial data range from minus amount to great amount such as the PER amount of Tesla in 2020, which was considered as an outlier. The great-range financial performance data allow a comparison of non-financial disclosure practice between great financial performers and worse financial performers.

In **Table 10** and **Figure 10**, we can have a quick look at the correlation of variables. The distribution of each variable is shown on **Figure 10**'s diagonal and the value of the correlation plus the significance level as stars were shown on the top of the diagonal. We can find that the specificity seems to have a significant correlation with the counting word indicator (lg_word) while weak correlations with the company size variables (lg_Sales, lg_Profits, and lg_Assets). Nevertheless, there are great correlations among variables of company size (lg_MV, lg_Sales, lg_Profits, and lg_Assets).

The bivariate scatter plots with a fitted line are displayed under the bottom of the diagonal. the scatter plot of specificity and lg_word showed us that there is seem to be a positive correlation between the specificity level and the word counted.

Table 9: Summary Statistics of Variables

	specificity	modification	lg_word	ROA	PER	PMR	lgMV	lgSales	lgProfit	lgAssets	change_ Sales	change_ Profits	change_ Assets	change_ MV
Min.	0.06274	0.003332	3.595	-0.065	-131.651	-0.1403	9.771	9.845	8.301	9.851	-0.25234	-27.3708	-0.09113	-0.6796
1st Qu.	0.10885	0.092024	4.336	0.0189	7.682	0.0301	10.261	10.458	9.041	10.466	-0.0494	-0.35045	-0.00302	-0.17488
Median	0.12609	0.217464	4.589	0.0347	11.849	0.0488	10.487	10.657	9.342	10.695	0.01825	-0.14009	0.06313	-0.02194
Mean	0.1233	0.226539	4.516	0.03716	13.193	0.04507	10.505	10.725	9.356	10.85	0.03054	-0.35777	0.243337	0.17686
3rd Qu.	0.13737	0.327016	4.722	0.0534	17.278	0.068	10.708	11.096	9.667	11.343	0.06259	0.08354	0.136942	0.24552
Max.	0.17162	0.623696	5.254	0.1244	88.58* (Note)	0.1374	11.851	11.448	10.356	11.811	1.64111	8.81566	17.12761	3.92441

*Note: RStudio found the maximum amount of PER without counting the PER amount of Tesla in 2020 as an outlier. The PER amount of Tesla in 2020 is 1029.13.

Table 10: Cross Matrix of Variables

	specificity	modification	lg_word	ROA	PER	PMR	lgMV	lgSales	lgProfit	lgAssets	change_Sales	change_Profits	change_Assets	change_MV
specificity	1	0.068	0.528	-0.131	-0.135	-0.012	0.107	0.225	0.228	0.229	-0.018	-0.02	0.047	-0.042
modification	0.068	1	-0.015	0.051	0.054	0.094	0.071	0.048	0.02	0.055	0.155	0.079	-0.059	-0.043
lg_word	0.528	-0.015	1	-0.249	-0.017	-0.132	0.28	0.461	0.333	0.467	-0.122	-0.067	-0.021	-0.054
ROA	-0.131	0.051	-0.249	1	0.127	0.895	-0.164	-0.42	0.019	-0.487	0.154	0.38	0.077	-0.128
PER	-0.135	0.054	-0.017	0.127	1	0.153	-0.139	-0.142	-0.664	-0.123	-0.075	0.093	-0.027	0.078
PMR	-0.012	0.094	-0.132	0.895	0.153	1	0.081	-0.203	0.325	-0.227	0.122	0.551	0.037	-0.063
lgMV	0.107	0.071	0.28	-0.164	-0.139	0.081	1	0.693	0.734	0.725	-0.02	0.032	-0.104	0.371
lgSales	0.225	0.048	0.461	-0.42	-0.142	-0.203	0.693	1	0.79	0.979	-0.076	-0.002	-0.037	-0.078
lgProfit	0.228	0.02	0.333	0.019	-0.664	0.325	0.734	0.79	1	0.766	0.062	0.166	-0.005	-0.089
lgAssets	0.229	0.055	0.467	-0.487	-0.123	-0.227	0.725	0.979	0.766	1	-0.109	-0.036	-0.06	-0.003
change_Sales	-0.018	0.155	-0.122	0.154	-0.075	0.122	-0.02	-0.076	0.062	-0.109	1	0.109	0.816	-0.072
change_Profits	-0.02	0.079	-0.067	0.38	0.093	0.551	0.032	-0.002	0.166	-0.036	0.109	1	-0.012	-0.046
change_Assets	0.047	-0.059	-0.021	0.077	-0.027	0.037	-0.104	-0.037	-0.005	-0.06	0.816	-0.012	1	-0.073
change_MV	-0.042	-0.043	-0.054	-0.128	0.078	-0.063	0.371	-0.078	-0.089	-0.003	-0.072	-0.046	-0.073	1

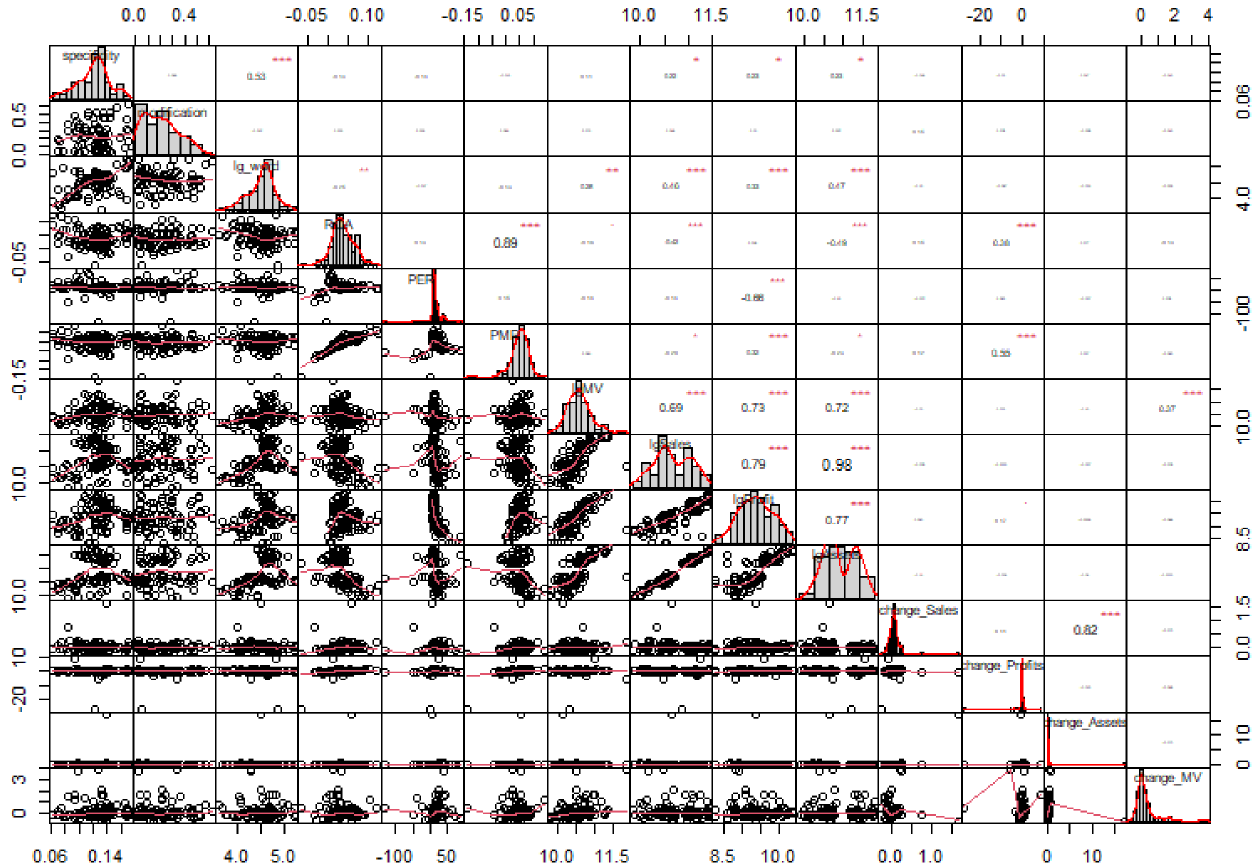


Figure 10: Correlation of Variables*

*Note: Each significance level is associated to a symbol : p-values(0, 0.001, 0.01, 0.05, 0.1, 1) <=> symbols("***", "**", "*", "."," ")

4.5.2 Visualization of Statistics Data

To have a quick look at the data, we drew the scatter plots and boxplots of specificity score and modification score.

From the scatter plots of specificity Score on firm-level (**Figure 11**), and the boxplot of specificity Score on firm-level (**Figure 12**), the level specificity of each firm's reports seems to be stable over time but vary from a firm.

Besides, **Figure 13** reveals that the majority of firms' modification is shown in a decreasing trend while the difference of modification among firms shown in **Figure 14** is not obvious.

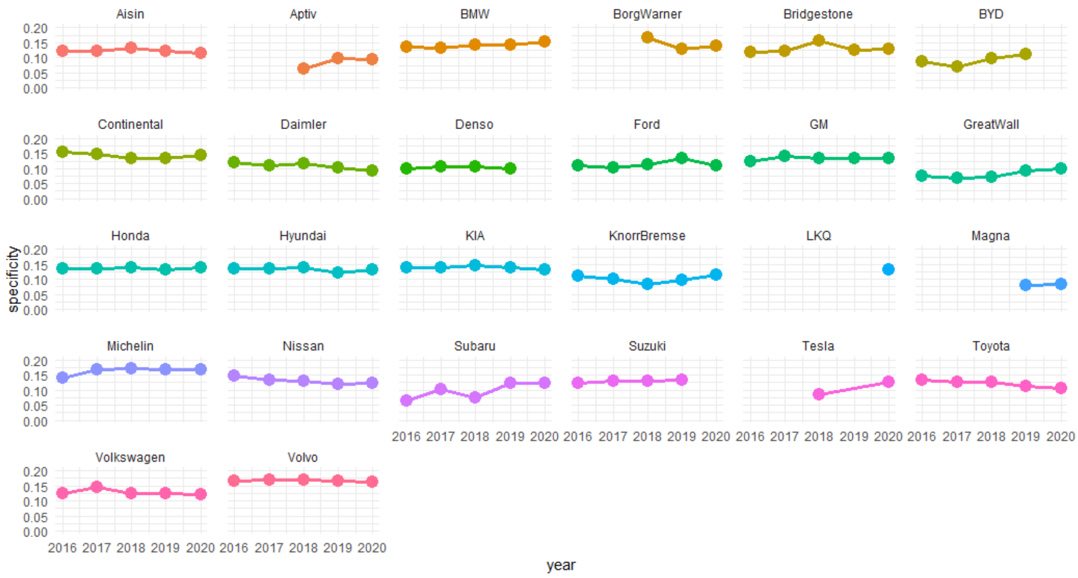


Figure 11: Scatter plots of Specificity Score on firm-level

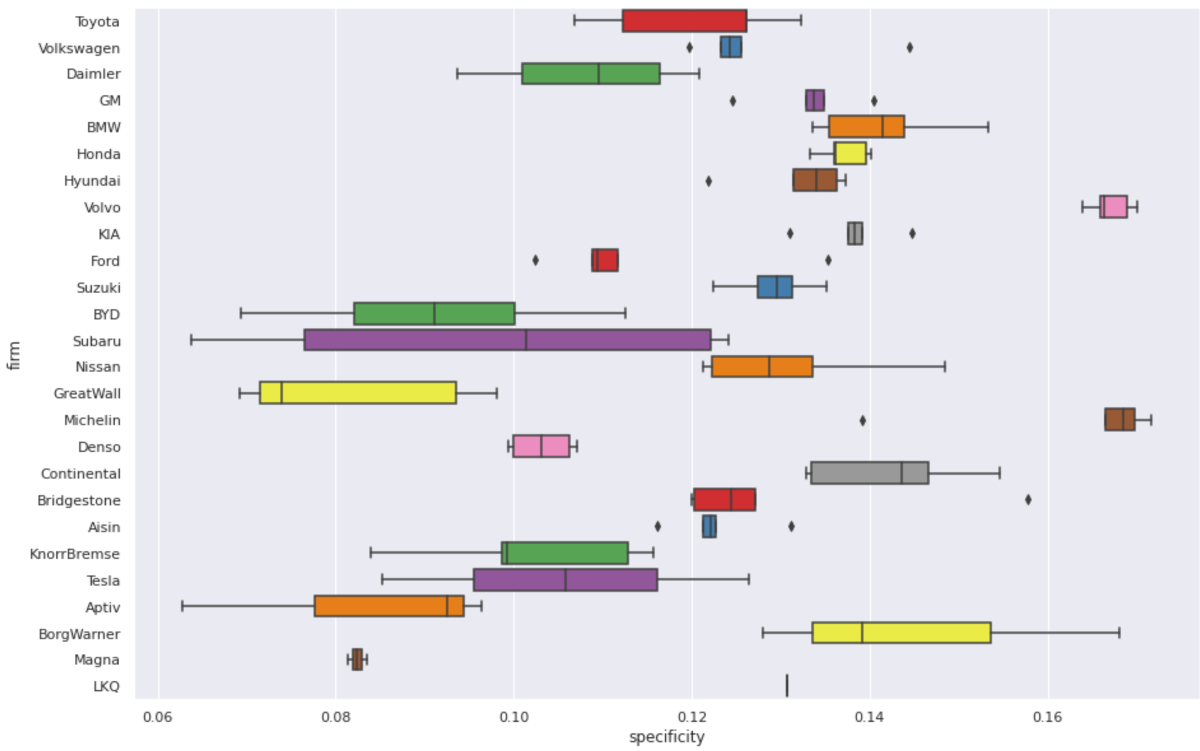


Figure 12: Boxplot of Specificity Score on firm-level

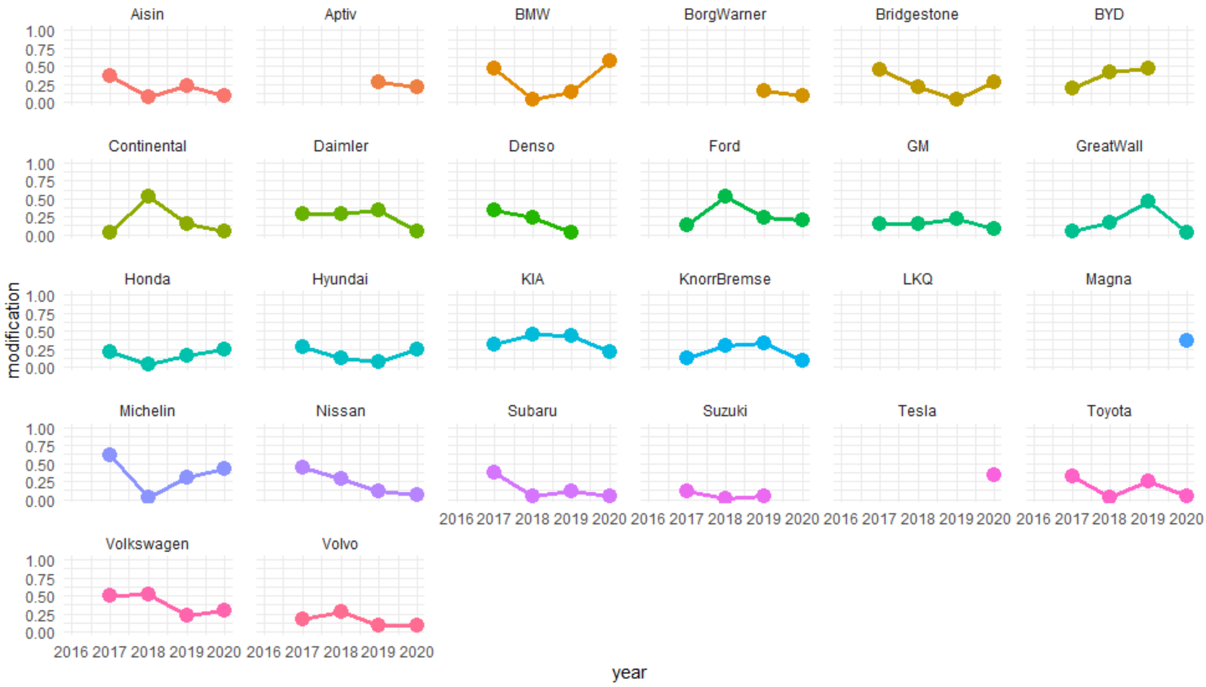


Figure 13: Scatter plots of Modification Score on firm-level

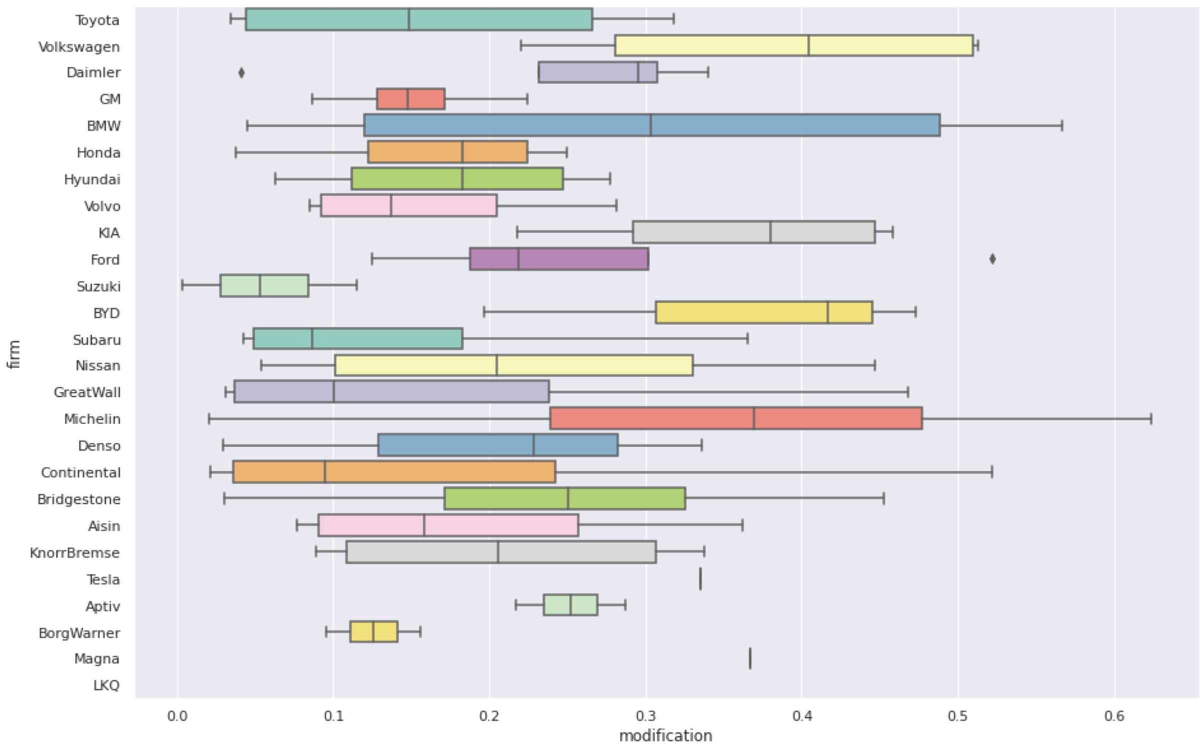


Figure 14: Boxplot of Modification Score on firm-level

Moreover, we also observed the quality measurements data grouped into five areas.

The specificity of reports disclosed by firms in China is shown at a lower level than that of reports disclosed by firms in Japan, the EU, North America, and South Korea.

Besides, the modification levels of reports disclosed by firms in these five areas almost range from 0.05 to 0.45 without great differences.

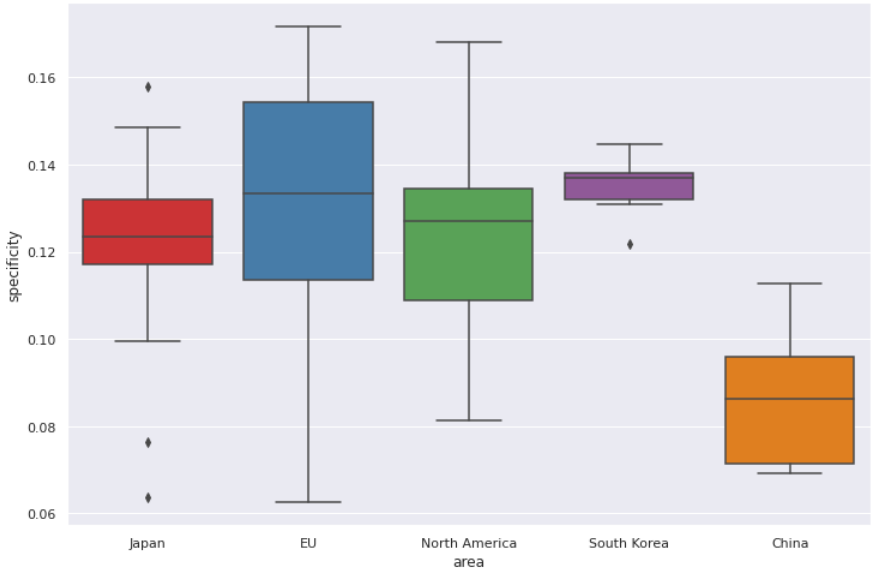


Figure 15: Boxplot of Specificity Score on area-level

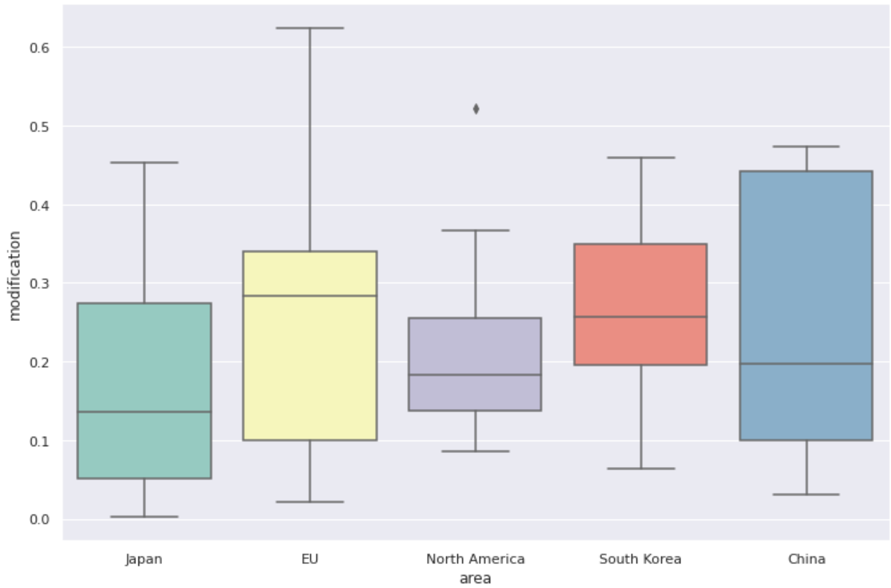


Figure 16: Boxplot of Modification Score on area-level

4.5.3 Visualization of high-dimensional vectors using T-SNE

We used a technique for dimensionality reduction called t-SNE (t-Distributed Stochastic Neighbor Embedding) (Van der Maaten et al., 2018) to have a visualization of the high-dimensional datasets generated by BERT bellowed.

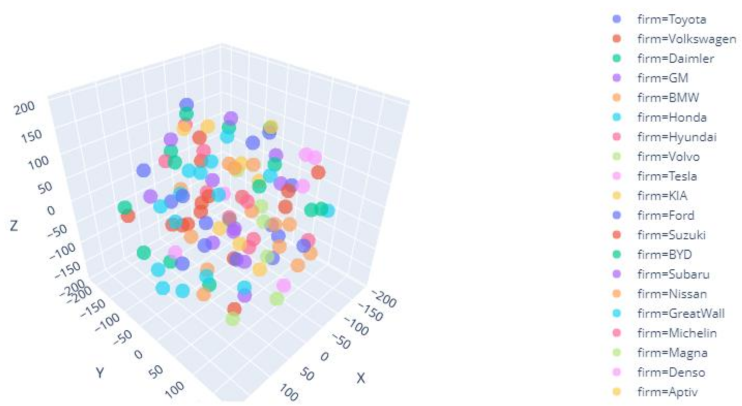


Figure 17: visualization of vectors generated by BERT on firm-level

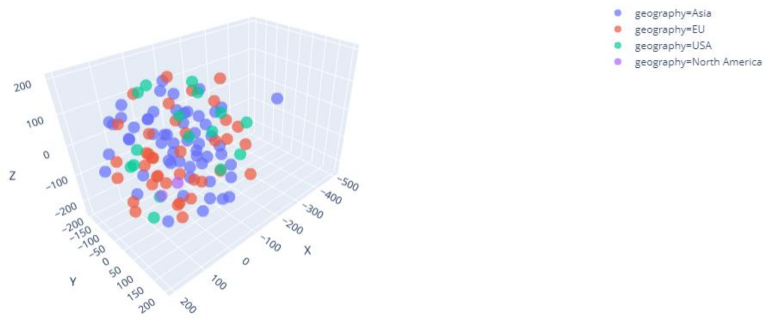


Figure 18: visualization of vectors generated by BERT on continent grouping level

The visualization of T-SNE shows us that the vector generated by BERT have different distances. From **Figure 17**, we can see that the ball in the same color, which means that the reports represented by them are disclosed by the same company, gather together at shorter distances than that with balls in other colors. It suggests that the reports disclosed by the same company share the same textual characteristics. As same as **Figure 17**, the visualization of vectors generated by BERT on continent grouping level shown in **Figure 18** represents the similarity among different continents.

4.6 Research Design

With the data collected, we design the following process to test our hypotheses.

To examine **Hypothesis1** we start with an ANOVA analysis to verify that whether the quality measurements toward non-financial disclosure differs among firms in five area groups.

Next, the linear trend estimations were used to test Hypothesis 2, investigating tendencies change with time in the quality measurements, in this study, which refers to specificity and modification performance scores. The time-series analysis in sampling periods (2016-2020) is conducted for both modification performance and specificity performance.

Then, the test of Hypothesis 3 and Hypothesis 4 are linear regressions analysis to investigate the association of quality measurements with a variety of variables, including length of sampling reports, company size (market value, turnover, profit, and asset), profitability performance (ROA, PER, profit margin ratio), and change variables.

The Linear regressions were applied to investigate the association of non-financial disclosures' quality measurements with variables listed above. All linear regressions are implemented in the lm() method in R.

Based on the ANOVA analysis result of Hypothesis 1, we will apply the LMM (Linear Mixed Model) to examine the relationship between quality measurements and dependent variables, putting the area factors into random effect.

5. Result

5.1 Result of Specificity

H1-1. The level of <u>specificity</u> of non-financial disclosure differs among firms in five area groups.

Hypothesis 1 predicts that the quality measurements differ among firms in five area groups. Therefore, we used the two textual characteristics, specificity, and modification, to proxy for quality practice and implemented ANOVA analysis to see the impact from the area factor.

Table 11: result of ANOVA analysis

> anova(lm(specificity ~ area, data=data))
Analysis of Variance Table
Response: specificity

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
area	4	0.017172	0.0042929	8.8357	3.284e-06 ***
Residuals	108	0.052473	0.0004859		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

From **Table 11**, the ANOVA revealed a significant main effect on specificity for area factor, $p<.001$. Significant (at 1%) value was starred. It indicated that the prediction of H1-1 was supported. For the non-financial disclosure practices, the level of specificity differs among firms in five area groups.

From the result of ANOVA, we knew that the performance of providing specific information through non-financial reports like sustainability reports varies in firms from different areas. Then, Based on this observation, we conducted the testing models in Hypothesis 2-1, Hypothesis 3-1, and Hypothesis 4-1 with the LMM (linear mixed model), putting the area factor into random effect. Linear mixed-effects models were performed using open-source R-package lmerTest.

H2-1. The level of <u>specificity</u> of non-financial disclosure decrease over time.

We put both the area and period variables into fixed effect and conduct the linear regression. The result in **Table 12** revealed that the trend of the first textual characteristics, specificity, had an insignificant result (p = 0.997). Prediction in Hypothesis 2 was significantly rejected with an assumption of a tendency trend in the specificity score. From the result, we can conclude that the specificity performance of sampling firms is stable in time series. We also used a linear mixed model to check the linear trend with area random effect and obtained the same insignificant result (p = 0.974).

Table 12: result of linear trend estimation (Linear regression)

lm(specificity ~ period + area, data=data)

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.564e-02	8.494e-03	10.083	< 2e-16 ***
period	-6.580e-06	1.513e-03	-0.004	0.996537
areaEU	4.623e-02	8.222e-03	5.623	1.51e-07 ***
areaJapan	3.605e-02	8.211e-03	4.391	2.67e-05 ***
areaNorth America	3.530e-02	9.117e-03	3.872	0.000186 ***
areaSouth Korea	4.956e-02	1.018e-02	4.869	3.90e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 13: result of linear trend estimation (LMM)

lmer(specificity ~ period + (1|area), data=data)

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	1.193e-01	9.576e-03	6.113e+00	12.462	1.42e-05 ***
period	4.966e-05	1.512e-03	1.071e+02	0.033	0.974

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

H3-1. The level of specificity of non-financial disclosure is negatively associated with the length of disclosure.

Based on the result of Hypothesis 1-1, we put the area factor and lg_word variable into fixed effect and conducted linear regression.

The result in **Table 14** revealed that the p-value of lg_word is near zero. Prediction in Hypothesis 3-1 was significantly supported with an assumption of firms that disclose longer reports have a better performance on providing specific content.

As shown in **Table 14**, we find that our lg_word variable is positive and statistically significant ($p<0.001$) in the result. This indicates that consistent with the understanding from the scatter plot shown, the longer the disclosures, the better performance in the contextual specificity. Therefore, we preliminarily concluded that the level of specificity of non-financial disclosure is positively associated with the length of disclosure. The contrast result of Hypothesis 3-1 was been supported.

Table 14: result of liner regression in Hypothesis 3-1

```
lm(formula = specificity ~ lg_word + area, data = data)
Residuals:
     Min          1Q        Median          3Q         Max
-0.064774    -0.010249   0.001801     0.012629   0.060957
Coefficients:
                      Estimate    Std. Error    t value     Pr(>|t|)
(Intercept)          -0.039574    0.027368    -1.446      0.15110
lg_word               0.031172    0.006606     4.719      7.21e-06 ***
areaEU                0.027680    0.008440     3.280      0.00140 **
areaJapan             0.017686    0.008422     2.100      0.03808 *
areaNorth America     0.024808    0.008520     2.912      0.00437 **
areaSouth Korea       0.032724    0.009921     3.298      0.00132 **
---
Signif. codes:   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.02015 on 107 degrees of freedom
Multiple R-squared:   0.3764,      Adjusted R-squared:   0.3472
F-statistic: 12.91 on 5 and 107 DF,   p-value: 7.872e-10
```

To further explore the relationship between the length of sampling reports (lg_word) and specificity level, we implemented a test with the linear mixed model (LMM) in which the area factor was taken as a random effect. The result shown in **Table 15** was consistent with the result of linear regression (**Table 14**). That is, the level of specificity of non-financial disclosure is **positively associated** with the length of disclosure.

Table 15: result of LMM in Hypothesis 3-1

```
Formula: specificity ~ lg_word + (1 | area)
Number of obs: 113, groups:   area, 5
Random effects: 5 area groups
Fixed effects:
               Estimate    Std. Error        df           t value       Pr(>|t|)
(Intercept)   -0.032331    0.028664     97.399659     -1.128      0.262
lg_word        0.034271    0.006352    104.432540      5.395      4.3e-07 ***
---
Signif. codes:   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

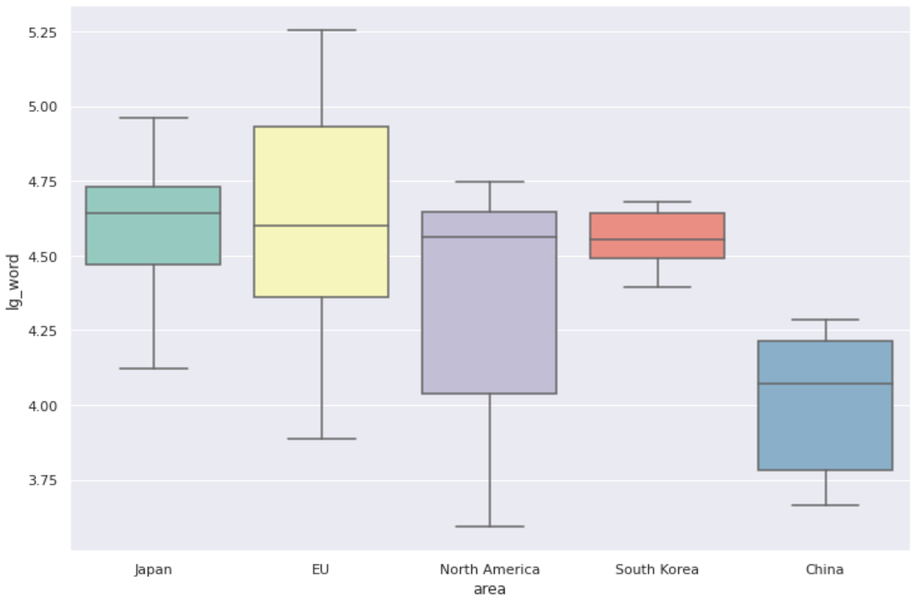


Figure 19: Boxplot of the length of reports on area-level

However, when we observed the relationship between the area factor and the length of reports, we found that the average of reports' length varies from five areas (**Figure 19**). Thus, to verify the impact of these two factors on the specificity performance, we additionally conducted an interaction test as shown in **Table 16**.

Table 16: result of Interaction of variables in Hypothesis 3-1

lm(formula = specificity ~ lg_word * area, data = data)

Residuals:

Min	1Q	Median	3Q	Max
-0.053950	-0.009475	0.001464	0.010667	0.050858

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.04879	0.11533	-0.423	0.673
lg_word	0.03347	0.02867	1.167	0.246
areaEU	-0.06919	0.12133	-0.570	0.570
areaJapan	0.25260	0.13745	1.838	0.069 .
areaNorth America	0.13348	0.12762	1.046	0.298
areaSouth Korea	0.36702	0.31496	1.165	0.247
lg_word:areaEU	0.02071	0.02981	0.695	0.489
lg_word:areaJapan	-0.05130	0.03294	-1.557	0.122
lg_word:areaNorth America	-0.02514	0.03128	-0.804	0.423
lg_word:areaSouth Korea	-0.07364	0.07041	-1.046	0.298

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.01867 on 103 degrees of freedom
Multiple R-squared: 0.4847, Adjusted R-squared: 0.4397
F-statistic: 10.77 on 9 and 103 DF, p-value: 1.232e-11

Table 17: result of ANOVA analysis in Hypothesis 3-1

anova(lm(specificity ~ lg_word * area, data=data))

Analysis of Variance Table

Response: specificity

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
lg_word	1	0.019434	0.0194341	55.7803	2.719e-11 ***
area	4	0.006777	0.0016943	4.8630	0.0012411 **
lg_word:area	4	0.007548	0.0018870	5.4161	0.0005348 ***
Residuals	103	0.035886	0.0003484		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The result of ANOVA analysis (**Table 17**) revealed a significant impact on specificity score from the interaction of length and area. P-value (0.0005348) was starred at a 1% significant level.

The interaction result shown in **Table 18** is a compilation of results from **Table 16.** For a clear understanding, we drew a straight line graph to visually check the interaction result (**Figure 19**).

Table 18: interaction result

Formula: $Specificity = \alpha + \beta_{area} * lg\_word + \varepsilon$

Area	Intercept: α	slope: β_{area}
China	0	0.03347
EU	-0.11798	0.05418
Japan	0.20381	-0.01783
North America	0.08469	0.00833
South Korea	0.31823	-0.04017

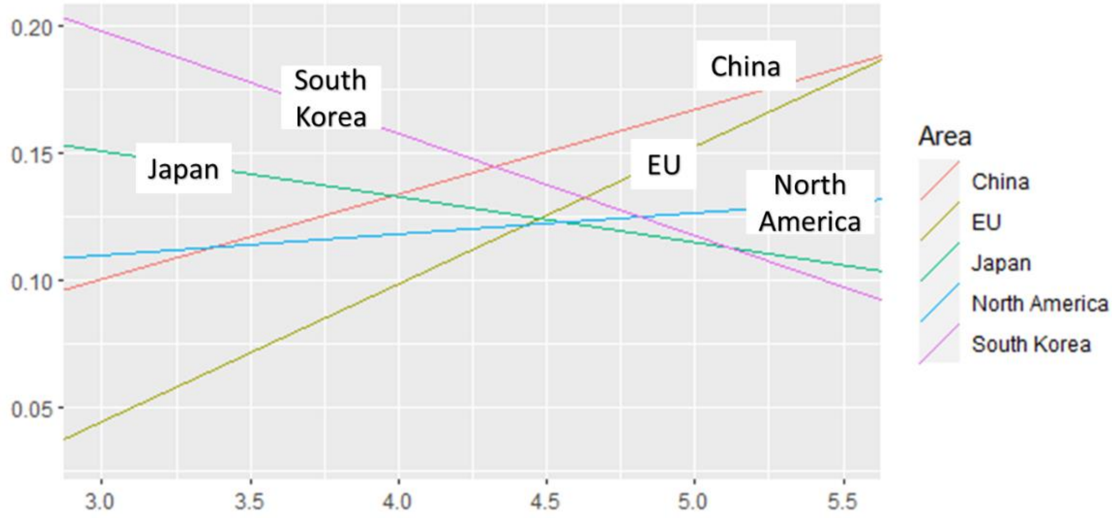


Figure 20: straight-line graph of interaction result

As shown in **Figure 20**, we found that in the cases of firms in China and firms in the EU, there is a positive correlation between the length of reports and specificity performance. It means that the longer the reports, the higher the specificity performance is. Regarding the boxplot of the length of reports on area-level (**Figure 19**), firms in China tend to disclose shorter and less specific reports while firms in the EU tend to disclose longer and more specific reports. Firms in the EU are superior performers in the quality of non-financial disclosure.

In the case of firms in North America, the specificity is weakly correlated with the length of reports. As the length of reports disclosed by firms in North America are not as long as that of reports disclosed by firms in the EU and South Korea, we can conclude that firms in North America have no attempt to use more extended reports or more specific reports to disclose their non-financial information.

When it comes to firms in Japan and firms in South Korea, they tend to use more extended reports, however, the interaction of specificity and length of reports supported the skepticism about the use of longer reports as a tool to dilute information as well as the use of sustainability reporting as a tool used to enhance reputation rather than accountability.

As discussed in the previous studies (Hummel and Schlick, 2016), voluntary disclosure theory suggested that superior sustainability performers tend to use high-quality sustainability disclosure to signal their superior performance while legitimacy theory indicated that poor sustainability performers prefer low-quality sustainability disclosure to disguise their true performance and to simultaneously protect their legitimacy.

In the previous review (Zamil et al.,2021), country-specific factors were taken into account when verifying drivers of corporate voluntary disclosure. For example, the level of economic and environmental, and social development, culture influences, legal systems, level of development in sustainability, the nature of market competition, political climates, and economic performance.

Rather than use large-length reporting to dilute the information, some of the firms in the Automotive and Parts sector who prefer the longer reports, such as firms in the EU tend to provide the higher communicative value of voluntary disclosures to users. They use the longer reporting to provide more and much specific information. Based on the analysis in previous studies, we believe that firms in the EU use these high-quality sustainability disclosures to signal their superior performance on sustainability. Meanwhile, firms in Asia should improve their performance on quality disclosing.

H4-1. There is a positive relationship between financial performance and the specificity level of non-financial disclosure.

In additional analyses, we replaced the period/lg_word variable with financial performance fixed effects. We used the value measures (i.e., company size, profitability, and change rate of financial performance) in the regressions. A p-value descriptive data (**Table 19**) were generated for all financial variables from which we can verify the significance of fixed effects from these variables.

Table 19: Result of financial variables

Variables	Pr(>|t|)	Formula
lgMV	0.264	lm(formula = specificity ~ lgMV, data = data)
lgSales	0.0178 *	lm(formula = specificity ~ lgSales, data = data)
lgProfit	0.0207 *	lm(formula = specificity ~ lgProfit, data = data)
lgAssets	0.0154 *	lm(formula = specificity ~ lgAssets, data = data)
ROA	0.171	lm(formula = specificity ~ ROA, data = data)
PER	0.159	lm(formula = specificity ~ PER, data = data)
PMR	0.902	lm(formula = specificity ~ PMR, data = data)
change_Sales	0.851	lm(formula = specificity ~ change_Sales, data = data)
change_Profits	0.838	lm(formula = specificity ~ change_Profits, data = data)
change_Assets	0.631	lm(formula = specificity ~ change_Assets, data = data)
change_MV	0.662	lm(formula = specificity ~ change_MV, data = data)

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1		

Regarding the result of simple regression analysis in **Table 19**, there is no obvious evidence to show the financial performance is relative with the specificity performance on non-financial disclosure, except for the lgSales, lgProfit, and lgAssets.

For further investigation of the correlation between these three variables and specificity, we conduct the further test as shown in **Table 20**. As we assume that there are differences in specificity performance among firms, we put the firm factor as a random effect; and include the financial variables as fixed effects in linear mixed models.

According to **Table 20**, it was found that not each company size variable but only the lgAssets variable had an obvious effect on the specificity score with firm random effect in the linear mixed model.

Table 20: LMM result of financial variables and specificity

Formula: specificity ~ lgSales + (1 | firm)

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-0.052094	0.106241	28.558039	-0.490	0.628
lgSales	0.016293	0.009955	28.478195	1.637	0.113

Formula: specificity ~ lgProfit + (1 | firm)

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	9.223e-02	4.501e-02	1.001e+02	2.049	0.0431 *
lgProfit	3.224e-03	4.810e-03	9.975e+01	0.670	0.5042

Formula: specificity ~ lgAssets + (1 | firm)

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-0.059843	0.087864	30.619987	-0.681	0.5009
lgAssets	0.016839	0.008142	30.528970	2.068	0.0472 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

As found in the part of data observation, the company size variables such as lgAssets are positively relative to the length of reports which had been considered to be much significantly correlative with specificity.

Thus, we can explain that in the same sector, large-scale companies with more employees and businesses involved. Thus, relative issues should be more so companies tend to report longer and more specific reports to convey information. In addition, in large companies, there are often holdings from institutional investors, information demand from these professional investors will drive companies to report more information to fulfill their accountability.

5.2 Result of Modification

H1-2. The level of modification of non-financial disclosure differs among firms in five area groups.

Hypothesis 1-2 predicts that the modification performance differs among firms in five area groups. Again, we implemented ANOVA analysis (**Table 21**) to see the impact from the area factor. Additionally, the firm factor was also been tested. The results were insignificant. It indicated that the prediction Hypnosis 1-2 was rejected. For the non-financial disclosure practices, there is no obvious evidence to show that the level of modification differs among firms in five area groups.

Table 21: result of ANOVA analysis

anova(lm(modification ~ area, data=data))

Analysis of Variance Table

Response: modification

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
area	4	0.13035	0.032587	1.3455	0.2602
Residuals	82	1.98591	0.024218		

anova(lm(modification ~ firm , data=data))

Analysis of Variance Table

Response: modification

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
firm	24	0.59456	0.024773	1.0094	0.4691
Residuals	62	1.52170	0.024544		

H2-2. The level of <u>modification</u> of non-financial disclosure decrease over time.

Tests in the table below are derived for Hypothesis 2-2 that a linear trend occurs in time series.

Table 22: result of linear trend estimation

```
lm(formula = modification ~ period, data = data)
Coefficients:
               Estimate   Std. Error   t value    Pr(>|t|)
(Intercept)    0.33047    0.05501      6.008      4.52e-08 ***
period        -0.02945    0.01487     -1.981      0.0508 .
---
Signif. codes:   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the result of linear trend testing in modification performance, significant (at 10%) values are pointed out, showing that the time tendency of the modification hypothesis was weakly supported. As the p-value is near 0.05 (p = 0.0508).

Besides, the coefficient of linear regression in modification performance for the time is a minus result (-0.02945), which means the modification of non-financial disclosure decrease over time. In other words, poorer practices of sampling reports occur in the sampling period.

H3-2. The level of <u>Modification</u> of non-financial disclosure is negatively associated with the length of disclosure.

We conducted a linear regression to test Hypothesis 3-2. The result in **Table 23** revealed that the p-value of lg_word is insignificant (p=0.89). Prediction in Hypothesis 3-2 was rejected with an assumption of firms who disclose longer reports will seldom modify much in reports. In other words, there is no obvious evidence to show a significant correlation between length of reports and modification performance. Hence, Hypothesis 3-2 was rejected.

Table 23: result of linear regression in Hypothesis 3-2

```
lm(formula = modification ~ lg_word, data = data)
Residuals:
     Min         1Q        Median        3Q        Max
  -0.22211    -0.13644    -0.01004     0.10040   0.40186
Coefficients:
               Estimate    Std. Error   t value    Pr(>|t|)
(Intercept)    0.261312    0.250784     1.042      0.30
lg_word       -0.007623    0.054855    -0.139      0.89
```

H4-2. There is a positive relationship between financial performance and the modification level of non-financial disclosure.

The following models were conducted, however, no obvious evidence to show the financial performance is relative to the modification practice on non-financial disclosure. Hence, Hypothesis 4-2 was rejected.

Table 24: Result of test for Hypothesis 4-2

Variables	**Pr(>|t|)**	**Formula (modification ~ Financial variables)**
lgMV	0.515	lm(formula = modification ~ lgMV, data = data)
lgSales	0.659	lm(formula = modification ~ lgSales, data = data)
lgProfit	0.858	lm(formula = modification ~ lgProfit, data = data)
lgAssets	0.612	lm(formula = modification ~ lgAssets, data = data)
ROA	0.64	lm(formula = modification ~ ROA, data = data)
PER	0.623	lm(formula = modification ~ PER, data = data)
PMR	0.387	lm(formula = modification ~ PMR, data = data)
change_Sales	0.16	lm(formula = modification ~ change_Sales, data = data)
change_Profits	0.473	lm(formula = modification ~ change_Profits, data = data)
change_Assets	0.59	lm(formula = modification ~ change_Assets, data = data)
change_MV	0.695	lm(formula = modification ~ change_MV, data = data)

5.3 Sum of Hypotheses and Result

Table 25: Sum of Hypotheses and Result

Hypotheses	Description	Result
Hypothesis 1-1	The level of specificity of non-financial disclosure differs among firms in five area groups.	Supported
Hypothesis 1-2	The level of modification of non-financial disclosure differs among firms in five area groups.	Rejected
Hypothesis 2-1	H2-1. The level of specificity of non-financial disclosure decrease over time.	Rejected
Hypothesis 2-2	H2-2. The level of modification of non-financial disclosure decrease over time.	Supported
Hypothesis 3-1	H3-1. The level of specificity of non-financial disclosure is negatively associated with the length of disclosure.	Contrast relationship is supported
Hypothesis 3-2	H3-2. The level of modification of non-financial disclosure is negatively associated with the length of disclosure.	Rejected
Hypothesis 4-1	H4-1. There is a positive relationship between financial performance and the specificity level of non-financial disclosure.	Rejected. (Lg_Aassets variable is weakly supported)
Hypothesis 4-2	H4-2. There is a positive relationship between financial performance and the modification level of non-financial disclosure.	Rejected.

6. Discussion and conclusion

6.1 Findings Discussion

The study aimed to evaluate the quality performance of non-financial disclosure in the automotive and parts sector. In particular, we applied text-mining techniques to explore the quality measurement and conducted a cross-area comparison.

According to the statistical analysis, there is no obvious correlation between modification performance and length of disclosure, area difference. The modification performance on firms' voluntary non-financial disclosures in the automotive and parts sector is under a decreasing time trend, and firms from different areas share the same tendency of sticky reporting. Though the automotive and parts sector is under a turning point, facing great changes in their operation and business environment, firms in this sector tend to simply use last year's disclosure as a template and make minor changes.

Moreover, in the aspect of providing specific information, specificity performance is significantly associated with the length of disclosure. The ANOVA analysis also suggested that the level of specificity of non-financial disclosure differs among firms in five area groups. We believe that some firms in this sector use their non-financial disclosures to provide more and much specific information, indicating that they precise the quality of voluntary non-financial disclosure and consider these disclosures as a great communication tool. Thus, this implies that firms with poor disclosure performance such as Chinese firms still require improvements in non-financial disclosures.

We picked up two sampling reports to do further observation. One is the Annual and Sustainability Report 2017 published by Volvo and the other one is Sustainable by Design 2019 published by Aptiv. The former won a high specificity score (specificity score = 0.17) in our study while the latter was recorded with a poor score (specificity score = 0.062). Information about Corporate Governance in Aptiv's report was mainly about policies and descriptions of the Board of Directors while Volvo's report used detailed meetings and events to show their governance actions and performances. For example, the organization structure figure was displayed. Also, photos that were not counted in the calculation of specificity score, were also used in Volvo's report. Our analysis indicates that some companies tend to focus more on discussing programs and initiatives than on providing specific action or performance data. Thus, we suggest that those firms with lower specificity scores should improve their reports' specificity to convey valuable community information to users of reports.

Mandatory disclosure of non-financial information is being considered by regions and countries, such as human capital information in the United States and climate-related information in France. Thus, the result about sticky reports and less specific performance in specific countries may affect regulators and standard setters in the future improvement of reporting requirements, which can lead to the implementation of quality disclosure, and improvement of valuable community between firms and stakeholders. For report users such as investors and analysts, the result also helps figure out whether the non-financial reports should be the desirable information source when they do some research or make decisions.

Additionally, results indicate that the company's size, profitability level, change of financial performance were found to be insignificantly associated with the extent of non-financial disclosure's quality.

6.2 Research Issues and Limitations

Our study does have some limitations.
First, in observational studies, there is a potential bias from the definition of specificity calculating under the study-specific data-processing method.

Besides, the vectors which were generated by BERT and then used to calculate the modification rate may not be accurate enough to represent the very large documents with similar topics.

Moreover, the research only used two quality measurements to evaluate the quality of sampling disclosures, and it ignores other quality dimensions like readability which can lead to misinterpretation.

Then, the sampling from each area is imbalanced. For example, there are only 2 companies picked up from China into the sample while 8 from Japan. Thus, careful attention needs to be taken when interpreting the results especially about China. The results should not be generalized to other companies from the same area.

Finally, the study was restricted to the global automotive and parts sector and 26 firms ranking in the Forbes list. Since the sample size was industry limited, we could not conduct a cross-industry comparison. Therefore, it is too easy to conclude whether the disclosure performance is poor or not.

6.3 For Future Study

In the future study, we recommend that more sampling from different countries/territories should be taken into consideration such as firms in Africa, Latin America, and the Middle East regions. Besides, cross-sector comparison with more quality measurements indexes should help contribute to great findings. Moreover, with the increase of the sample size, more detailed analysis may become possible. For example, observation of content related to a specific topic from each report, a study of a specific type of report (e.g. Environmental report), and comparison of companies with different roles (e.g. in the automotive and parts sector, the comparison of final assembly makers, first-tier suppliers, second-tier suppliers and third-tier suppliers) may be possibly achieved.

In addition, while the impact from the length of text files on modification scores in Brown's study (2010) was taken into consideration, impacts from lengths on vectors generated using BERT as well as those vectors-based modification scores in this paper are still unclear. Therefore, we have not conducted any length adjustment on modification scores in this study. Further study about the impact from the length of text files on modification scores using BERT is required.

Finally, even though results in this paper indicated that financial data were found to be insignificantly associated with the extent of non-financial disclosure's quality, further exploration are required. In detail, financial variables may be thought to alter the dependent or independent variables. Financial variables or other factors thought to be influential can be considered as control variables to minimize their effects on specific hypothesis testing.

Acknowledgment

This paper and the research behind it would not have been possible without the exceptional support of my supervisor, Prof. Takaki Hayashi. His enthusiasm, knowledge and exacting attention to detail have been an inspiration and encouraged me to overcome all challenges in learning about statistics and text mining. I would also like to show my gratitude to Prof. Yutaro Murakami, Prof. Hirokazu Kono and Prof. Takashi Iwamoto, for offering insightful comments and expertise that greatly assisted the research. Finally, it is with true pleasure that I acknowledge the contributions of my amazing peers in Takaki Hayashi lab, who have offered me valuable ideas, suggestions and encouragement.

Reference

[1] Arvidsson, Susanne. "Disclosure of non-financial information in the annual report: A management-team perspective." Journal of intellectual capital (2011).

[2] Tarquinio, Lara, and Stefanía Carolina Posadas. "Exploring the term "non-financial information": an academics' view." Meditari Accountancy Research (2020).

[3] Michelon, Giovanna, Silvia Pilonato, and Federica Ricceri. "CSR reporting practices and the quality of disclosure: An empirical analysis." Critical perspectives on accounting 33 (2015): 59-78.

[4] Buhr, Nola. "A structuration view on the initiation of environmental reports." Critical perspectives on Accounting 13.1 (2002): 17-38.

[5] Cho, Charles H., Dennis M. Patten, and Robin W. Roberts. "Corporate environmental financial reporting and financial markets." The Oxford handbook of business and the natural environment. 2011.

[6] Cho, Charles H., et al. "CSR disclosure: the more things change…?." Accounting, Auditing & Accountability Journal (2015).

[7] Milne, Markus J., and Rob Gray. "Future prospects for corporate sustainability reporting." Sustainability accounting and accountability. Routledge, 2010. 203-227.

[8] Gray, Rob, and I. M. Herremans. "Sustainability and social responsibility reporting and the emergence of the external social audits: The struggle for accountability." The Oxford handbook of business and the natural environment (2012): 405-424.

[9] Li, Feng. "Annual report readability, current earnings, and earnings persistence." Journal of Accounting and economics 45.2-3 (2008): 221-247.

[10] Loughran, Tim, and Bill McDonald. "IPO first-day returns, offer price revisions, volatility, and form S-1 language." Journal of Financial Economics 109.2 (2013): 307-326.

[11] Lang, Mark, and Lorien Stice-Lawrence. "Textual analysis and international financial reporting: Large sample evidence." Journal of Accounting and Economics 60.2-3 (2015): 110-135.

[12] Hope, Ole-Kristian, Danqi Hu, and Hai Lu. "The benefits of specific risk-factor disclosures." Review of Accounting Studies 21.4 (2016): 1005-1045.

[13] 金野有真. "有価証券報告書のテキスト情報が投資家に与えるインパクトの定量的分析." (2020).

[14] Paananen, Mari, Emmeli Runesson, and Niuosha Samani. "Time to clean up environmental liabilities reporting: disclosures, media exposure and market implications." Accounting Forum. Vol. 45. No. 1. Routledge, 2021.

[15] Seebeck, Andreas, and Devrimi Kaya. "The Power of Words: An Empirical Analysis of the Communicative Value of Extended Auditor Reports." European Accounting Review, forthcoming (2021).

[16] Zamil, Izdihar Abdullah, et al. "Drivers of corporate voluntary disclosure: a systematic review." Journal of Financial Reporting and Accounting (2021).
[17] Gamerschlag, Ramin, Klaus Möller, and Frank Verbeeten. "Determinants of voluntary CSR disclosure: empirical evidence from Germany." Review of Managerial Science 5.2-3 (2011): 233-262.
[18] Giannarakis, Grigoris. "The determinants influencing the extent of CSR disclosure." International Journal of Law and Management (2014).
[19] Cowen, Scott S., Linda B. Ferreri, and Lee D. Parker. "The impact of corporate characteristics on social responsibility disclosure: A typology and frequency-based analysis." Accounting, Organizations and society 12.2 (1987): 111-122.
[20] Abd Rahman, Nor Hawani Wan, Mustaffa Mohamed Zain, and Norashfah Hanim Yaakop Yahaya Al‐Haj. "CSR disclosures and its determinants: evidence from Malaysian government link companies." Social Responsibility Journal (2011).
[21] Modapothala, Jashua Rajesh, Biju Issac, and E. Jayamani. "Appraising the corporate sustainability reports–text mining and multi-discriminatory analysis." Innovations in Computing Sciences and Software Engineering. Springer, Dordrecht, 2010. 489-494.
[22] Giannarakis, Grigoris. "Corporate governance and financial characteristic effects on the extent of corporate social responsibility disclosure." Social Responsibility Journal (2014).
[23] Eliza Haverstock. "Inside The Global 2000: The Value Of The World's Largest Public Companies Soars, As Sales And Profits Falter." https://www.forbes.com/sites/elizahaverstock/2021/05/13/inside-the-global-2000-the-value-of-the-worlds-largest-public-companies-soar-as-sales-and-profits-falter/?sh=42e840ad26d4 (2021)
[24] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
[25] Varun. "Calculating Document Similarities using BERT, word2vec, and other models" https://towardsdatascience.com/calculating-document-similarities-using-bert-and-other-models-b2c1a29c9630 (2020)
[26] Weischedel, Ralph, et al. "OntoNotes release 5.0 with OntoNotes DB Tool v0. 999 beta." (2012).
[27] Hummel, Katrin, and Christian Schlick. "The relationship between sustainability performance and sustainability disclosure–Reconciling voluntary disclosure theory and legitimacy theory." Journal of accounting and public policy 35.5 (2016): 455-476.
[28] Brown, Stephen V., and Jennifer Wu Tucker. "Large‐sample evidence on firms' year‐over‐year MD&A modifications." Journal of Accounting Research 49.2 (2011): 309-346.
[29] Rizk, Riham, Robert Dixon, and Anne Woodhead. "Corporate social and environmental reporting: a survey of disclosure practices in Egypt." Social Responsibility Journal (2008).

[30] Aljifri, Khaled. "Annual report disclosure in a developing country: The case of the UAE." Advances in accounting 24.1 (2008): 93-100.
[31] Adnan, Shayuti Mohamed, David Hay, and Chris J. van Staden. "The influence of culture and corporate governance on corporate social responsibility disclosure: A cross country analysis." Journal of Cleaner Production 198 (2018): 820-832.
[32] Ali, Waris, and Jedrzej George Frynas. "The role of normative CSR-promoting institutions in stimulating CSR disclosures in developing countries." Corporate Social Responsibility and Environmental Management 25.4 (2018): 373-390.
[33] De Villiers, Charl, and Ana Marques. "Corporate social responsibility, country-level predispositions, and the consequences of choosing a level of disclosure." Accounting and Business Research 46.2 (2016): 167-195.
[34] Van der Maaten, Laurens, and Geoffrey Hinton. "Visualizing data using t-SNE." Journal of machine learning research 9.11 (2008).