

Impact of Sustainability Disclosures on Financial Performance: A Natural Language Processing Perspective

Gladie Lui
 ESCP Business School
glui@escp.eu

Jeremy Chia
 ESCP Business School
Jeremy.Jerome.Chia@edu.escp.eu

Connie Shum
 Pittsburg State University
cshum@pittstate.edu

Abstract

Sustainability reporting has emerged as an increasingly important topic since the trend linking financial and non-financial performance in the capital market attracted great attention in recent years. This study employs Natural Language Processing (NLP) to examine three characteristics of sustainability reports including major component topics, readability, and sentiments of words used. Findings from the NLP analysis are regressed against corporate financial performance measures to examine the relation between sustainability report characteristics and firm performance. Results show that specific thematic topics of sustainability reports relate to different financial performance indicators. For instance, community-related disclosures have a positive relation with Return on Assets (ROA) and with Market Value (MV), methodology-related topics have a positive relation with MV, resource-related topics have a negative relation with MV, and governance and climate-related topics both have a positive relation with the Zmijewski score (ZJS), suggesting that these factors contribute to a firm's financial viability.

Keywords: Sustainability Reports, Text Mining, Natural Language Processing, Financial Performance, Corporate Social Responsibility

1. Introduction

In the last decade, stakeholders have been pressuring firms to take more initiative with their sustainability agendas globally (UNGC 2019). This growing attention paid to sustainability issues led to a booming demand for firms' non-financial information disclosure. In the literature, treatment of sustainability overlaps with that for Corporate Social Responsibility (CSR) and Environment, Social and Governance (ESG). Even though CSR and ESG are often used interchangeably, they are different conceptually because CSR emerged from social concerns whereas corporate governance originated from agency problems. CSR focuses on accountability with a

stakeholder-based approach while ESG focuses on tracking and measuring with a risk- or performance-based approach (Apiday, 2023). The term ESG was introduced in a 2004 United Nations (UN) whitepaper *Who Cares Wins*. Since then ESG has become a ubiquitous concept. In this paper, ESG is treated as a process to meet the dynamicity of stakeholders' expectations with sustainability being the goal reflected in a business approach to creating long-term value. It is a proposed way to categorize CSR initiatives initially. Furthermore, sustainability reports include both ESG and CSR disclosures in this paper. However, with respect to the sustainability reporting landscape, in the absence of precise and binding ESG disclosure standards, firms have considerable discretion regarding the content and format of presentation. Since the motivation of this paper is to examine the shifting underlying attitudes of corporations, therefore, a firm's CSR performance remains the focus.

CSR performance is increasingly used to draw conclusions about the quality of organizations' management, identify their exposure to risk, and assess their ability to leverage business opportunities. This leads to the question: Does CSR affect firm value? In the traditional viewpoint of shareholder capitalism, maximization of shareholder wealth should be a company's primary goal (Friedman, 1970), but CSR activities hinder corporate value by incurring opportunity cost. However, from a stakeholder perspective, it is argued that companies need to be responsive to their stakeholders' needs, and the resulting improvement in firm-stakeholder relationship will be financially beneficial in the long run (Chandler, 2016; Li et al., 2018).

Sustainability reporting, a medium through which companies primarily communicate their CSR performance to stakeholders, is a complex phenomenon that cannot be explained by just a single theory (Cormier et al., 2005; Tagesson et al., 2009). This study focuses on observational inferences, yet it is necessary to maintain an openness that there may be mediator or moderator variables that define the relation between sustainability disclosure and

financial performance (Blanco et al., 2013). Nevertheless, the more prevalent theories that attempt to explain this relation are the legitimacy (Cho and Patten, 2007), stakeholder (Huang and Kung, 2010), and signaling theories (Connelly et al., 2011).

Findings of previous studies that examine the relation between CSR reports and corporate financial performance (CFP) have been equivocal. Studies linking CSR and firm performance identify several ways in which CSR creates firm value, such as lower cost of capital (Tan et al., 2020), less restricted access to funding, lower agency costs due to stakeholder engagement, less information asymmetry due to greater transparency (Cheng et al., 2014), increased customer and employee satisfaction (Servaes and Tamayo, 2013), and higher level of institutional investor ownership (Dimson et al., 2015). On the other hand, some researchers question the role of CSR and its departure from the goal of firm value maximization. For instance, Brammer et al. (2006) report that investment in CSR is destructive in terms of shareholder value. Furthermore, with a sample of Canadian companies, Richardson and Welker (2001) document a negative association between ESG disclosure and market value.

Reporting requirements have been changing rapidly. In 1997, Global Reporting Initiative (GRI), an independent global impact reporting standard setter, was formed with the support of the United Nations Environment Program. GRI's standards have a broad base of stakeholder focus—investors, employees, suppliers, customers, and community. In 2015, 17 Sustainable Development Goals (SDGs) were developed with the aim of achieving them by 2030. The SDGs were adopted by member states of the UN as a blueprint to achieve a sustainable future and they cover a wide range of dimensions such as poverty reduction, public health, food supply, and environmental sustainability. To monitor progress of achieving the SDGs, 248 key performance indicators (KPIs) are used. In addition, the International Sustainability Standards Board released exposure drafts for sustainability reporting standards with the aim of harmonizing existing standards in 2022.

With a clear trend towards mandatory sustainability reporting amidst its complicated reporting environment, the first purpose of this paper is to take stock of the status of sustainability reporting. Specifically, via Natural Language Processing (NLP) Analysis, this study aims to identify what the most prominent topics are in sustainability reports. In addition, a sentiment score and a readability score are assigned to each report to assess the sentiment and readability of these reports. The second purpose of this paper is to examine the relation between the

characteristics of these sustainability reports and CFP. To accomplish this second objective, input variables from the NLP analysis in part one are regressed against three financial performance indicators: Return on Assets (ROA), Market Value (MV), and Zmijewski Score (ZJS) (Zmijewski, 1984).

Findings from this study suggest that sustainability disclosure has positive effect on firms' performance. It is expected that this study will provide insights into how sustainability reports can be assessed, enabling stakeholders to understand what CSR factors are important so that they can encourage companies to adopt more positive sustainability initiatives amid a climate emergency.

2. Methodology

This study is a longitudinal analysis of companies' sustainability reporting for financial periods ending in 2015 to 2021 and their relation with financial performance. Singapore's capital market is chosen for this study for two reasons. First, in 2016, the Singapore Stock Exchange (SGX) introduced mandatory sustainability reporting requirements for all listed companies. Sustainability reports are required to be disclosed at the same time as financial statements. Of the 30 companies included in the Straits Time Index (STI), which is regarded as the benchmark index for the stock market in Singapore, all of them have been submitting sustainability reports since 2015. Second, as reported by a disclosure scoring study by GRI and National University of Singapore (NUS) (2022), Singapore excelled in consistently tracking and disclosing metrics across historical periods. Given that 99% of all listed companies adopted GRI, the Singaporean mandate and standardized sustainability reporting environment provide a unique setting, which aids trend analysis.

2.1. Research questions (RQs)

This study addresses the following three research questions:

RQ1: What are the most prominent topics in sustainability reports for an exploratory topic modelling analysis?

RQ2: How do those topics identified in RQ1 relate to CFP?

RQ3: How do sustainability reports' readability score and sentiment score relate to CFP?

2.2. Data sources and sample

The sample in this study consists of 206 sustainability reports from companies included in the STI covering the period 2015 to 2021. STI, which is a market capitalization-weighted index, tracks the performance of the top 30 companies listed on the Singapore Exchange (SGX). The start of the time frame is the financial period ending in 2015 when a significant number of companies listed in Singapore began publishing sustainability reports. Where the sustainability report was unavailable, the relevant section on sustainability in the annual report was extracted.

There were 206 reports in the corpus. Most companies (19 out of 30, 63%) in the STI maintained a separate sustainability report from 2015 to 2021 (observation period). Three companies (10%) continually disclosed sustainability topics in the annual report throughout the observation period. The other companies began disclosing sustainability topics as separate reports some time during the observation period. As shown in Figure 1, there was a significant increase in the average number of pages in the reports. Figure 2 reveals a similar trend in regard to number of words used in the reports. The average number of words per report tripled from 11,972 in 2015 to 33,888 in 2021.

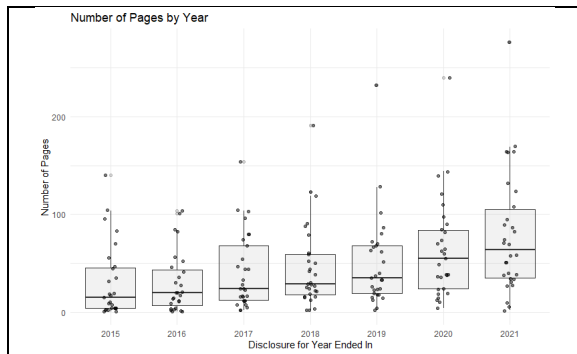


Figure 1. Number of pages in sustainability reports by financial year.

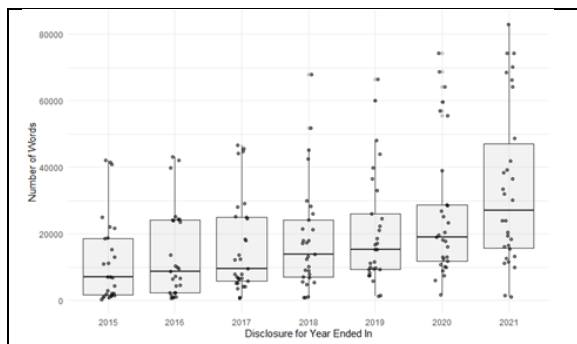


Figure 2. Number of words in sustainability reports by financial year.

Number of words is plotted against number of pages in Figure 3. It can be seen that the majority of reports tend to have a lower number of pages and words, and the distribution of pages and words is right skewed by a few companies that have an exceptionally extensive amount of reporting.

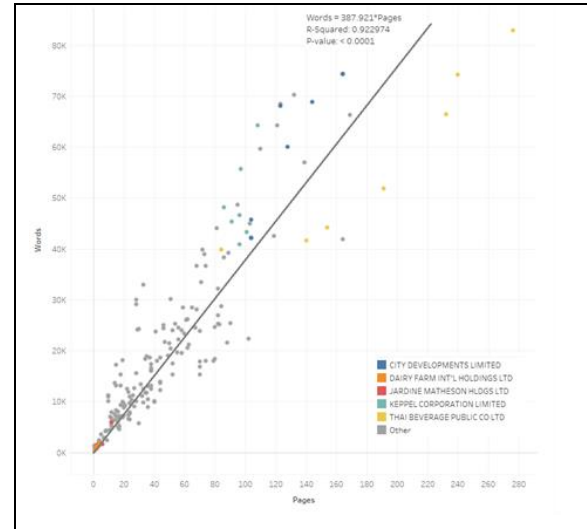


Figure 3. Number of words versus number of pages in sustainability reports.

3. Explanatory variables identification: Textual pre-processing

To address research questions RQ1, RQ2 and RQ3, several NLP methods are employed – Tokenization and Latent Dirichlet Allocation (LDA) for topic modelling, readability assessment, and sentiment analysis. NLP can be defined as a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis to achieve human-like language processing for a range of tasks or applications (Liddy, 2001). Table 1 shows the textual pre-processing steps applied to sustainability reports.

Figure 4 shows how the research questions are to be answered.

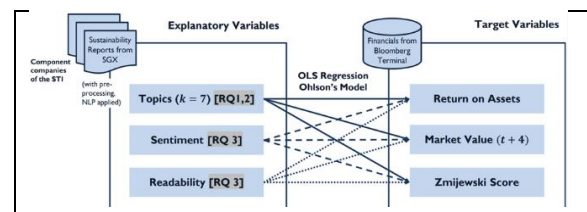


Figure 4. Data sources and methodology.

Table 1. Textual pre-processing steps applied to sustainability reports.

Tokenization	Lemmatization	Stop word removal				
<ul style="list-style-type: none"> The first step in modelling text data is to split words into token, which is a meaningful unit of text to be used. 	<ul style="list-style-type: none"> The process of grouping together inflected forms of a word to analyze as a single item. 	<ul style="list-style-type: none"> For topic modelling, to remove non-meaningful words that do not add value to the results (i.e., stop words). 				
<p>Example: Original text: “The reduction of total water consumption and overall water intensity are also key sustainability goals of the Manager.”</p>	<p>Example: ‘goal’ and ‘goals’ should be lemmatized to the root word ‘goal’ to represent one token for this analysis.</p>	<ul style="list-style-type: none"> Used stop word dictionaries provided by Lewis et al. (2004) and Feinerer et al. (2008) to remove stop words. 				
<p>As tokens: “The” “reduction” “of” “total” “water” “consumption” “and” “overall” “water” “intensity” “are” “also” “key” “sustainability” “goals” “of” “the” “Manager”</p>	<p>Preferred over stemming (a crude heuristic process that removes derivational affixes and chops off words ends) because lemmatization produces more interpretable words (Schütze et al., 2008)</p>	<ul style="list-style-type: none"> Proposed removing stop words: <table border="1"> <tr> <td>Category words</td> </tr> <tr> <td>Company name: ‘Singtel’</td> </tr> <tr> <td>Country name: ‘Singapore’</td> </tr> <tr> <td>Website and others: ‘www,’ ‘com’</td> </tr> </table>	Category words	Company name: ‘Singtel’	Country name: ‘Singapore’	Website and others: ‘www,’ ‘com’
Category words						
Company name: ‘Singtel’						
Country name: ‘Singapore’						
Website and others: ‘www,’ ‘com’						

4. Explanatory variables: Three characteristics of sustainability reports

This study explores three characteristics of companies’ sustainability reports: major topic components, sentiment, and readability.

4.1 Major topic components: RQ1 and RQ2

Using Gibbs Sampling method for Latent Dirichlet Allocation (LDA) in Topic Modelling, the first objective in addressing RQ1 is to find the balance

between having enough topics to identify key themes for separate analysis but not too many that the themes are not interpretable. To determine the optimal number of topics, “CaoJuan2009” (Cao et al., 2009) and “Deveaud2014” (Deveaud et al., 2014) are adopted in this study. Results are presented in Figure 5. A minimal value for “CaoJuan2009” is more optimal, whereas a maximal value for “Deveaud2014” is more optimal. The “Deveaud2014” metric begins to hover when the number of topics increases beyond seven. Although the “CaoJuan2009” metric still decreases, the decreases after the seventh topic are marginal compared to the initial values. The method FindTopicsNumber () from R package is used.

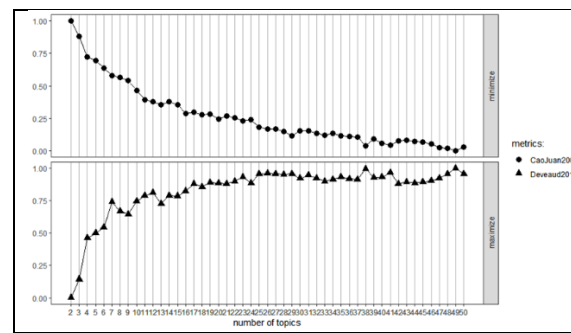


Figure 5. “CaoJuan2009” and “Deveaud2014” metric scores.

In the interest of having more interpretable topics, the lowest parameter where an increase in number of topics brings a significant improvement of the metric is seven topics and it provides the optimal trade-off between performance and interpretability. After examining the top ten words by the identified topics, the proposed interpretation of these topic components (TC) is suggested in Table 2.

The themes identified in this study align with disclosure themes required in the GRI. GRI 3 specifies that the material topics are the ones that represent an organization’s most significant impact on the economy, environment, and people, including impact on human rights. Three of the seven optimal topics identified are related to people. There are also topics on environment – resources and climate. GRI 2 contains disclosure guidelines for organizations to provide information about their reporting practices, activities, and workers. One of the seven themes identified pertains to methodology and disclosures.

Given that the GRI is adopted by 99% of listed companies in Singapore (GRI and NUS, 2022), the themes identified from topic modelling are expected to be influenced by the disclosure themes required by the GRI, which is in accordance with the findings of Kang and Kim (2022).

Table 2. Interpretation of topics.

Topic	Theme	Words
1	Social / Community	Community, Social, Support, Development, Children, Activities
2	Environment / Resources	Energy, Emissions, Waste, Water, Consumption, GHG
3	Social / Customers	Customer, Safety, Service, Security, Quality
4	Social / Employees	Employees, Work, Safety, Development, Training
5	Methodology	GRI, Material, Assurance, Information, Standards
6	Governance	Board, Corporate, Governance, Committee, Risk, Policy
7	Environment / Climate	Climate, Impact, Green, Building

As shown in Figure 6, the distribution of topics across all reports is generally consistent across all years. The variance in proportion is dependent on the company's industry. Figure 7 shows two clear examples of differing distribution between ComfortDelGro, a transportation company, and Keppel Corporation, a conglomerate.

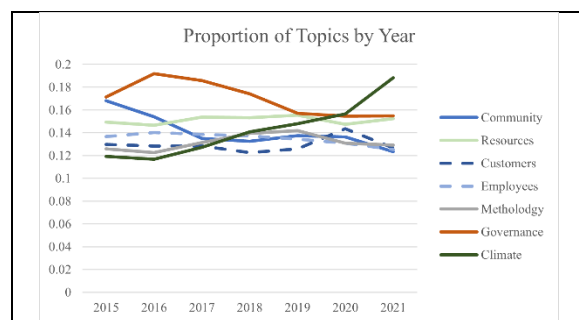


Figure 6. Distribution of topics across years.

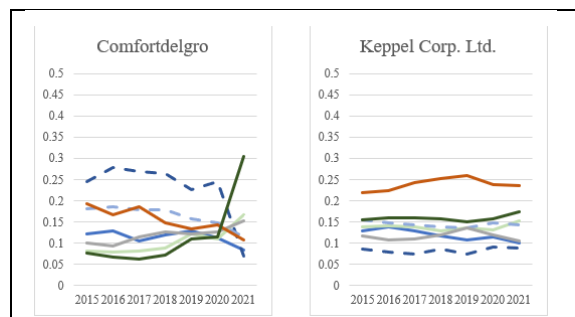


Figure 7. Distribution of topics by companies: ComfortDelGro and Keppel.

Since the implementation of climate disclosure requirements by SGX, the proportion of climate-related disclosures have increased, while 40% of content is dedicated to disclosures on social-related dimensions.

Effects of the signaling theory are eminent in the focus of each industry. Keppel Corporation has businesses in Offshore and Marine industry, which was badly plagued by a corruption scandal. That explains its emphasis on governance topics—to dispel a negative reputation. ComfortDelGro focused on customer-centric disclosures before pivoting sharply to climate-themed disclosures, indicating an intent to be sustainable.

4.2 Sentiment

Sentiment Analysis is a common technique used to detect the effectual states, emotions, and valence of texts. Taking the arithmetic average of lexicons AFINN, NRC, Bing, Jockers (General), and Loughran (Jockers) as a gauge, each report is given a sentiment score (S). The `get_sentiments()` method from R package is used.

Prior studies find that companies are generally consistent in their deployment of positive and negative words over time, keeping their messaging consistent under the signaling theory (Kang and Kim, 2022). However, when specific events occur, the "obfuscation hypothesis" can be inferred, where companies make unfavorable news more difficult to decode (Rutherford, 2003), making it worthwhile to examine the relation.

The average sentiment per page has been consistent throughout the years, although it has decreased slightly but returned to 2015 levels for the financial year ending in 2021. This concurs with Boiral's (2013) finding that more positive language is used in years when there is bad news; e.g., during the COVID-19 pandemic. Figure 8 demonstrates this pattern.

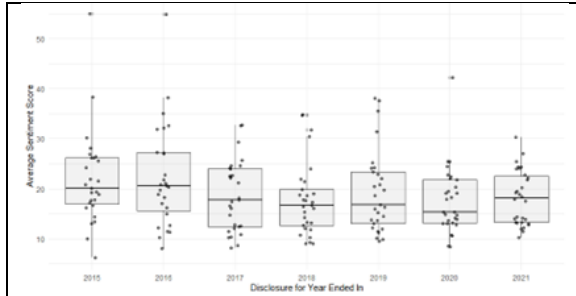


Figure 8. Average Sentiment Score, adjusted for number of pages by year.

4.3 Readability

The original Flesch Reading Ease Score (RE) (Flesch, 1948) is used for its simplicity. Other factors are not included in the mix to avoid autocorrelation in the regression analysis. Lexical density is assessed as part of these metrics. Higher lexical density is interpreted as having higher textual complexity due to a higher conceptual load. The readability () method of the R package is used.

A readability assessment is used to identify companies whose annual reports can be read more easily, which signals a safer investment (Butler and Keselj, 2009). The readability of corporate reports and disclosures may be a product of management motivation to obfuscate poor firm performance (Butler and Keselj, 2009) and even fraud (Goel et al., 2010).

Figure 9 reveals that there is a negative relation between number of words used and the readability of sustainability reports.

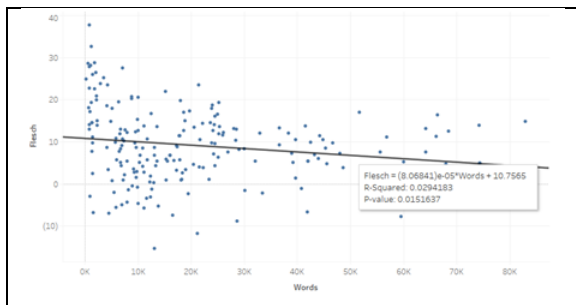


Figure 9. Readability Scores by number of words used in sustainability reports.

5. Target variables

The second objective of this study is to determine the relation between sustainability disclosures and CFP. ROA, MV, and ZJS are used to measure CFP. Financial performance information is extracted from Bloomberg Terminal and downloaded by ticker number. The ratios and standardized financial information, relied on as target variables, are derived

from annual reports filed with the Singapore Exchange.

5.1 Return on assets (ROA)

ROA is a measure of a company's ability to use its assets efficiently to generate profits. It is one of the standard metrics used by Ching et al. (2017) to assess company performance, and by Wahyuningrum et al. (2021) as an indicator of profitability when analyzing Singapore companies.

The summary statistics for ROA is presented in Table 3.

Table 3. Summary statistics for ROA.

Measure	Value
Count of Reports	200
Average ROA	4.65%
Min. ROA	-11.36%
Max. ROA	19.35%
Std. Dev. ROA	4.43%

5.2 Market value four months after the disclosure date (MV_{t+4})

MV_{t+4} reflects investors' perception of the companies' values four months after disclosure materials were released. This was the approach used by Loh et al. (2017). The summary statistics for MV_{t+4} is presented in Table 4.

Table 4. Summary statistics for MV_{t+4} .

Measure	Value
Count of Reports	200
Average MV_{t+4}	15,231
Min. MV_{t+4}	966
Max. MV_{t+4}	88,142
Std. Dev. MV_{t+4}	16,849

5.3 Zmijewski Score (ZJS)

ZJS assesses a company's risk of bankruptcy based on profitability, solvency, and liquidity. Rodgers et al. (2013) used this as a proxy for financial viability. More-positive scores indicate that companies are likely to go bankrupt. The summary statistics for ZJS is presented in Table 5.

Table 5. Summary statistics for ZJS.

Measure	Value
Count of Reports	200
Average ZJS	-1.808
Min. ZJS	-4.146
Max. ZJS	0.833
Std. Dev. ZJS	1.270

6. Models and results

This study seeks to determine the relation between sustainability disclosures and CFP. Table 6 shows the variables of interest.

Table 6. Variables used in determining the relation between sustainability disclosures and corporate financial performance.

Explanatory variables	Target variables
Topic Composition (TC) (RQ2)	Return on Assets (ROA)
Flesch Reading Ease Score (RE) (RQ3)	Market Value (MV_{t+4})
Weighted-Sentiment Score (S) (RQ3)	Zmijewski Score (ZJS)

These variables are iterated in different combinations in response to RQ2 and RQ3. To analyze MV_{t+4} , the Ohlson (1995) model is employed, which is also used by Loh et al. (2017). In addition, the Flesch Reading Ease Score (RE) and Weighted-Sentiment Score (S) are analyzed together. Ordinary least squares regression (OLS) is preferred for the analysis of ROA and ZJS.

To address RQ2 (How do the most prominent topics relate to CFP?), the following models are used:

$$\text{OLS } ROA_t = \beta_0 + \beta_1 TC_{t,2} + \beta_2 TC_{t,2} + \dots + \beta_k TC_{t,k} + \epsilon$$

$$\text{OLS } ZJS_t = \beta_0 + \beta_1 TC_{t,2} + \beta_2 TC_{t,2} + \dots + \beta_k TC_{t,k} + \epsilon$$

$$\text{Ohlson's } MV_{t+4} = \beta_0 + \beta_1 BV_t + \beta_2 EARN_t + \beta_3 EARN_t * NEG_t + \beta_4 TC_1 + \beta_5 TC_2 + \dots + \beta_{k+3} TC_3 + \epsilon$$

To address RQ3 (How do sustainability report's readability score and sentiment score relate to CFP?), the following models are employed:

$$\text{OLS } ROA_t = \beta_0 + RE_t + S_t + \epsilon$$

$$\text{OLS } ZJS_t = \beta_0 + RE_t + S_t + \epsilon$$

$$\text{Ohlson's } MV_{t+4} = \beta_0 + \beta_1 BV_t + \beta_2 EARN_t + \beta_3 EARN_t * NEG_t + \beta_4 RE_t + \beta_5 S_t + \epsilon$$

6.1 Results: Analysis of ROA (RQ2, RQ3)

Table 7 reports the results of the analysis of ROA. Results show that the extent of community-related disclosures has a positive relation with a company's ROA.

Table 7. Analysis of ROA.

	RQ 2		RQ 3		Combined	
	β	p-value	β	p-value	β	p-value
Intercept	6.33 E ⁻²	< 2 E ⁻¹⁶ (***)	5.08 E ⁻²	3.79 E ⁻⁸ (***)	8.97 E ⁻²	2.81 E ⁻¹³ (***)
S: Community	4.32 E ⁻⁶	0.000695 (***)	-	-	4.93 E ⁻⁶	0.00014 (***)
E: Resources	-4.82 E ⁻⁷	0.842	-	-	1.74 E ⁻⁷	0.942
S: Customers	-2.87 E ⁻⁶	0.099	-	-	-3.02 E ⁻⁶	0.080
S: Employees	-5.49 E ⁻⁶	0.260	-	-	-6.38 E ⁻⁶	0.190
Methodology	1.71 E ⁻⁸	0.997	-	-	-4.06 E ⁻⁶	0.426
G: Governance	-4.51 E ⁻⁸	0.976	-	-	3.53 E ⁻⁷	0.815
E: Climate	-1.54 E ⁻⁶	0.339	-	-	5.42 E ⁻⁹	0.998
Readability	-	-	9.26 E ⁻⁵	0.796	-4.97 E ⁻⁴	0.157
Sentiment	-	-	-2.65 E ⁻⁴	0.503	-9.19 E ⁻⁴	0.0323 (*)
F-Statistic, (p-value)	5.671 (5.73 E ⁻⁶) (***)		0.259 (0.772)		5.254 (2.17 E ⁻⁶) (***)	
Adjusted R ²	0.141		-0.0075		0.161	

The results support the legitimacy theory. Community-related disclosures influence society's perceptions of a company and have an empowering effect on improving the firm's financial performance (Hooghiemstra, 2000). A possible inference is that when companies focus on community initiatives, they are more recognized by Singaporean society. Therefore, corporations can use community efforts as a form of marketing. For example, Keppel Corporation makes regular donations and participates in social initiative through its philanthropic arm, Keppel Volunteers. In turn, it enjoys more efficient deployment of its capital as consumers and businesses seek to be associated with such a brand and do more business with Keppel.

There is no statistically significant relation between other disclosure areas and a company's ROA. Although this finding is not conclusive, it may suggest that other factors do not affect a company's ability to operate efficiently to produce returns for companies included in the STI. A sustainability report's readability and sentiment score does not significantly influence the company's ROA. By extension, the sustainability report, as a communication medium of the company, does not have a statistically significant relation with the company's internal operational efficiency to profitability.

6.2 Results: Analysis of MV_{t+4} (RQ2, RQ3)

Table 8 reports the results of the analysis of MV_{t+4} . Results show that companies that are more compliant are more positively perceived by investors

– evidenced by the impact of the extent of methodology of disclosure. The relation with community-related disclosure is supported by the stakeholder theory in that investors believe a company that serves its stakeholders is poised to perform better than one that does not serve its stakeholders. These findings are reflected in Singapore government’s efforts for an “all of society” approach in which firms are encouraged to contribute to social advancement.

Table 8. Analysis of MV_{t+4} .

	RQ 2		RQ 3		Combined	
	β	p-value	β	p-value	β	p-value
<i>Intercept</i>	1.06 E ⁵	0.371	6629.45	0.000627 (***)	2.38 E ⁵	0.333
<i>BV_t</i>	3.219 E ⁻¹	0.000288 (***)	0.342	0.000228 (***)	3.384 E ⁻¹	0.000185 (***)
<i>EARN_t</i>	6.037	6.92 E ¹¹ (***)	6.47	1.41 E ⁻¹¹ (***)	5.974	1.72 E ⁻¹⁰ (***)
<i>EARN_t + NEG_t</i>	-5.427	0.0201 (*)	-6.31	0.00862 (**)	-5.433	0.0203 (*)
<i>S: Community</i>	5.63 E ⁻¹	0.0418 (*)	-	-	5.60 E ⁻¹	0.0464 (*)
<i>E: Resources</i>	-1.66	0.00144 (**)	-	-	-1.65	0.00167 (**)
<i>S: Customers</i>	3.29 E ⁻¹	0.393	-	-	2.89 E ⁻¹	0.0455
<i>S: Employees</i>	-7.72 E ⁻²	0.940	-	-	-2.43 E ⁻¹	0.817
<i>Methodology</i>	2.79	0.00883 (**)	-	-	2.63	0.0188 (*)
<i>G: Governance</i>	-3.34 E ⁻¹	0.293	-	-	-32.31 E ⁻¹	0.486
<i>E: Climate</i>	-1.73 E ⁻¹	0.627	-	-	-4.51 E ⁻²	0.906
<i>Readability</i>	-	-	4.702	0.951	48.37	0.533
<i>Sentiment</i>	-	-	-183.58	0.0312 (*)	-89.06	0.346
<i>F-Statistic, (p-value)</i>	51.12 (< 2.2 E ⁻¹⁶) ***		79.46 (< 2.2 E ⁻¹⁶) ***		43.19 (< 2.2 E ⁻¹⁶) ***	
<i>Adjusted R²</i>	0.735		0.703		0.734	

A statistically significant positive relation between resources and market value is also observed. Since Singapore is a resource-limited country, which relies on the region for imports, how companies deal with the finite resources is important for shareholders’ perceptions.

Furthermore, in relation to sentiment, the more positive the words used, the lower the observed MV, suggesting that more positive words were used when negative events occurred. MV reflects investors’ perceptions of a company’s business prospects. These values are readily available for the sample companies given that they are all listed on the SGX. The operative word is ‘perceptions,’ based on presumably publicly available information. This means that disclosure documents, i.e., financial and sustainability reports, are expected to have an influence.

More disclosures on methodology used is seen with higher MV. The null hypothesis is rejected with a 0.8% statistical significance, suggesting a statistically significant positive relation between these two variables. Companies with higher MV are naturally making more extensive disclosures on

methodology used. Given their larger size, it is expected that they have more resources and more considerable ability to do so. In the context of other disclosure themes, it is important to Singapore investors that the regulator’s requirements are met. This is evident in that the coefficient for the extent of methodology-related disclosure ($\beta=2.79$) is the highest among all other topics.

In an individual model with readability and sentiment scores, the sentiment score has a statistically significant negative relation with MV -- the more positive the words used, the lower the observed MV. This finding agrees with the view Boiral (2013) holds, where more positive words are used when negative events occurred.

6.3 Results: Analysis of ZJS (RQ2, RQ3)

Table 9 reports the results of the analysis of ZJS. Results show that the more a firm discloses climate-related activities, the more financially viable it is. This is encouraging support for the modern view of using triple bottom line of profits, people, and the planet as a barometer for a company’s success.

Table 9. Analysis of ZJS.

	RQ 2		RQ 3		Combined	
	β	p-value	β	p-value	β	p-value
<i>Intercept</i>	-1.949	< 2 E ⁻¹⁶ (***)	-2.676	< 2 E ⁻¹⁶ (***)	-2.748	2.79 E ⁻¹⁵ (***)
<i>S: Community</i>	4.55 E ⁻⁵	0.201	-	-	1.86 E ⁻⁵	0.601
<i>E: Resources</i>	5.15 E ⁻⁵	0.453	-	-	2.68 E ⁻⁵	0.690
<i>S: Customers</i>	5.75 E ⁻⁵	0.243	-	-	5.97 E ⁻⁵	0.214
<i>S: Employees</i>	-2.07 E ⁻⁴	0.135	-	-	-2.01 E ⁻⁴	0.140
<i>Methodology</i>	-3.09 E ⁻⁴	0.0273 (*)	-	-	-1.72 E ⁻⁴	0.230
<i>G: Governance</i>	1.25 E ⁻⁴	0.00334 (**)	-	-	1.26 E ⁻⁴	0.00333 (**)
<i>E: Climate</i>	2.18 E ⁻⁴	3.55 E ⁻⁶ (***)	-	-	1.81 E ⁻⁴	0.000251 (***)
<i>Readability</i>	-	-	0.162	0.103	2.91 E ⁻²	0.00330 (**)
<i>Sentiment</i>	-	-	0.0372	0.000818 (***)	2.09 E ⁻²	0.0815 (.)
<i>F-Statistic, (p-value)</i>	6.474 (7.40 E ⁻⁷) ***		7.111 (0.001043) **		6.612 (3.34 E ⁻⁸) ***	
<i>Adjusted R²</i>	0.162		0.0578		0.2024	

This finding is supported by the resource theory and transaction cost economics. In recent years, companies that do more for the climate and build more resources present themselves as more responsible and are more financially viable in the long run.

Governance factors similarly have a statistically significant positive relation with ZJS. The more governance factors a company discloses, the more financially viable the firm is. If a company makes disclosures about its governance, it seems to reflect better management, which in turn leads to greater business viability.

The readability aspect of a sustainability report has the most statistically significant relation with the Zmijewski score compared to other financial metrics. While the relation is not statistically significant enough to reject the null hypothesis, it suggests that the readability of such reports has an impact on a company's long-term financial viability. The signaling theory supports this inference.

7. Limitations and future studies

Results in this study provide a basis for further studies. Performing industry differentiation in this study is not possible due to the small sample size. Future studies could use a larger sample and focus on the disclosure effects for critical industries identified by SGX that have a high impact – financial, agricultural, energy, materials and buildings, and transportation. These five industries are required by SGX to present climate-related disclosures in line with the mandate imposed by the Task Force on Climate-Related Disclosure (TCFD) in 2023

In this study, all reports across the observation period of 2015 to 2021 are treated as part of the same sample space. However, SGX launched several sustainability regulations during this period, which may offer opportunities to examine the pre- and post-implementation of the changes.

The analysis relating to MV raises questions about perceptions of Singaporean investors. There are inferences about investors' perceptions of the importance of methodology disclosures, community-related initiatives, and resource-related topics and their effect on MV. These inferences should be validated through a survey of investors.

Mediating and moderating factors influencing MV can also be another potential direction of future study.

8. Conclusions

This study proposes a robust methodology for quantifying and assessing a company's CSR disclosures by using NLP techniques applied directly to sustainability reports for companies included in the STI. Results show that specific topics of sustainability disclosure have statistically significant relation with different financial performance factors. For example, community-related disclosures have a positive relation with ROA and with MV, methodology-related disclosures have a positive relation with MV, resource-related disclosures have a negative relation with MV, and governance and climate-related disclosures both have a positive relation with the Zmijewski score, suggesting that these factors

contribute to a firm's financial viability. These observations are also supported by legitimacy theory, stakeholder theory, and signalling theory. However, there is no conclusive evidence that readability nor sentiments of words used in sustainability reports have any impact on CFP. These factors are complex, which may also include mediating and moderating variables.

The findings of this study, together with empirical evidence from other studies, suggest that there are benefits of CSR disclosures and they have positive effects on financial performance.

It is hoped that this study will provide insights into how CSR disclosures can be assessed, enabling stakeholders to understand what CSR factors are important so that they can encourage companies to adopt more of such positive initiatives.

9. References

- Apiday. <https://www.apiday.com/blog-posts/what-are-the-differences-between-corporate-social-responsibility-csr-and-environmental-social-governance-esg>
- Blanco, B., Guillamon-Saorin, E., & Guiral, A. (2013). Do non-socially responsible companies achieve legitimacy through socially responsible actions? The mediating effect of innovation. *Journal of Business Ethics*, 117, 67-83.
- Boiral, O. (2013). Sustainability reports as simulacra? A counter-account of A and A+ GRI Reports. *Accounting, Auditing & Accountability Journal*, 26(7), 1036-1071.
- Brammer, S., Brooks, C., & Pavelin, S. (2006). Corporate social performance and stock returns: UK evidence from disaggregate measures. *Financial Management*, 35, 97-116.
- Butler, M., & Keselj, V. (2009). Financial forecasting using character n-gram analysis and readability scores of annual reports. *Advances in Artificial Intelligence: 22nd Canadian Conference on Artificial Intelligence, Canadian AI 2009 Kelowna, Canada, May 25-27, 2009 Proceedings*, 22, 39-51.
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model Selection. *Neurocomputing*, (7-9), 1775-1781.
- Chandler, D. (2016). *Strategic Corporate Social Responsibility: Sustainable Value Creation*. London, England: Sage Publications.
- Cheng, B., Ioannou, I., & Serafeim, G. (2014). Corporate social responsibility and access to finance. *Strategic Management Journal*, 35(1), 1-23.
- Ching, H. Y., Gerab, F., & Toste, T. H. (2017). The quality of sustainability reports and corporate financial performance: Evidence from Brazilian Listed Companies. *Sage Open*, 7(2). <https://journals.sagepub.com/doi/epub/10.1177/2158244017712027>
- Cho, C. H., & Patten, D. M. (2007). The role of environmental disclosures as tools of legitimacy: A

- research note. *Accounting, Organizations, and Society*, 32(7-8), 639-647.
- Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, 37(1), 39-67.
- Cormier, D., Magnan, M., & Van Velthoven, B. (2005). Environmental disclosure quality in large German companies: Economic incentives, public pressures, or institutional conditions? *European Accounting Review*, 14(1), 3-39.
- Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective Latent Concept Modelling for ad hoc information retrieval. *Document Numerique*, 17(1), 61-84.
- Dimson, E., Karakas, O., & Li, X. (2015). Active Ownership. *Review of Financial Studies*, 28(12), 3225-3268.
- Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software*, 25(5), 1-54.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221-233.
- Friedman, M. (1970). The social responsibility of business is to increase its profits. *New York Times Magazine*, 32(13), 122-126.
- Global Reporting Initiative (GRI), & National University of Singapore (NUS). (2022). *Climate Reporting in ASEAN: State of Corporate Practices, July 2022*. <https://www.globalreporting.org/media/oujbt3ed/climate-reporting-in-asean-state-of-corporate-practices-2022.pdf>
- Goel, S., Gangolly, J., Faerman, S. R., & Uzuner, O. (2010). Can linguistic predictors detect fraudulent financial filings? *Journal of Emerging Technologies in Accounting*, 7(1), 25-46.
- Hooghiemstra, R. (2000). Corporate communication and impression management – New perspectives why companies engage in corporate social reporting. *Journal of Business Ethics*, 27, 55-68.
- Huang, C.-L., & Kung, F. (2010). Drivers of environmental disclosure and stakeholder expectation: Evidence from Taiwan. *Journal of Business Ethics*, 96, 435-451.
- Kang, H., & Kim, J. (2022). Analyzing and visualizing text information in corporate sustainability reports using Natural Language Processing methods. *Applied Sciences*, 12(11). <https://www.mdpi.com/2076-3417/12/11/5614>
- Lewis, D. D., Yang, Y., Russell-Rose, T., & Li, F. (2004). RCv1: A new benchmark collection for text categorization research. *Journal of Machine Learning Research*, 5, 361-397.
- Li, Y., Gong, M., Zhang, X., & Koh L. (2018). The impact of environmental, social, and governance disclosure on firm value: The role of CEO power. *The British Accounting Review*, 50(1), 60-75.
- Liddy, E. D. (2001). Natural Language Processing. In *Encyclopedia of Library and Information Science* (2nd ed.). NY: Marcel Decker, Inc.
- Loh, L., Thomas, T., & Wang, Y. (2017). Sustainability reporting and firm value: Evidence from Singapore-Listed companies. *Sustainability*, 9(11). <https://www.mdpi.com/2071-1050/9/11/2112>
- Ohlson, J. A. (1995). Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Research*, 11(2), 661-687.
- Richardson, A. J., & Welker, M. (2001). Social disclosure, financial disclosure and the cost of equity capital. *Accounting, Organizations and Society*, 26(7-8), 597-616.
- Rodgers, W., Choy, H. L., & Guiral, A. (2013). Do investors value a firm's commitment to social activities? *Journal of Business Ethics*, 114, 607-623.
- Rutherford, B. A. (2003). Obfuscation, textual complexity, and the role of regulated narrative accounting disclosure in corporate governance. *Journal of Management and Governance*, 7, 187-210.
- Servaes, H., & Tamayo, A. (2013). The impact of corporate social responsibility on firm value: The role of customer awareness. *Management Science*, 59(5), 1045-1061.
- Tagesson, T., Blank, V., Broberg, P., & Collin, S. (2009). What explains the extent and content of social and environmental disclosures on corporate websites: A study of social and environmental reporting in Swedish listed corporations. *Corporate Social Responsibility and Environmental Management*, 16(6), 352-364.
- Tan, W., Tsang, A., Wang, W., & Zhang, W. (2020). Corporate social responsibility (CSR) disclosure and the choice between bank debt and public debt. *Accounting Horizons*, 34(1), 151-173.
- United Nations Global Compact (UNGC). (2019). *The United Nations Global Compact – Accenture strategy CEO study on sustainability*.
- Wahyuningrum, I., Oktavilla, S., Putri, N., Solikhah, B., Djajadikerta, H., & Tjahjaningsih, E. (2021). Company financial performance, company characteristics, and environmental disclosure: Evidence from Singapore. *IOP Conference Series: Earth and Environmental Science*, 623. <https://iopscience.iop.org/article/10.1088/1755-1315/623/1/012065/pdf>
- Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22, 59-82.