ESCP
BUSINESS SCHOOL

BERLIN
LONDON
MADRID
PARIS
TURIN
WARSAW

A Natural Language Processing Analysis of Sustainability Reports and its Relationship with Corporate Financial Performance: Empirical Evidence from Singapore's Listed Companies

Jeremy Chia

ESCP Business School
MSc Big Data and Business Analytics, 2021/2022

Supervised by:

Associate Professor Man Ching Gladie LUI

ESCP Business School
Financial Reporting & Audit

We are in the fight of our lives, and we are losing … And our planet is fast approaching tipping points that will make climate chaos irreversible.

"We are on a highway to climate hell with our foot on the accelerator."

António Guterres
Secretary-General of the United Nations

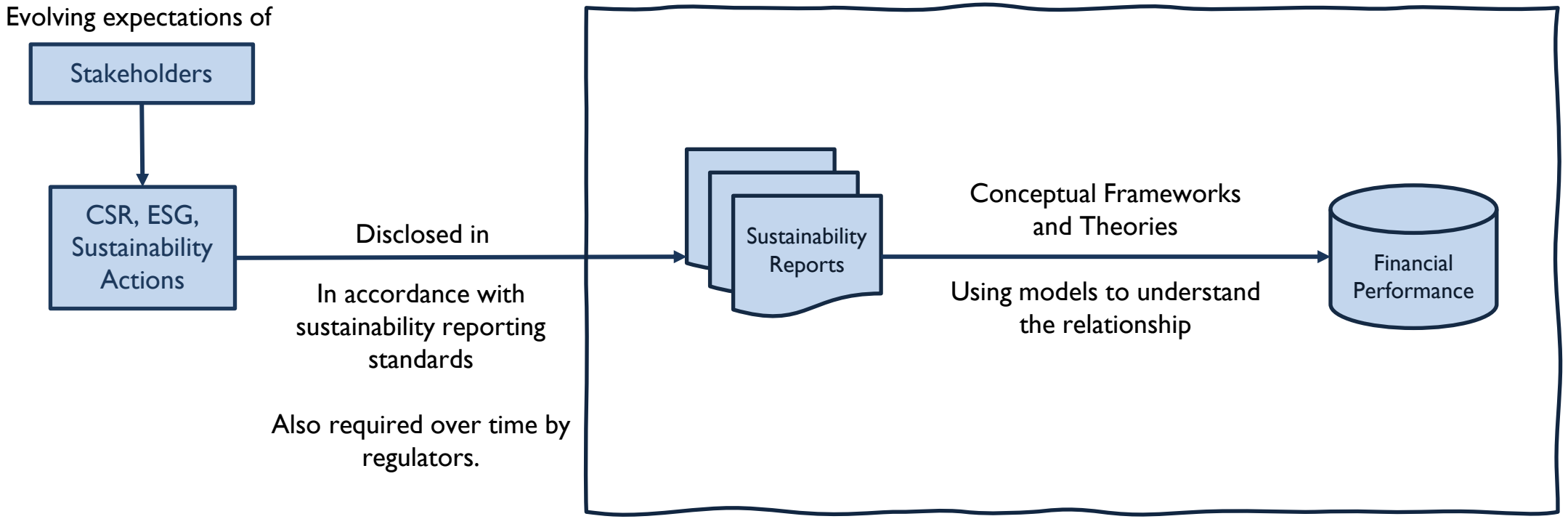
Nov 7, 2022

BERLIN
LONDON
MADRID
PARIS
TURIN
WARSAW

ESCP
BUSINESS SCHOOL

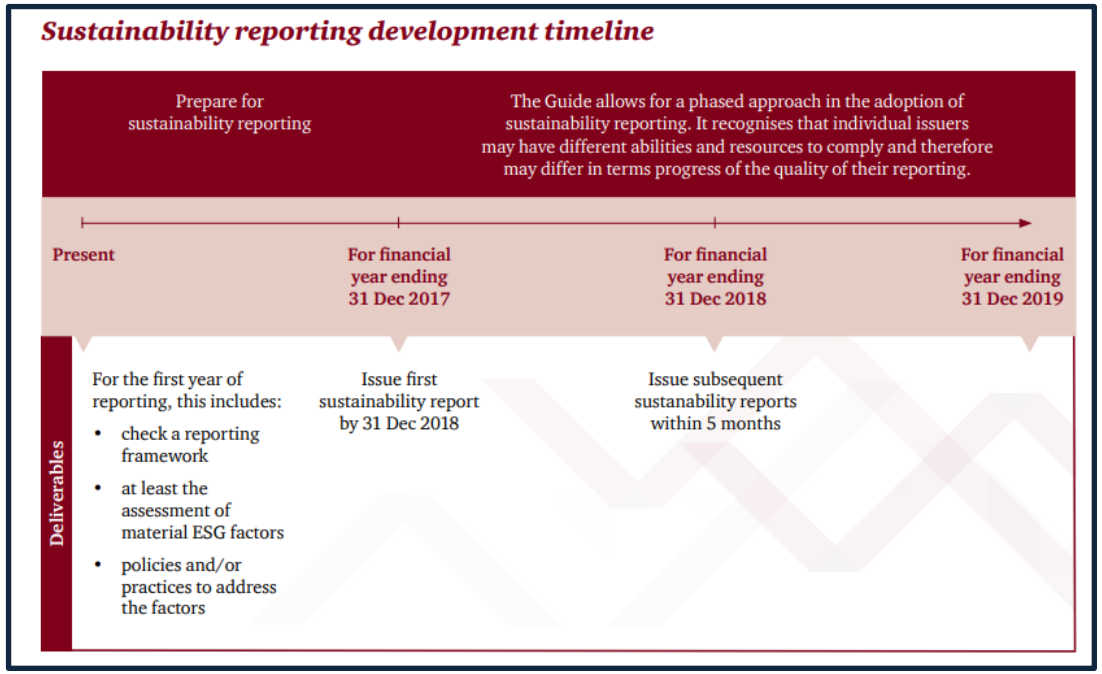
An Overview

What is this study about?

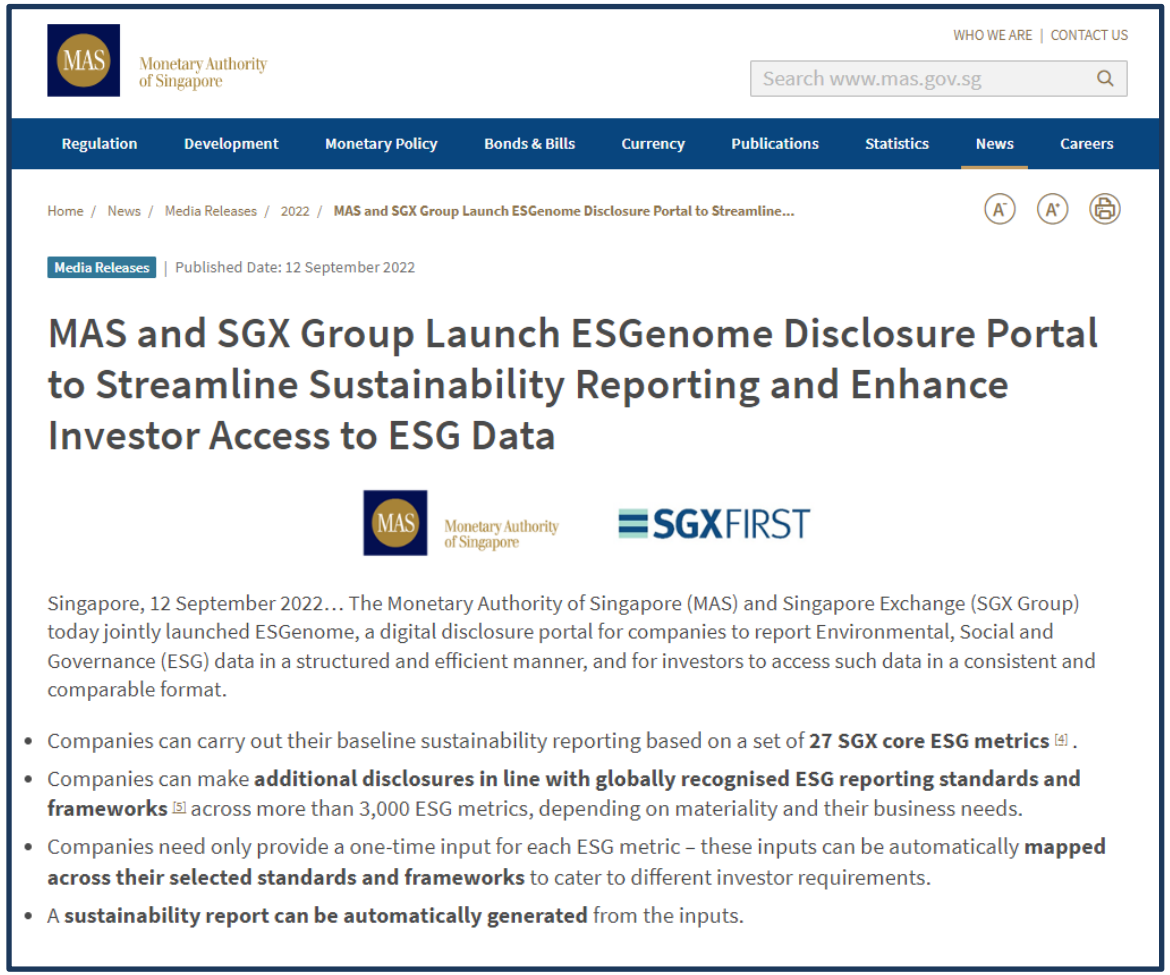


Singapore, Companies, and Disclosures

What has Singapore and what have companies in Singapore done so far?



Timeline for initial launch of sustainability reporting requirements in Singapore (PwC, 2016)



BERLIN
LONDON
MADRID
PARIS
TURIN
WARSAW

ESCP BUSINESS SCHOOL

CSR vs ESG, Past and Present

How have attitudes towards CSR, ESG changed?

1940s: companies' growth and political instability led corporations to be seen as institutions with social relevance (Heald, 1970).

1960s: CSR was a response to labour-management conflict. (Jenkins, 2009) In addition, there have been calls for responsible actions by companies concerning social aspects (Gomez-Carrasco et al., 2016).

1970s: era of "managing corporate social responsibility" (Carroll, 2015). The first Earth Day was celebrated in 1970 in response to the 1969 oil spill in Santa Barbara.

1980s: Concerns about environmental influence from human behaviour began to emerge (Salvioli, 2000). The goal of CSR policy morphed to focus on improving companies' image and reputation and securing social justification to practice (Carrol, 2008).

1990s: institutionalisation of CSR. The emerging view was that corporate ethical behaviour would increase corporate competitiveness by establishing sustainable and productive relationships with stakeholders (Jones, 1995).

2000s: Corporate policies were changing in response to the immense public interest (Smith, 2011). "Doing good" became a competitive advantage as potential investors began to view ESG initiatives as essential factors (Vibert, 2019). Change of "CSR from being a minimal commitment to becoming a strategic necessity, which can translate into a sustainable competitive advantage." (Warther & Chandler, 2005).

2010s: Acceleration of institutionalisation of CSR. More standard-setting bodies emerged for comparability. Porter and Kramar (2011) propose that – "the purpose of the corporation must be redefined as creating shared value".

Sustainability Reporting Standards

How may sustainability disclosures be standardised?

New sustainability reporting standards proposed by the IFRS Foundation – exposure drafts for standards IFRS S1 and IFRS S2 were released in mid-2022.

Governance (IFRS S1:11a, 12-13): Providing disclosures on processes, controls, and procedures used to monitor and manage sustainability-related risks and opportunities.

Strategy (IFRS S1:11b, 14-24): Company's strategy for addressing significant sustainability-related risks and opportunities and the effects of significant sustainability-related risks and opportunities on the financial position, financial performance, and cash flows.

Risk management (IFRS S1:11c, 25-26): The process by which sustainability-related risks and opportunities are identified, assessed, and managed.

Metrics and targets (IFRS S1:11d, 27-35): How the entity measures, monitors, and manages its significant sustainability-related risks and opportunities.

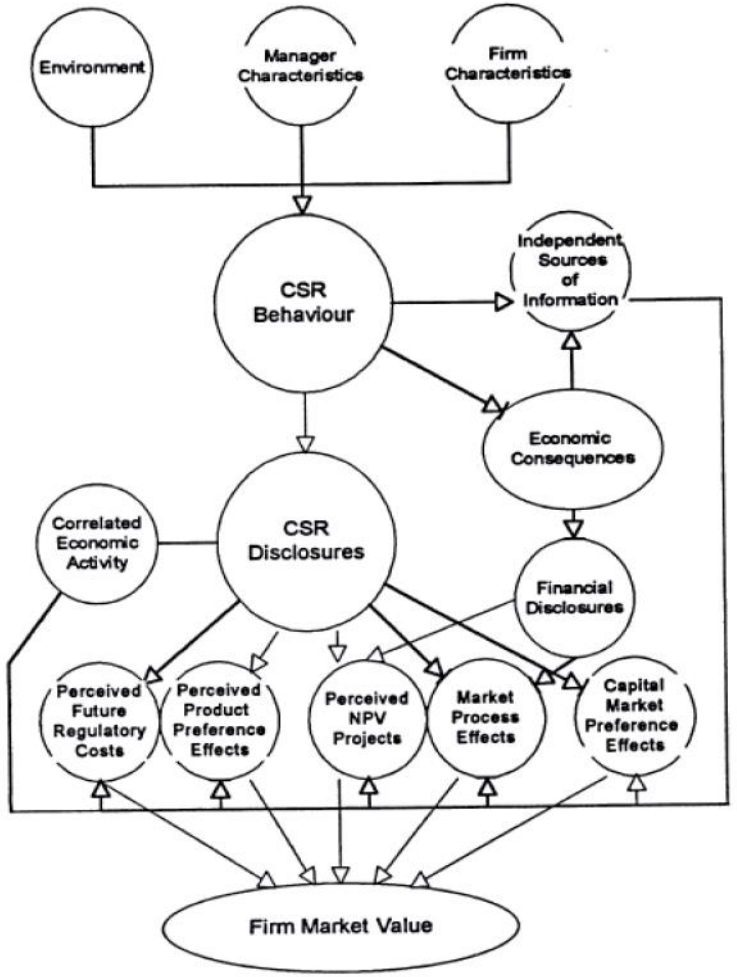
	GRI	IIRC	SASB	SDG	TCFD
Indonesia	93%	4%	16%	93%	5%
Malaysia	73%	35%	11%	74%	19%
Philippines	82%	17%	35%	86%	38%
Singapore	99%	8%	10%	65%	18%
Thailand	89%	13%	10%	95%	27%
Vietnam	65%	8%	2%	42%	0%

Adoption of Various Sustainability Reporting Frameworks for Listed Companies in Reports issued for Financial Periods ending in 2021 (GRI and NUS, 2022)

Theories in CSR and CFP

How may these two aspects of companies' be related?

- **Legitimacy theory**: Companies need to act in congruence with society to uphold their business activities. (O'Donovan, 2002).
- **Stakeholder theory**: This suggests that firms accede to the needs of stakeholders – more than merely the needs of shareholders (Ruf et al., 2001) – because of stakeholders' expectations of a firm.
- **Signalling theory**: CSR activities and ESG management can also be seen as a type of signal companies send to their stakeholders (Backhaus et al., 2002). People can react to a firm's CSR investment by seeking employment with the firm instead of just purchasing its products. (Greening and Turban, 2000).
- **Resource theories**: a company's internal resources decide the source of corporate competitive advantage (McWilliams and Siegel, 2011).
- **Transaction cost economics**: firms with good CSR perceptions have low-cost implicit claims (Cornell and Shapiro, 1987; Peloza, 2006) for lower risks involved.



A model of the capital market impacts of corporate social responsibility (Richardson et al., 1999)

Prior Approaches

How was this topic studied previously?

Explanatory Variables

- Loh et al. (2017) used a measurement scheme developed by the ASEAN CSR Network, which proposes 23 criteria in the scheme over governance, economic, environmental, and social indicators.
- Wahyuningrum et al. (2021) used an abridged form of content analysis to assign scores based on the number of disclosures – the number of sentences and the number of pages.
- Aggarwal (2013) used scores extracted from private databases.

Target Variables

- Financial metrics used were based on generally-available information.
- Return on Assets (ROA), Return on Equity (ROE), and Profit before Tax (PBT), employed in several studies (Aggarwal, 2013; Ching et al., 2017; Kasbun et al., 2016)
- Rodgers et al. (2013) and Rodgers et al. (2017) used the Zmijewski score (Zmijewski, 1984) as a proxy for financial viability.

Models Used

- Some studies chose to regress listed firms at a point in time to determine if more disclosures in the year had a positive relationship with financial performance (Kasbun et al., 2017; Aggarwal, 2013; Motwani and Pandya, 2016; Ching et al., 2017; Jones et al., 2007).
- Others conducted a longitudinal study comparing companies' performance across years (Tsang, 1998; Wahyuningrum et al., 2021).
- Loh et al. (2017) used the Ohlson model (Ohlson, 1995) as a baseline model.

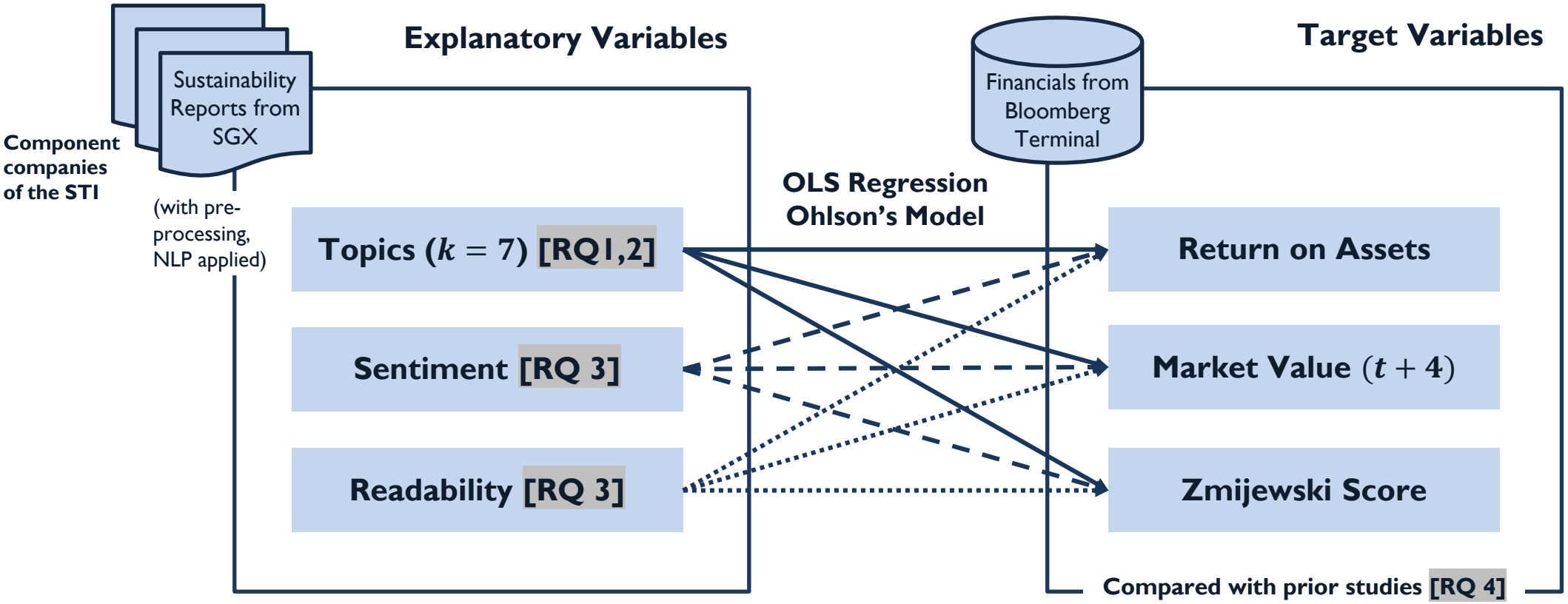
Research Questions

What are the questions this study proposes to address?

1. What are the **most prominent topics** in sustainability reports from an exploratory topic modelling analysis? Will this resemble the disclosure themes in frameworks like the GRI?
2. How do these **topics** identified in RQ 1 relate to corporate financial performance?
3. How do a sustainability report's **readability** and **sentiment** score relate to corporate financial performance?
4. Are there material **differences** in the relationship between CSR disclosures and CFP at this point, vis-à-vis **prior studies** conducted by Tsang, 1998 and Loh et al., 2017?

Data Sources and Methodology

How will the research questions be answered?



Textual Pre-processing

What were the textual pre-processing steps applied to sustainability reports?

Tokenisation

- Splitting text in sustainability reports into units for analysis – each token is one word.

Original Text:
"The reduction of total water consumption and overall water intensity are also key sustainability goals of the Manager."

As Tokens:
"The", "reduction", "of", "total", "water", "consumption", "and", "overall", "water", "intensity", "are", "also", "key", "sustainability", "**goals**", "of", "the", "Manager".

(Extracted from Page 41 of Ascendas REIT's 2019 Sustainability Report)

Lemmatisation

- For topic modelling, grouping words together based on their inflected forms using the default English lemmatisation list (Měchura, 2016).
- For example, 'runs', 'running', 'ran' are all forms of the word 'run', therefore run is the lemma of all these words.
- Preferred over stemming, because lemmatising produces more interpretable words.

Stop Word Removal

- For topic modelling, to remove non-meaningful words which do not value-add to the results (i.e. stop words).
- Used stop word dictionaries provided by Lewis et al. (2004) and Feinerer et al. (2008) to remove stop words.
- Proposed removing other stop words:

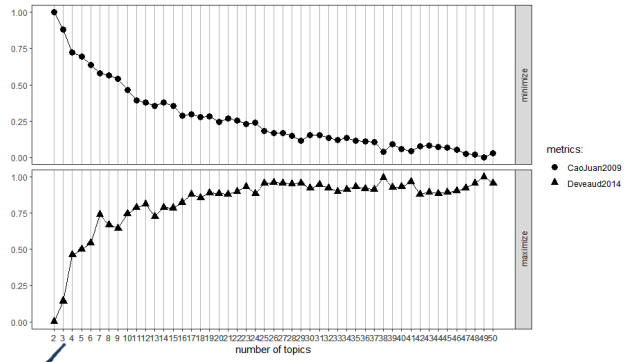
Category	Words
Company	'singtel', 'wilmar', 'cdl', 'sia'
Country	'singapore', 'taiwan', 'thailand'
Websites	'www', 'com'

Explanatory Variables

What factors are considered for sustainability reports?

Topics ($k = 7$) [RQ1]

- Determining the key themes in all sustainability reports using LDA.
- Optimising for the number of topics – balancing between separation and interpretability. (Cao et al., 2009 and Deveaud et al., 2014)



Sentiment

- Measuring the sentiment of sustainability reports using lexicons.
- Taking the arithmetic average of the following lexicons: AFINN, NRC, Bing, Jockers (General), Loughran (Jockers) as a gauge.
- Each report is given a sentiment score.
- Obfuscation hypothesis (Rutherford, 2003): companies make unfavourable to decode using more positive language.

Readability (Flesch, 1948)

- Flesch Reading Ease Score, a classical readability measure will be used.
- Longer sentence lengths (ASL) and syllables per words (ASW) decrease the readability.
- Each report is given a readability score.

$$\textbf{\textit{Readability Ease}} = 206.835 - (1.015 * ASL) - (84.6 * ASW)$$

BERLIN
LONDON
MADRID
PARIS
TURIN
WARSAW

ESCP BUSINESS SCHOOL

Target Variables

What factors are considered for financial performance?

Return on Assets

- Measure of a company's ability to use its assets efficiently to generate profits (Wahyuningrum et al., 2021).
- Higher ROAs indicate better returns on each investment dollar spent.

$$\textbf{Return on Assets} = \frac{Net\ Income}{Total\ Assets}$$

Market Value $(t + 4)$

- Market value fourth months after the disclosure date (Loh et al., 2017).
- Investors' perception of the company's valuation after disclosure materials were released and considered.

$$\textbf{\textit{Market Value}} = Share\ Price * (\#\ Total\ Outstanding\ Shares)$$

Zmijewski Score

(Zmijewski, 1984)

- Assesses the company's risk of bankruptcy based on profitability, solvency, and liquidity (Rodgers, 2017).
- Higher scores are indicative of a higher risk of defaults.

$$\textbf{\textit{Zmijewski score}} = -4.336 - 4.513 * \frac{Net\ income}{Total\ assets} + 5.679 * \frac{Total\ liabilities}{Total\ assets} + 0.004 * \frac{Current\ assets}{Current\ liabilities}$$

Model and Hypothesis

What was being tested, and how?

[RQ2] - How do the most prominent topics relate to CFP?

OLS $ROA_t = \beta_0 + \beta_1 TC_{t,2} + B_2 TC_{t,2} + \cdots + B_k TC_{t,k} + \epsilon$

OLS $ZJS_t = \beta_0 + \beta_1 TC_{t,2} + B_2 TC_{t,2} + \cdots + B_k TC_{t,k} + \epsilon$

Ohlson's $MV_{t+4} = \beta_0 + B_1 BV_t + \beta_2 EARN_t + \beta_3 EARN_t * NEG_t + (\beta_4 TC_1 + \beta_5 TC_2 + \cdots \beta_{k+3} TC_3) + \epsilon$

[RQ3] - How do a sustainability report's readability and sentiment score relate to CFP?

OLS $ROA_t = \beta_0 + RE_t + S_t + \epsilon$

OLS $ZJS_t = \beta_0 + RE_t + S_t + \epsilon$

Ohlson's $MV_{t+4} = \beta_0 + B_1 BV_t + \beta_2 EARN_t + \beta_3 EARN_t * NEG_t + (\beta_4 RE_t + \beta_5 S_t) + \epsilon$

Model and Hypothesis

What was being tested, and how?

For overall model significance (F-test):

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

$$H_1: \text{1 or more parameters} \neq 0$$

For individual variable significance (t-test):

$$H_0: \beta_p = 0$$

$$H_1: \beta_p \neq 0$$

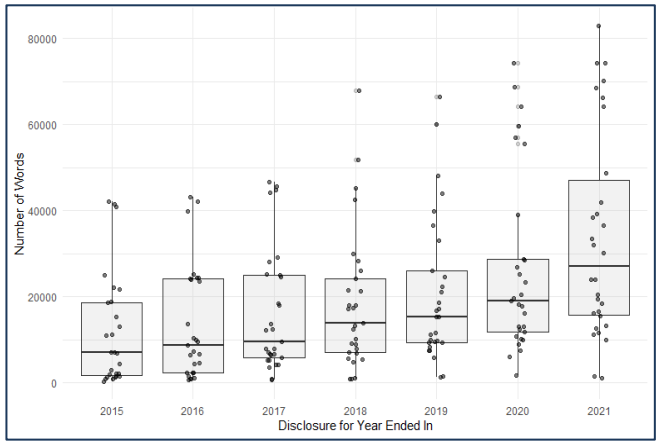
Whereas β_p represents the regression coefficients, and the significance level is 5%.

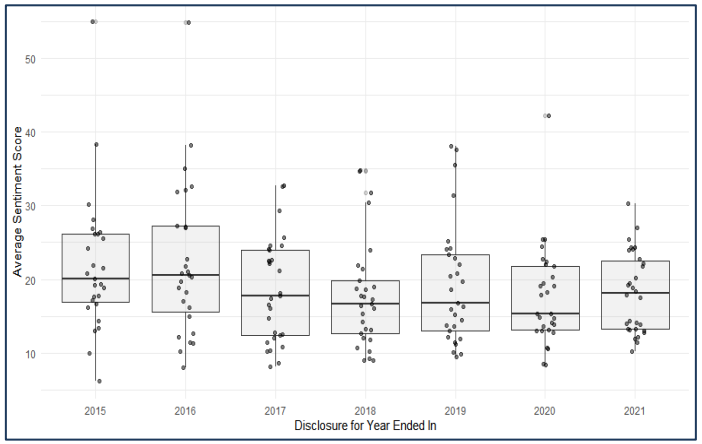
Insights from Exploratory Data Analysis

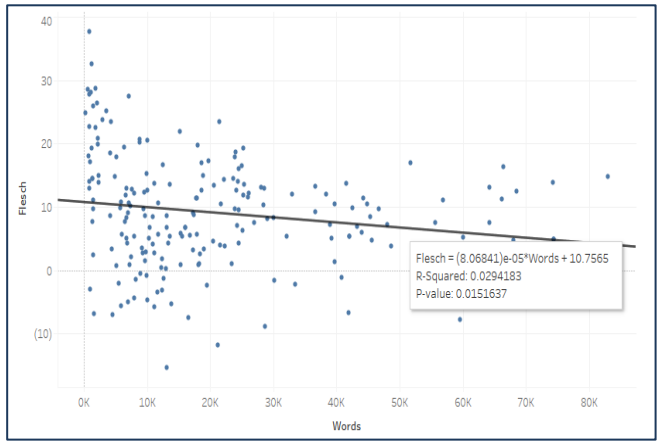
What were some trends seen from the data?

In the observation period of 2015-2021:

- The number of words, w in sustainability reports has increased. ($\overline{w}_{2015} = 11{,}972$ vs $\overline{w}_{2021} = 33{,}888$)
- Company reports' sentiment, s are generally positive ($\bar{s} = 19.3$, $\min s = 6.2$)
- Dipped towards 2020 before coming back up in 2021.
- Readability, RE, fell by 65% from 2015 to 2021. ($RE_{2015} = 14.1$, $RE_{2021} = 5.1$)
- The more words, the less readable the sustainability report.







BERLIN
LONDON
MADRID
PARIS
TURIN
WARSAW

ESCP BUSINESS SCHOOL

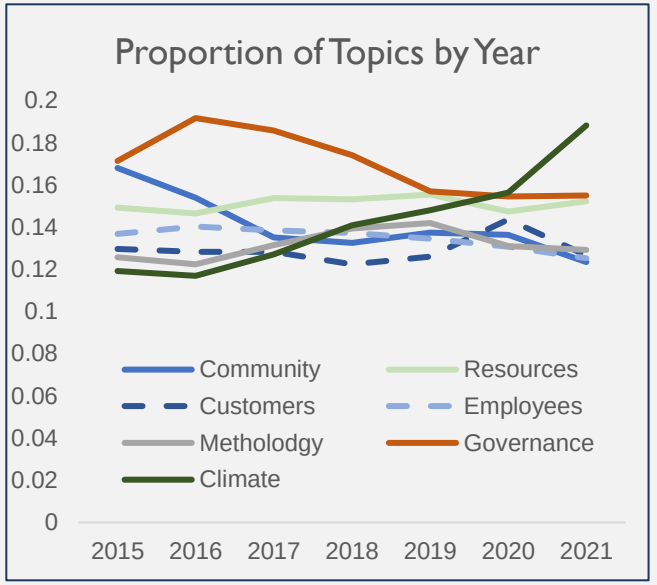
Key Topics

[RQ1]: What are the most prominent topics in sustainability reports?

In the observation period of 2015-2021:

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

- Focus in sustainability reports has shifted from governance to climate disclosures.



BERLIN
LONDON
MADRID
PARIS
TURIN
WARSAW

ESCP BUSINESS SCHOOL

Analysis on Market Value

What is a sustainability report's influence on the company's valuation by investors? [RQ2,3]

- The financial indicators currently still have the most statistically significant influence on investors' valuation of the company – this validates Ohlson's model used in Loh et al. (2017),
- Understanding this through the **legitimacy theory**, companies that are more compliant are more positively perceived by investors – evidenced by the impact of the extent of methodology disclosures has.
- The relationship on community-related disclosures is supported by the **stakeholder theory** – in that investors believe that a company which serves its stakeholders is poised to perform well, and is thus valued more.
- Sentiments: The more positive the words used were, the lower the observed market value. This tends towards the view Hrasky (2012) and Boiral (2013) held – more positive words were used when negative events occurred.

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

BERLIN
LONDON
MADRID
PARIS
TURIN
WARSAW

ESCP BUSINESS SCHOOL

Analysis on Return on Assets

What is a sustainability report's influence on the company's profitability? [RQ2,3]

- The extent of community-related disclosures, has a positive relationship with the company's return on assets.
- This result supports the legitimacy theory – that companies need to act in congruence with society to uphold their business activities (O'Donovan, 2002).
- A sustainability report's readability and sentiment score does not significantly influence the company's return on assets.
- By extension, the sustainability report, as a communication medium of the company, does not have a statistically significant relationship with the company's internal operational efficiency

	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	

BERLIN
LONDON
MADRID
PARIS
TURIN
WARSAW

ESCP BUSINESS SCHOOL

Analysis on Zmijewski Score

What is a sustainability report's influence on the company's valuation by investors? [RQ2,3]

- The extent of climate and governance disclosures had more positive disclosures have a statistically significant positive relationship with the Zmijewski score.
- This is encouraging empirical support for the modern view of the triple bottom line of profits, people, and the planet as a barometer for a company's success.
- This finding is supported by the **resource theory** and **transaction cost economics**.
- The readability aspect of a sustainability report has the most statistically significant relationship to the Zmijewski score compared to other financial metrics. However, it is still not statistically significant.

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

BERLIN
LONDON
MADRID
PARIS
TURIN
WARSAW

ESCP BUSINESS SCHOOL

Comparison with Prior Studies

What are the similarities and differences with these findings? [RQ4]

Tsang (1998):

- Analysed disclosures from 1986 to 1995 (prior to sustainability disclosures being mandated for listed companies).
- In Tsang (1998), from 1986 to 1995, community-related disclosures were more prominent than environmental disclosures. (5.37% of sentences for climate vs. 0.24% of sentences about environmental-related disclosures).
- This is vastly different in 2021, where environmental disclosures are more than twice in proportion to community-related disclosures. (~35% of sustainability reports on environmental-related disclosures vs. 15% for community-related disclosures)

Loh et al. (2017):

- Used the Ohlson's model to investigate the relationship between a sustainability reporting score and the market value.
- Both studies evidence statistically significant positive relationships between the extent and quality of sustainability disclosures and the company's market value four months after the disclosure date.
- In this study showed more specifically which topics had impacts on the market value of a company – methodology (+), community (+), and resources (-). Whereas the others did not evidence a statistically significant relationship.

Limitations of this Study

What are some opportunities for further studies?

Industry differentiation:

- Rodgers et al. (2013) and Loh et al. (2017) considers differences between industries, or characteristics of the companies (e.g. Government-Linked, Family Businesses, and others)
- Without limiting the scope to companies on the STI, a wider sample can be considered for the study to look at all companies listed on the SGX.

Time-series comparison:

- All reports (from 2015 to 2021) were treated as within the same sample space.
- More regulations were introduced progressively from 2016 by SGX – from mandatory sustainability reporting, to climate reports, and adopting the TCFD's recommendations. It is worth investigating the differences segregated by year.

Influencing factors on market value:

- This study has identified statistically significant relationships between the extent of methodology (+), community (+), and resources (-) disclosures and a company's market value, and leaves room for investigating further into why this is so.
- A qualitative study on the perception of Singapore's retail and institutional investors will be beneficial for understanding this relationship.

Causality:

- Correlation is not causation – this study shows a statistically significant relationship but does not study causation.
- There is a possibility to update the conceptual framework proposed in Richardson et al. (1999) with controlled studies.

ESCP
BUSINESS SCHOOL

BERLIN
LONDON
MADRID
PARIS
TURIN
WARSAW

A Natural Language Processing Analysis of Sustainability Reports and its Relationship with Corporate Financial Performance: Empirical Evidence from Singapore's Listed Companies

Jeremy Chia

ESCP Business School
MSc Big Data and Business Analytics, 2021/2022

Supervised by:

Associate Professor Man Ching Gladie LUI

ESCP Business School
Financial Reporting & Audit