Detecting Greenwashing Through Environmental Performance: Evidence from Corporate Disclosures and Environmental Violations in Malaysia

BY

LU LI HENG

A research project submitted in partial fulfilment of the requirement for the degree of

BACHELOR OF FINANCE (FINANCIAL TECHNOLOGY) WITH HONOURS

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF ACCOUNTANCY AND MANAGEMENT
DEPARTMENT OF FINANCE

DECEMBER 2025

LU LI HENG	GREENWASHING	BFinance (HONOURSS) FT	DECEMBER 2025

This final year project report is submitted in partial fulfillment of the requirements for the degree of Bachelor of Finance (Financial Technology) with Honours at Universiti Tunku Abdul Rahman (UTAR). This final year project report represents the work of the author, except where due acknowledgment has been made in the text.

DECLARATION

I hereby declare that:

(1) This undergraduate research project is the end result of my own work and that due acknowledgement has been given in reference to ALL sources of information be they printed, electronic, or personal.

(2) No portion of this research project has been submitted in support of any application for any other degree or qualification of this or any other university, or other institutes of learning.

(3) The word count of this research report is 21752.

Name of Student:	Student ID:	Signature:
LU LI HENG	2206129	Li Heng

Date: 16th Dec 2025

ACKNOWLEDGEMENT

I would like to sincerely express my deepest gratitude to all those who have supported me throughout the completion of this thesis.

First and foremost, I wish to convey my heartfelt appreciation to my supervisor, Dr. Lim Boon Keong, for his unwavering guidance, encouragement, and expertise. From the earliest stages of topic formulation to the final completion of this research, his mentorship has been invaluable. Despite his demanding schedule, he has always been generous with his time, offering constructive feedback, insightful comments, and patient supervision. His commitment and meticulous attention to detail have greatly enhanced the quality of this thesis, and I am deeply indebted to him for his dedication and support.

I am also profoundly grateful to my parents for instilling in me the values of hard work and perseverance, and for providing the opportunities that have shaped my academic path. Their continuous encouragement has been the foundation of my success.

In addition, I would like to extend my heartfelt thanks to my **close friends**, who have supported me with their encouragement, empathy, and companionship throughout the research process. Their presence has provided much-needed balance and motivation during difficult times.

Lastly, I would like to acknowledge the professors, lecturers, and administrative staff of the Faculty of Accountancy and Management, Universiti Tunku Abdul Rahman (UTAR). Their dedication, guidance, and the facilities they provided have enriched my academic journey and contributed to the completion of this thesis.

TABLE OF CONTENTS

Page

Copyright Page....................iii

Declaration....................iv

Acknowledgement....................v

List of Tables....................ix

List of Figures....................x

List of Abbreviations....................xi

Preface....................xiii

Abstract....................xiv

Chapter 1: Introduction....................1

1.1 Research Background....................2

1.2 Problem Statement....................7

1.3 Research Objectives....................9

1.4 Research Questions....................10

1.5 Significance of the Study....................11

1.6 Scope of the Study....................12

1.7 Limitation of the Study....................13

1.8 Organization of the Thesis....................14

Chapter 2: Literature Review....................15

2.1 Introduction....................15

2.2 Concept of Greenwashing....................16

2.3 Theoretical Framework....................19

2.3.1 Legitimacy Theory....................19

2.3.2 Signaling Theory....................21

2.3.3 Agency Theory....................23

2.4 Empirical Literature Review and Identification of Gap 25
2.4.1 Beyond the Numbers: The Reliability Crisis in ESG Ratings 25
2.4.2 Measuring Corporate Disclosure Quality: From Words to Numbers ...26
2.4.4 Empirical Findings 30
2.5 Conceptual Framework & Hypotheses Development.................................45
2.5.1 Conceptual Framework 45
Chapter 3: Methodology 47
3.1 Introduction...47
3.2 Research Design...47
3.3 Sample Selection and Data Sources...48
3.3.1 Target Population...48
3.3.2 Sample Period 48
3.4 Variable Measurement and Operationalization 50
3.4.1 Dependent Variables...50
3.4.2 Independent Variables 52
3.4.3 Control Variables...57
3.5 Formulation of Hypotheses...58
3.6 Statistical Model 58
3.6.1 Panel Multiple Regression Model...58
3.6.2 Model Equation and Variable Definitions 59
3.6.3 Software Used...61
3.7 Data Analysis Plan...61
3.7.1 Descriptive Analysis 61
3.7.2 Pearson Correlation Analysis...62
3.7.3 Diagnostics Test...63
3.8 Ethical Considerations 64
3.9 Summary...65
Chapter 4: Data Analysis 66
4.1 Introduction...66

4.2 Descriptive Analysis 66

4.3 Correlation Analysis 69

4.4 Diagnostic Tests 72

4.4.1 Hausman Test 72

4.4.2 Test for Heteroskedasticity 74

4.4.3 Test for Serial Correlation 75

4.4.5 Multicollinearity Test (Variance Inflation Factor – VIF) 77

4.4.6 Normality Test (Jarque Bera Test) 77

4.5 Model Results 78

4.5.1 Interpretation of Model Results 78

4.5.2 Justification for the Absence of Adjusted R^2 in Random Effects Estimation 81

4.6 Hypotheses Testing 82

4.7 Summary of Key Findings 83

Chapter 5: Discussion and Conclusion 85

5.1 Introduction 85

5.2 Discussion of Key Findings 85

5.3 Implications of the Findings 89

5.3.1 Theoretical Implications 89

5.3.2 Practical and Policy Implications 89

5.4 Limitations of the Findings 90

5.5 Recommendations for Future Research 92

5.6 Concluding Remarks 94

References 96

Appendices 112

LIST OF TABLES

Page

Table 2.1 Empirical Findings for E Pillar 30

Table 2.2 Empirical Findings for E Sub-pillar: Emissions 31

Table 2.3 Empirical Findings for E Sub-pillar: Resource Use 34

Table 2.4 Empirical Findings for E Sub-pillar: Innovation 36

Table 2.5 Empirical Findings for Readability Score 38

Table 2.6 Empirical Findings for Disaggregation Score 40

Table 2.7 Empirical Findings for Environmental Violation 42

Table 3.1 Dependent Variable Definition 52

Table 3.2 Independent Variables Definition 56

Table 3.3 Control Variables Definition 57

Table 3.4 Pearson Correlation Range of Coefficient 63

Table 4.1 Descriptive Analysis 67

Table 4.2 Pearson Correlation Matrix 70

Table 4.3 Multicollinearity Test Result 77

Table 4.4 Regression Results for Model 1 and Model 2 78

LIST OF FIGURES

Page

Figure 2.1 Proposed Conceptual Framework..45

Figure 3.1 Construction of Greenwashing Score..50

Figure 3.2 Construction of Readability Scores...54

Figure 3.3 Construction of Disaggregation Score...55

Figure 4.1 Hausman Test Result for Model 1..73

Figure 4.2 Hausman Test Result for Model 2..73

Figure 4.4 Breusch-Pagan-Godfrey Test Result for Model 2...............................74

Figure 4.5 White Test Result for Model 1 75

Figure 4.6 White Test Result for Model 2 75

Figure 4.7 Serial Correlation LM Test Result for Model 1 76

Figure 4.8 Serial Correlation LM Test Result for Model 2 76

Table 4.4 Regression Results for Model 1 and Model 2.......................................78

LIST OF ABBREVIATIONS

ASEAN	Association of Southeast Asia Nations
BNM	Bank Negara Malaysia
CCPT	Climate Change and Principle-Based Taxonomy
CSR	Corporate Social Responsibility
CSRD	Corporate Sustainability Reporting Directive
DQ	Disaggregation Quality
DTA	Debt-to-asset
EU	European Union
FE	Fixed Effects
FRE	Flesch Reading Ease
FTMT100	FTSE Bursa Malaysia Top 100
GLS	Generalized Least Squares
GRI	Global Reporting Initiative
GW	Greenwashing
IFRS	International Financial Reporting Standards
ISSB	International Sustainability Standards Board
JB	Jarque-Bera
LM	Lagrange Multiplier
ML	Machine Learning
MTB	Market-to-Book
NLP	Natural Language Processing
NSRF	National Sustainability Reporting Framework
OCR	Optical Character Recognition
OLS	Ordinary Least Squares
PCSE	Panel-Corrected Standard Errors
PLC	Public Listed Company
RE	Random Effects
REITs	Real Estate Investment Trusts
ROA	Return on Assets
SEC	Securities and Exchange Commission
SFDR	Sustainable Finance Disclosure Regulation
SRI	Sustainable and Responsible Investment

TCFD	Task Force on Climate-related Financial Disclosures
US	United States
VADER	Valence Aware Dictionary and Sentiment Reasoner
VIF	Variance Inflation Factor
XBRL	eXtensible Business Reporting Language

PREFACE

Environmental crises such as climate change and resource depletion have reshaped the corporate environment, making sustainability a central concern for stakeholders globally. As a result, Environmental, Social, and Governance (ESG) ratings have become important tools for investment decisions and regulatory compliance, especially in countries like Malaysia that are shifting from voluntary to mandatory sustainability reporting. However, this growing pressure to appear environmentally responsible has also increased the risk of greenwashing, where companies overstate their environmental achievements without making real operational improvements. Therefore, it is essential to examine the gap between corporate "green talk" and actual "green walk," and to investigate how environmental performance, disclosure quality, and regulatory violations influence greenwashing behaviour in the Malaysian market.

ABSTRACT

Environmental concerns such as climate change have intensified pressure on companies to demonstrate credible sustainability practices. However, this pressure has also increased the risk of greenwashing, where firms overstate their environmental commitment in disclosures without corresponding operational improvements. This study examines the determinants of greenwashing among Malaysia's top non-financial public-listed companies from 2019 to 2024. A novel Greenwashing Score is developed by comparing the positive sentiment of sustainability report narratives, derived using machine-learning techniques, with firms' actual environmental ratings. Using a sample of 82 firms and panel regression analysis, the study investigates whether environmental performance, disclosure quality, and environmental violations influence the degree of "talk–walk" decoupling.

The results show that substantive environmental performance, particularly emissions management and resource-efficient practices, significantly reduces greenwashing. In contrast, disclosure quality indicators such as readability and financial statement disaggregation, as well as recorded environmental violations, do not meaningfully explain greenwashing behaviour. Furthermore, the introduction of stricter sustainability reporting requirements in 2022 does not significantly alter these relationships. Overall, the findings indicate that genuine environmental action is the strongest predictor of credible sustainability communication in the Malaysian market. The study contributes to ESG literature by offering a replicable measure of greenwashing and providing practical insights for regulators, investors, and firms seeking to enhance sustainability reporting integrity.

Chapter 1: Introduction

This chapter established the context of the study, defined the central problem, and articulated the significance of the research. It constructed a narrative that justified the necessity of the study by moving from the global challenges of greenwashing to the specific Malaysian setting, before outlining the objectives, research questions, and scope of the work. The chapter concluded with the organization of the thesis to guide the reader through its flow.

1.1 Research Background

1.1.1 Greenwashing

The concept of greenwashing was first introduced by environmentalist Jay Westerveld in 1986. He used the term to criticise hotels that encouraged guests to reuse towels as an environmental initiative, while in reality making little or no genuine effort to reduce waste or improve their environmental practices. Since then, greenwashing has evolved into a broader phenomenon describing situations where companies overstate, exaggerate, or misrepresent their environmental credentials, creating a gap between symbolic communication and actual sustainability performance (Delmas & Burbano, 2011; Marquis & Qian, 2013). As sustainability reporting becomes more widespread, stakeholders increasingly rely on ESG disclosures to assess firms' environmental responsibility, elevating concerns that companies may strategically craft narratives to appear greener without making genuine operational improvements.

1.1.2 From Classification Risk to Enforcement Gaps: ESG Challenges in Mature and Emerging Markets

Environmental, social, and governance (ESG) ratings became fundamental to capital allocation in recent years, yet persistent gaps were observed between disclosure and substantive performance (Sklavos et al., 2025; Yu et al., 2020). In developed markets include European countries and the United States, a major challenge involves classification problems and differences between rating agencies. Because each agency uses its own method, their ESG ratings often do not match, making it difficult to compare companies fairly (Amel-Zadeh & Serafeim, 2018). To address inconsistencies in sustainability reporting, EU regulators established regulatory instruments such as the SFDR, CSRD, and the EU Taxonomy. However, cases of "classification greenwashing" still occurred. This happens when investment funds are re-labelled as ESG-friendly even though their actual portfolios did not change in any meaningful way (Kölbel et al., 2020).

The Securities and Exchange Commission (SEC) of the United States began investigating and acting against companies and investment funds that were suspected of exaggerating their climate strategies or giving misleading information about their ESG practices (SEC, 2022). These actions show that there are still weaknesses in enforcement. Some companies use positive-sounding narratives and voluntary sustainability frameworks to create the appearance of strong environmental performance, even when there is little real evidence to support it (Dyck et al., 2018). This indicates that even advanced markets can still be exposed to greenwashing when verification and monitoring systems are not strong enough.

In emerging markets, the gap between what companies report and what they do is wider, mainly because enforcement systems are weaker and environmental monitoring is less consistent. Research on Chinese companies shows a clear gap between their CSR reports and their real environmental performance, especially in industries that produce high levels of pollution (Liu & Wang, 2022). Even though many Chinese companies publish detailed CSR reports, their actual operations do not fully match what they report. This shows that the disclosures are often driven more by reputation than by real environmental action. Environmental penalties, weak project verification, and strong pressure to appear responsible all encourage companies to focus on symbolic compliance rather than real environmental improvements (Vigneau & Adams, 2023).

Overall, evidence from both developed and emerging markets shows that even though sustainability reporting frameworks are now more structured, the main problems are weak verification and enforcement. This global situation is important for understanding Malaysia, which is moving from voluntary CSR practices to more formal and stricter sustainability reporting requirements.

1.1.3 A Unique Malaysian Case: ESG Ratings, Readability, Disaggregation, and Violations

Malaysia has made good progress in strengthening ESG governance through several initiatives, including key national sustainability frameworks such as Bursa Malaysia's Sustainability Reporting Guide, the Securities Commission Malaysia's

Sustainable and Responsible Investment (SRI) Roadmap, and Bank Negara Malaysia's Climate Change and Principle-Based Taxonomy (CCPT) (Bursa Malaysia, 2022; Securities Commission Malaysia, 2019; Bank Negara Malaysia, 2022). Since 2022, Main Market listed companies have been mandated to adopt TCFD-aligned climate-related disclosures, with full implementation required by FY2025. These disclosures cover key areas such as corporate governance, business strategy, risk management, and climate-related metrics and targets. The scope of these requirements will be expanded by 2027 to include ACE Market companies as well as selected large non-listed entities (PwC, 2024). At the same time, Malaysia's regulatory framework is increasingly aligned with international standards, including the ISSB's IFRS S1 and S2, reflecting a clear transition toward globally recognised sustainability reporting practices.

Despite these advances, recent empirical work for Malaysia suggests that the credibility of sustainability communication remains uneven. Case-study evidence in the oil and gas sector identifies sustainability reports that highlight net-zero and human-rights commitments while external indicators (emissions data, regulatory actions and advertising bans) point to ongoing environmental and social controversies, raising concerns about ESG greenwashing in practice (Jamil & Wahyuni, 2024). Survey- and sector-based studies in Malaysian property development and consumer markets similarly document perceived greenwash and image-driven environmental claims (Quoquab, 2021; Shahudin et al., 2015). Although large Malaysian PLCs show limited greenwashing via readability and tone, recent evidence stresses that only some channels of manipulation have been tested and that other forms of selective or incomplete disclosure remain a concern (Kishan & Azhar, 2025). Earlier studies in the Malaysian context have similarly documented a negative relationship between the intensity of environmental disclosures and firm profitability, indicating that such reporting may be motivated more by legitimacy-seeking behaviour than by genuinely strong environmental or financial performance (Smith et al., 2007).

These dynamics make Malaysia an ideal context for developing a quantitative, firm-level greenwashing measure that integrates, first, environmental ratings, second, disclosure quality indicators such as readability and disaggregation, and last but not least, environmental violations. Understanding whether Malaysian firms' sustainability disclosures align with their environmental performance provides critical insights for policymakers, investors, and regulators during Malaysia's transition toward mandatory sustainability reporting and international ESG standards.

1.1.4 Justification for Focusing on the Environmental Ratings

While ESG refers to Environmental (E), Social (S), and Governance (G), this study focuses exclusively on the Environmental pillar. The decision is grounded in theoretical, empirical, and methodological considerations supported by past research.

Greenwashing is mainly an environmental issue. The term was first introduced to criticise companies that made misleading environmental claims (Westerveld, 1986). Modern studies also describe greenwashing as a mismatch between what a company says about its environmental efforts and what it does, not its social or governance activities (Lyon & Montgomery, 2015). Because of this, studying greenwashing must focus on whether environmental disclosures ("green talk") match real environmental performance ("green walk"). This makes the Environmental pillar the most suitable area for evaluating greenwashing.

Second, research shows that environmental claims are the most likely to be exaggerated or selectively reported compared with social or governance claims. Studies from different institutional contexts show that firms with weak environmental performance may still present highly positive environmental narratives to appeal to investors, regulators, and consumers. Greenwashing is commonly defined as a situation where a company has weak environmental performance but presents its environmental actions in an overly positive or misleading way through its communications and publicity (Delmas & Burbano, 2011). Empirical evidence from the US and Europe documents ambitious environmental policy statements that are only partially implemented (Ramus & Montiel, 2005; Hummel & Schlick, 2016), while cross-country and China-based studies find that firms with poorer environmental performance have higher probability to adopt symbolic CSR or greenwashing strategies (Marquis & Qian, 2014; Testa et al., 2018; Zhang et al., 2022)."

Taken together, these theoretical and empirical reasons justify restricting the analysis to the Environmental pillar, ensuring conceptual clarity, methodological robustness, and direct alignment with the definition and real-world manifestations of greenwashing.

1.2 Problem Statement

Although Malaysia has improved its sustainability reporting rules, recent studies show that what companies say in their sustainability reports does not always match their actual environmental performance. For example, some Malaysian firms highlight goals such as “net-zero” or present very positive environmental messages, but external evidence, like emissions data or environmental fines, suggests that their real practices are still weak (Jamil & Wahyuni, 2024; Smith et al., 2007). Other research also finds that some companies use promotional language or selective disclosure, which can mislead stakeholders and create the risk of greenwashing (Quoquab, 2021; Shahudin et al., 2015).

This issue is not unique to Malaysia. Many studies from the US, EU, and China show that even with stricter sustainability reporting rules, companies may still exaggerate their environmental achievements or present information that sounds positive but is not fully supported by their real actions. For example, Delmas and Burbano (2011) explain that companies with weak environmental performance sometimes use attractive environmental messages to improve their image. Ramus and Montiel (2005) also find that firms often publish ambitious environmental policies but do not fully carry them out in practice. In China, research by Marquis and Qian (2014) shows that some firms issue detailed CSR or environmental reports mainly to satisfy external expectations, even though their operational performance remains poor. These studies suggest that stronger reporting rules alone do not guarantee honest sustainability reporting because companies can still choose what to disclose, emphasise positive information, and leave out negative details. As a result, stakeholders may receive an incomplete or overly optimistic picture of a company’s true environmental performance.

In Malaysia, there is still no clear and combined method to determine whether a company’s environmental ratings, sustainability report quality, and environmental violations accurately represent its real environmental performance. Most existing studies only look at one aspect at a time. For example, Kishan and Azhar (2025) study only two textual features, which are readability and tone, in sustainability reports of major Malaysian public-listed companies. Quoquab et al. (2021) create a perception-based measure of greenwashing in the sustainable property sector, but they do not connect it to environmental ratings or violation records. Jamil (2024) uses a case study on a single oil and gas company by comparing its sustainability reports with external ESG controversies, but this approach cannot be applied widely across many firms. Earlier research by Smith et al. (2007) mainly examines the volume of environmental disclosures and how it relates to company characteristics, rather than using a multi-dimensional method to detect greenwashing.

This study will integrate environmental ratings, disclosure quality (readability and disaggregation), and environmental violations to examine whether companies reported environmental performance is consistent with their actual practices. By combining these three areas, this study provides a more complete way to detect possible greenwashing among Malaysian firms. Collectively, these elements serve as the foundation for the study's research objectives.

1.3 Research Objectives

To answer these questions, the study will pursue the following objectives:

RO1: To evaluate the extent to which the environmental sub-pillars (Emissions, Resource Use, and Innovation) of Environmental scores reflect substantive sustainability performance, or whether they are associated with greenwashing behavior.

RO2: To assess how the quality of sustainability disclosures, measured through textual readability and the level of disaggregation, can relate to greenwashing

RO3: To investigate the influence of environmental violations on greenwashing, and whether regulatory breaches correlate with inconsistencies between reported environmental performance and actual practices.

1.4 Research Questions

To address concerns regarding information credibility in Asian ESG markets, this thesis formulates a central research question, accompanied by several subsidiary questions that break the issue down into empirically testable components.

RQ1: Do higher environmental pillar and its sub-pillar scores from Refinitiv reflect lower greenwashing?

RQ2: How does the quality of sustainability disclosures, measured in terms of textual readability and the level of disaggregation, influence the likelihood of greenwashing in corporate sustainability reporting?

RQ3: What is the relationship between environmental regulatory violations and greenwashing behavior, and do firms with such violations exhibit greater inconsistencies between reported sustainability performance and actual environmental outcomes?

1.5 Significance of the Study

This study is important because it develops a clear and measurable Firm-year Greenwashing Score that compares what companies communicate in their sustainability reports with their actual environmental ratings. The "talk" component is captured using machine learning (ML), which scans and analyses PDF sustainability reports to detect the level of positive tone in the text. This allows the study to quantify how optimistic a company's environmental disclosures are, based on automated and replicable methods.

By comparing this tone with the environmental pillar (E-pillar) score, the study provides a practical way to detect whether companies' claims match their real performance. This is an area that is still underexplored in Malaysia. The research also breaks down the Environmental pillar into emissions, resource use, and innovation, helping to show which dimensions reflect genuine performance and which may be linked to greenwashing.

The study further improves on past research by including readability (how easy the report is to understand) and disaggregation (how detailed the data is). These indicators give a more complete picture of disclosure quality, beyond relying on only one metric. The addition of environmental violations also serves as an objective check on whether companies' reports are consistent with their real-world behaviour.

The data used in the study come from publicly available sources such as Refinitiv ESG and Bursa Malaysia, and the ML, text-processing, and coding procedures are fully transparent and repeatable. This means that other researchers can apply the same method to future samples or even other ASEAN countries. The approach is also compatible with new reporting standards like ISSB and TCFD, making the study a useful reference for regulators, investors, and researchers aiming to reduce greenwashing and improve the reliability of sustainability reporting.

1.6 Scope of the Study

This study focuses on firm-level greenwashing among the Top 100 non-financial companies listed on Bursa Malaysia. It examines whether the environmental information companies disclose (symbolic signals) is consistent with their actual environmental performance (substantive outcomes). The unit of analysis is the firm-year, covering FY2019 to FY2024, which includes the years before and after Malaysia strengthened its sustainability reporting rules in 2022. The initial sample consists of the FTSE Bursa Malaysia Top 100 (FTMT100) companies based on market capitalisation in September 2025, excluding financial institutions and REITs. After data cleansing and removing firms with missing information, 82 companies remain in the final sample. The study includes companies from industries with significant environmental impacts such as energy, plantations, industrials, consumer sectors, and telecommunications.

The data are taken from three sources. Firstly, the environmental pillar (E) and sub-pillars such as emissions, resource use, and innovation are taken from Refinitiv. Secondly, annual and sustainability reports downloaded from companies' websites and Bursa announcements provide text needed to compute the sentiment (PositiveTone), readability (e.g., Flesch–Kincaid or Flesch Reading Ease), and disclosure disaggregation (granularity for financial/ESG measures). Positive tone is extracted using machine learning (ML) techniques applied to PDF reports. Thirdly, environmental violations are determined from Refinitiv Environmental Controversies which are coded as a binary variable.

The dependent variable is a decoupling measure, the Greenwashing Score, specified as

$GW\ Score_{it} = z(Positive\ Tone_{it}) - z(EPillar_{it})$ industry standardization. The independent variables are the E-pillar and its sub-pillars, readability, disaggregation, and the violation dummy; the controls are firm size, profitability, leverage and growth with industry/year fixed effects. The study combines text analytics (sentiment analysis, readability, and disclosure granularity) with panel regression techniques (fixed and random effects, with Hausman testing). The final outputs include a reproducible Greenwashing Score, regression results, and several robustness checks.

1.7 Limitation of the Study

This study has several limitations that should be considered when interpreting the results. First, although Refinitiv ESG ratings are produced using structured, audited methodologies, the study relies on one rating provider. Different providers may adopt slightly different scoring principles, meaning that results may vary if another ESG database is used. The limitation here relates to comparability across rating systems, not the credibility of the Refinitiv scores themselves. Text-based measures may also be affected by boilerplate wording, translation issues, or differences in how companies draft their sustainability reports. The disaggregation measure reflects the level of detail in disclosures, but not necessarily their accuracy. The environmental violation variable is coded as a simple binary indicator, which may miss minor incidents or delays in regulatory reporting.

Second, because this study uses observational data, there are internal validity concerns. Some relationships may involve endogeneity may change their reporting tone or attempt to improve their ESG ratings, which complicates causal interpretation when firms are facing public pressure. Other unobserved factors, such as supply-chain exposure, assurance quality, or reverse causality (e.g., increased scrutiny triggering violations rather than violations revealing greenwashing), may also influence the findings.

Third, external validity is limited. The results mainly apply to large Malaysian non-financial companies and cannot be assumed to hold for SMEs, financial institutions, or companies in other countries. Malaysia is selected because it is undergoing a shift from voluntary to mandatory sustainability reporting, making it an important context for studying greenwashing behaviour. However, other markets with stronger enforcement or different reporting cultures may show different patterns. The study period (2019–2024) is also relatively short, and relationships may evolve as new standards such as ISSB and TCFD become more widely adopted.

Finally, some statistical limitations exist, including potential multicollinearity among the environmental sub-pillars, missing data that reduce the number of firms to 82 and may affect statistical power, and multiple regression models that may increase the risk of Type I error.

1.8 Organization of the Thesis

The remainder of the thesis is organised as follows. Chapter 2 reviews the existing literature, develops the research hypotheses, and presents the conceptual framework. Chapter 3 describes the research methodology, including data collection procedures and empirical analysis techniques. Chapter 4 reports the descriptive statistics and correlation analysis of the dataset, presents the results of diagnostic tests, and reports the regression findings. Finally, Chapter 5 concludes the study by outlining the discussions of the regression results, implications, limitations, and recommendations for future research.

Chapter 2: Literature Review

This chapter outlines the academic foundation for the thesis. It starts with defining and deconstructing the concept of greenwashing. It then establishes a multi-pronged theoretical framework to explain the underlying motivations for this corporate behavior. Finally, it reviews the empirical literature to identify the specific research gaps that this study aims to fill, culminating in the development of formal, testable hypotheses.

2.1 Introduction

The literature review was organized into three parts. First, it defined and deconstructed the concept of greenwashing to establish the boundaries of analysis. Second, it reviewed relevant theories such as legitimacy, signaling, and agency, that explain corporate motivations behind greenwashing. Third, it synthesized empirical findings from global and Malaysian studies, highlighting gaps that this study addressed. The chapter culminated in the development of the conceptual framework and hypotheses.

2.2 Concept of Greenwashing

Greenwashing is generally defined as making unsupported or misleading claims about a company's environmental efforts or sustainability performance (Hu et al., 2023). It involves presenting information in a way that creates a false impression of strong environmental, social, or governance (ESG) performance (Rhou & Singal, 2020). The term was first introduced by environmentalist Jay Westerveld in the 1980s. Borrowed from the idea of "whitewashing," it refers to giving the public a positive environmental image that does not match the company's actual practices (Testa et al., 2023). Greenwashing is often described as a gap between what a company says about sustainability (the "talk") and what it actually does in practice (the "walk"). This means a company may appear environmentally friendly through statements and reports, even if little real improvement is happening internally.

There are several key characteristics of greenwashing. The first is unsupported or false claims, where companies make environmental statements that are not backed by data or are entirely untrue (Stewart, 2025). Second, misleading representation occurs when companies create the wrong impression about being environmentally responsible or sustainable (Yu et al., 2020). Keilmann and Koch (2023) explain that a large gap between what companies claim and what they actually do is known as the "talk–walk gap," often measured by comparing ESG disclosures with actual ESG performance.

Greenwashing can take different forms. The first type is claim greenwashing, which refers to false or exaggerated environmental claims (Lyon & Montgomery, 2015). The second is executional greenwashing, where companies use green-themed branding, colours, or imagery to appear environmentally friendly (Parguel et al., 2011). The third form is selective disclosure, where companies highlight positive information while hiding negative details (Lyon & Maxwell, 2011). Finally, decoupling describes a situation where companies stated commitments do not match their actual practices (Boxenbaum & Jonsson, 2008).

According to Zervoudi et al. (2025), greenwashing often occurs through written claims that are exaggerated, vague, unclear, or unsupported, especially when describing a product's environmental benefits. These claims can take different forms, including product-centered, procedure-centered, image-centered, environmental-fact claims, mixed claims, false promises, and falsified data. Product-centered claims focus on describing a product as having environmentally friendly features. Procedure-centered claims highlight processes that are claimed to be eco-friendly. Image-centered claims associate the company with environmental values to gain public approval. Environmental-fact claims generalize environmental benefits in a way that may mislead consumers. Mixed claims combine several types of claims to create an impression of being environmentally responsible without real commitment. False promises refer to environmental promises that companies do not intend to fulfil. Falsified data involves providing inflated or fabricated information to appear greener than the company truly is.

In addition, executional greenwashing uses visuals like the colour green, nature symbols, or environmental imagery to create the impression of sustainability without any real environmental actions behind it (Wu et al., 2025). Font et al. (2012) describe selective disclosure as highlighting only positive environmental or social achievements while leaving out negative information. This practice has been referred to as "reporting the news but not worrying," as companies present only the favourable parts of their performance.

Decoupling, in Corporate Social Responsibility (CSR) and firm behaviour, refers to the disconnection between an organization's formal commitments to social and environmental responsibility and the actual practices implemented within its operations. This concept originates from institutional theory, where firms often adopt CSR policies symbolically to gain legitimacy, comply with external pressures, or enhance reputation, while substantive changes to business practices may lag behind (Meyer & Rowan, 1977; Bromley & Powell, 2012). In this sense, CSR decoupling highlights the difference between what companies "say" and what they "do." One form of CSR decoupling is policy-practice decoupling, where organization formally adopts a policy but does not implement it in practice, or when implementation is so weak or symbolic that there is little real effect (Jabbouri et al., 2022). In their words, it refers to "a mismatch between adopted policies and internal organizational practices," where the policy is "symbolically adopted," but the organization either fails to follow through or monitors/enforces it poorly. The second form is means-ends decoupling. This is a newer, more subtle kind of gap, even when the policy *is* implemented. For example, the "means" are deployed and the organization still fails to achieve the policy's intended goals or outcomes "ends". That means the activities (practices) happen, resources are put in, but the outcomes fall short of what the policy promised or aimed for (Jabbouri et al., 2022). A particularly prominent type of CSR decoupling is communication–action decoupling, often associated with greenwashing. In such cases, companies heavily emphasize CSR activities in reports, advertising, and stakeholder communications, while the real environmental and social impact remains limited or even harmful (Marquis & Qian, 2013). This form of decoupling undermines stakeholder trust and raises concerns about the credibility of CSR reporting. For instance, Amel-Zadeh and Serafeim (2018) note that ESG ratings frequently reward disclosure quality rather than substantive performance, creating opportunities for firms to highlight symbolic actions while concealing weak practices.

Pizzetti et al. (2019) explain that greenwashing can be either active or passive. Active greenwashing happens when a company deliberately provides false, exaggerated, or misleading information. This is intentional and dishonest. Passive greenwashing, on the other hand, occurs when a company leaves out certain information, which unintentionally makes its environmental performance look less negative than it really is. Zervoudi et al. (2025) further describe two types of misleading behaviour, which are negative greenwashing, where companies try to make consumers believe their product has less environmental impact than it does, and positive greenwashing, where companies exaggerate small environmental benefits to improve their reputation and distract attention from more serious environmental issues. Webb (2023) also highlights the concept of "Greenwishing". This term, introduced by Duncan Austin in 2019, refers to overly optimistic or hopeful beliefs about the success of environmental initiatives. Greenwishing is not intentional dishonesty; instead, it reflects wishful thinking that makes environmental efforts seem more effective than they truly are, even when the real impact is much smaller.

2.3 Theoretical Framework

2.3.1 Legitimacy Theory

According to Huang et al. (2024), legitimacy theory is an important concept for understanding why organisations behave in certain ways, especially in relation to corporate social responsibility (CSR) and ESG disclosures. Legitimacy refers to society's belief that a company's actions are appropriate and acceptable based on commonly shared norms, values, and expectations (Suchman, 1995). This means that companies try to gain and maintain approval from different stakeholders because social acceptance is necessary for their long-term survival.

Suchman (1995) describes three types of legitimacy, which are pragmatic, moral, and cognitive. Pragmatic legitimacy is based on the self-interest of stakeholders. Stakeholders judge a company according to whether it benefits them. Lenowitz (2022) further divides this into three forms, which are exchange legitimacy, influence legitimacy, and depositional legitimacy. Exchange legitimacy refers to when stakeholders receive direct benefits from the company. Influence legitimacy is when stakeholders believe the company will act in their broader interests. Depositional legitimacy shows that stakeholders view the company as trustworthy or having good character. Zervoudi et al. (2025) note that more studies are still needed to understand how greenwashing affects this dimension.
Uniquely, Cognitive legitimacy is based on shared cultural values and expectations. A company is considered legitimate if its actions fit naturally into societal norms (Seele & Gatti, 2015).

Legitimacy theory is widely used to explain why companies engage in greenwashing (Forliano et al., 2025; Treepongkaruna et al., 2024; Esposito et al., 2025; Uyar et al., 2020). Companies may make unsupported or misleading claims about their environmental or ESG performance to appear environmentally responsible and shape stakeholder perceptions (Hu et al., 2025). This behaviour is more common in companies that face public criticism or pressure due to weak environmental or social performance (Patten, 2002). Companies may also increase voluntary environmental disclosures to make them more positive or favourable to maintain legitimacy even when their actual performance is poor (Treepongkaruna et al., 2024). This leads to **"**cheap talk**"** or decoupling, where what the company says ("the talk") does not match what it does ("the walk").

If companies fail to meet stakeholder expectations regarding environmental or economic performance, they risk losing legitimacy. This can expose them to public criticism, political pressure, and higher operational costs (Esposito et al., 2025). Poor sustainability performance can also cause reputational damage and loss of trust. Companies can use disclosures strategically to influence how stakeholders view their environmental and social performance (Font et al., 2012). This is often seen in industries such as energy, where firms must maintain legitimacy while facing increasing pressure to transition to sustainable practices (Choi & Shepherd, 2005). Overall, legitimacy theory highlights how companies adjust their communication and behaviour to appear socially acceptable. In many cases, this motive leads to greenwashing, where firms prefer to appear environmentally responsible rather than make real improvements.

Legitimacy theory is especially relevant in Malaysia, where new regulations have increased sustainability reporting requirements, but enforcement remains inconsistent. Bursa Malaysia now requires TCFD-aligned reporting, and Bank Negara Malaysia has introduced its Principles-Based Taxonomy to guide climate-related risk assessment (Bursa Malaysia, 2022; Bank Negara Malaysia, 2022). Recent sustainability auditing research also highlights that although reporting standards have strengthened, enforcement and verification still vary widely across Malaysian industries (Haikal et al., 2025). As a result, companies face strong pressure to show alignment with global sustainability norms, which may lead them to rely more on symbolic disclosures than on genuine environmental improvements (Fahmi et al., 2022).

For example, plantation companies accused of deforestation or labour issues often highlight carbon-reduction efforts or biodiversity initiatives in their sustainability reports. A similar pattern is observed in Malaysia's oil and gas sector, where firms continue expanding fossil-fuel operations while emphasising renewable-energy investments in their disclosures (Jamil & Wahyuni, 2024). These practices demonstrate how disclosure can be used as a legitimacy tool to satisfy regulators and maintain investor confidence despite operational weaknesses.

Family-owned conglomerates and government-linked companies also face especially strong societal expectations. Research shows that Malaysian firms tend to increase positive environmental disclosures when responding to legitimacy threats such as environmental controversies, NGO criticism, or regulatory sanctions (Ngu & Amran, 2021; Wan-Hussin et al., 2021). This behaviour reflects the "cheap talk" mechanism described in legitimacy-theory literature, where companies present favourable narratives to protect their legitimacy even if underlying practices remain unchanged (Fahmi et al., 2022).

Overall, legitimacy theory helps explain why Malaysian companies may engage in greenwashing. It highlights how a disclosure-driven system can encourage firms to prioritise reporting for social acceptance, while existing assurance and enforcement mechanisms do not always ensure that disclosures reflect actual performance (Hikal et al., 2025). In such an environment, greenwashing becomes a rational survival strategy, making legitimacy theory a key lens for understanding Malaysia's ESG landscape.

2.3.2 Signaling Theory

According to Connelly et al. (2010), signalling theory explains how companies communicate with external stakeholders when there is information asymmetry. In simple terms, companies know more about their true environmental performance than the public does. Because of this, firms with stronger environmental results are more likely to share voluntary CSR or ESG information to show that they are performing well, while weaker firms tend to disclose less (Clarkson et al., 2011; Mahoney et al., 2012). In this situation, sustainability disclosures act as a **signal** of quality. These signals are usually reliable for strong performers but can be misused by poor performers.

Signalling theory has often been used to study how "green claims" are used to show environmental responsibility (Kim & Lyon, 2014). Companies may use sustainability statements, CSR reports, or social media announcements to send signals about their environmental commitment, but these signals are not always accurate (Testa et al., 2023). Greenwashing occurs when companies intentionally send misleading signals, especially in markets where stakeholders cannot easily verify the company's true environmental practices (Xu et al., 2025). Low-performing companies may use selective disclosure to imitate successful firms, while high-performing companies are encouraged to share genuine sustainability actions to improve their reputation and firm value (Clarkson et al., 2007; Connelly et al., 2010). In this way, greenwashing becomes a misleading signalling strategy used to improve image without making real environmental progress (Lyon & Maxwell, 2011).

Researchers also explain that signalling theory links greenwashing to disclosure bias. Companies often emphasise positive results while minimising or leaving out negative information to create a favourable image (Guo et al., 2015). This selective reporting can mislead investors and stakeholders into believing that a company is more environmentally responsible than it actually is. However, as regulations strengthen and verification processes become more common, these biased signals become less effective. Wu et al. (2020) argue that stricter disclosure rules and better transparency reduce the benefits of greenwashing and encourage companies to improve their real sustainability performance.

In summary, signalling theory shows that environmental communication can serve two purposes: it can hide poor performance for low-performing firms, but it also provides a genuine way for high-performing firms to demonstrate real sustainability achievements. This dual behaviour highlights the importance of strong regulation and independent assurance to ensure that disclosures reflect actual performance.

In Malaysia, signalling theory helps explain why companies often highlight positive environmental initiatives in their ESG disclosures while downplaying weaker areas. With stricter reporting requirements from Bursa Malaysia and Bank Negara Malaysia, firms face pressure to disclose climate and sustainability information, but weaker enforcement allows them to manage these disclosures strategically. According to the news from Revon Media in 2023, Sarawak Energy initiated several tree-planting activities to promote sustainable environmental conservation across its assets, facilities and project areas in the State. This indicates that companies more likely in plantation and energy are those who frequently promote reforestation or renewable-energy projects as their core businesses are related to practices like deforestation and fossil-fuel expansion. Because investors rely heavily on disclosure-based ESG ratings, companies are encouraged to improve the tone and volume of their reports rather than make real environmental improvements. As a result, Malaysian firms may opt for selective disclosure to appear more sustainable, protect their reputation, and meet reporting expectations even when their true performance remains limited or unobservable.

2.3.3 Agency Theory

According to Jensen and Meckling (1979), agency theory explains the relationship and potential conflicts between principals (such as shareholders) and agents (such as managers). Because managers control daily operations, while shareholders only provide capital, their interests may not always align. The theory assumes that managers may prioritise their own short-term goals over the shareholders' long-term objectives. This happens due to information asymmetry, where managers have more information about the company than shareholders, giving them the opportunity to act in their own interest.

Agency theory helps explain the internal pressures that may lead managers to engage in greenwashing (Testa et al., 2023). As agents, managers may choose to present an overly positive environmental image to satisfy short-term stakeholder expectations, even if this conflicts with long-term shareholder interests or does not reflect real environmental improvements. Tsang et al. (2022) note that this can involve adjusting reports to appear "greener" than the company's actual practices. Executives may also prioritise short-term profits and view greenwashing as a "low-cost, high return" way to enhance reputation without making meaningful operational changes.

Greenwashing is strongly influenced by information asymmetry, where managers take advantage of the gap between disclosed ESG information and a firm's actual performance (Hu et al., 2023). Research shows that companies often engage in selective disclosure, highlighting positive environmental outcomes while hiding negative ones to create an overly favourable image (Nyilasy et al., 2013). Rating disagreements across ESG providers can further increase information opacity and encourage such behaviour (Hutton et al., 2009). Wang et al. (2025) argue that greenwashing reflects a form of moral hazard, where managers focus on short-term gains rather than long-term value. This may include diverting funds away from genuine environmental projects or disguising non-green activities as sustainable (Lian et al., 2023). Managers may also portray a “green” image to hide poor environmental or social performance and mislead stakeholders (Lee & Raschke, 2023).

Agency theory explains that self-interested managers may engage in greenwashing when information asymmetry allows them to hide weak environmental performance. It also stresses the need for strong governance and monitoring systems to reduce these risks. Better oversight limits managerial discretion and helps ensure that ESG practices are genuine rather than merely symbolic.

In Malaysia, efforts such as Bursa Malaysia’s *Sustainability Reporting Guide* and the Securities Commission’s push for stronger ESG practices represent progress toward better governance. However, regulatory oversight remains in a developmental stage, meaning that further improvements like requiring external assurance of ESG disclosures are important for reducing opportunities for greenwashing (Bursa Malaysia, 2022).Market discipline also plays a role. Institutional investors, credit rating agencies, and ESG rating providers can strengthen monitoring, as large investors increasingly use shareholder engagement and voting rights to demand greater transparency and accountability, aligning managerial behaviour with long-term value creation (Dyck et al., 2018). Internally, stronger control mechanisms such as linking executive compensation to verified sustainability outcomes can reduce moral hazard. When managerial rewards depend on actual environmental performance instead of symbolic reporting, the conflict between managers and shareholders is reduced (Flammer et al., 2019).

Overall, agency theory helps explain why managers who are motivated by self-interest and enabled by information asymmetry may engage in greenwashing. It also shows that effective governance and external monitoring are essential to limit such behaviour and ensure sustainability reporting reflects real long-term value rather than symbolic communication.

2.4 Empirical Literature Review and Identification of Gap

This section critically evaluates the existing empirical research on ESG ratings, corporate disclosure, and greenwashing, with a particular focus on the Malaysian context. The aim is to synthesize key findings and, more importantly, to identify the specific gaps in the literature that this thesis is designed to address.

2.4.1 Beyond the Numbers: The Reliability Crisis in ESG Ratings

The rapid growth of ESG ratings has raised many concerns about how reliable and meaningful these scores are. Berg et al. (2022) highlight that different rating agencies often produce very different scores for the same company. They refer to this problem as "aggregate confusion," caused by differences in measurement, scope, and weighting, where measurement differences alone explain more than half of the variation. Other studies report similarly low correlations, sometimes around 0.60, which shows that ESG ratings are far less consistent than credit ratings, which often reach correlations of 0.99 (MIT Sloan, 2021).

Another issue is that many ESG ratings focus on how environmental or social factors affect a company's financial risk ("outside-in") rather than measuring the company's actual impact on society or the environment ("inside-out"). This can result in companies with harmful environmental practices still receiving high ESG scores (Harvard Law School Forum on Corporate Governance, 2022). ESG ratings also tend to reward companies for the amount and quality of their disclosures rather than their real environmental performance. As a result, firms may invest more in improving their reporting than in making meaningful sustainability improvements. Critics also point out that some rating agencies use aggressive or opaque rating methods to increase market share and may face conflicts of interest when they sell consulting services to the same companies they evaluate.

Empirical studies in Malaysia show that ESG practices are still largely shaped by disclosure, as many ESG assessments depend on company-reported data when verification mechanisms are still developing (Jamil & Wahyuni, 2024). Governance structures such as independent boards and CSR committees can help guide firms toward more transparent reporting and reduce greenwashing risks. However, because Malaysia's ESG landscape is still evolving, with ongoing improvements in regulatory oversight and growing academic attention (Jamil & Wahyuni, 2025; Yu et al., 2020; Dorfleitner & Utz, 2023). ESG assessments may place greater emphasis on the quality and completeness of disclosures. This suggests that disclosure-focused practices remain an important part of how Malaysian companies communicate sustainability performance.

A clear research gap exists because prior studies have not directly linked Environmental Performance to greenwashing by combining it with textual measures such as readability and disaggregation alongside environmental violations. No existing research systematically integrates these dimensions into a single framework, leaving the relationship between Environmental Performance and greenwashing largely unexplored.

2.4.2 Measuring Corporate Disclosure Quality: From Words to Numbers

The accounting and finance literature provides several objective and replicable ways to assess the quality of corporate disclosures, moving beyond subjective judgement. Scaltrito (2015) explains that one common method uses linguistic readability measures, such as the Flesch-Kincaid Grade Level or the Gunning Fog Index, to determine how easy or difficult a text is to understand. The underlying idea is that well-performing firms usually communicate in a clearer manner, while weaker performers may use more complex language to hide poor results, a strategy known as intentional vagueness (Hu et al., 2024). These findings suggest that companies with more readable ESG disclosures are generally less likely to engage in greenwashing, especially when information asymmetry is high.

In addition to readability measures, researchers also use sentiment and content analysis through Natural Language Processing (NLP) to evaluate disclosure tone and themes. These methods classify statements as positive, neutral, or negative and identify how often certain topics appear (Motz et al., 2025). By analysing how frequently companies highlight environmental achievements compared to risks or weaknesses, NLP provides a more detailed understanding of the symbolic nature of corporate sustainability communication (Calamai et al., 2025).

Another methodology involves constructing disclosure indices. For example, Almaqtari et al. (2023) examine how board structure, IT governance, innovation, FinTech, and green finance relate to sustainability performance in Indian banks, finding that strong board and IT practices support innovation and FinTech adoption, while green initiatives show limited short-term impact. Their study applies this approach by using secondary data from reputable sources and panel data analysis to assess how board attributes and firm characteristics shape disclosure levels.

Finally, the literature introduces an objective way to measure financial transparency known as Disaggregation Quality (DQ), developed by Chen et al. (2015). This measure captures how detailed a firm's reporting is by counting the number of non-missing line items in its balance sheet and income statement using Compustat data. Higher DQ values are linked to lower information asymmetry, smaller differences in analyst forecasts, tighter bid–ask spreads, and a lower cost of equity. These relationships support DQ as a useful indicator of disclosure clarity and reporting detail.

Although many methods exist to assess disclosure quality, most studies use these measures separately. Readability reflects the clarity of a firm's language, while disaggregation reflects the level of detail in its financial reporting; however, these two dimensions are rarely examined together in greenwashing research. This gap is important in settings like Malaysia, where ESG assessments rely heavily on company disclosures (Jamil & Wahyuni, 2024).

This study uses both financial disaggregation measures and readability indices for three main reasons. First, readability reflects the symbolic side of disclosure.

Companies under scrutiny may change the complexity or tone of their sustainability reports to create a more favourable image (Hu et al., 2024). Second, higher levels of detail and transparency help reduce information asymmetry and limit opportunities for managers to report selectively; disaggregation therefore represents the substantive side of disclosure (Chen et al., 2015). By combining these measures, the study assesses whether firms that provide clearer and more detailed disclosures are less likely to show a gap between their sustainabiltiy communication and their actual performance. Third, using both metrics together offers a more robust and multidimensional indicator of disclosure quality, as disaggregation captures structural transparency in financial reporting, while readability can reveal deliberate ambiguity in sustainability narratives.

Together, these factors create a comprehensive framework for assessing how disclosure practices influence greenwashing. This approach is particularly relevant in Malaysia, where businesses face strong pressure to signal ESG compliance, but independent verification mechanisms are still developing. By balancing the symbolic and substantive aspects of disclosure, the methodology offers an objective and replicable way to detect greenwashing in emerging markets.

2.4.3 Greenwashing and ESG in the Malaysian Context

The literature on ESG and greenwashing in Malaysia, while expanding, remains underdeveloped compared to that in more mature markets such as North America and Europe (Jamil & Wahyuni, 2025). Nevertheless, few important themes have emerged.

Greenwashing is a recognised issue in Malaysia’s corporate sector. Case studies show that firms in sensitive industries, such as oil and gas, palm oil, and energy, have at times overstated or misrepresented their sustainability efforts. For instance, Jamil and Wahyuni (2024) found that a major oil and gas company continued to exhibit signs of greenwashing despite extensive reporting. Such evidence suggests that sustainability disclosures in Malaysia do not always reflect genuine environmental performance.

Second, regulatory and institutional efforts in Malaysia are becoming stronger. Bursa Malaysia and the Securities Commission now require listed companies to report ESG information, including climate risks, under the Sustainability Reporting Framework aligned with TCFD guidelines (Bursa Malaysia, 2022). Malaysia has also introduced the Sustainable and Responsible Investment (SRI) Roadmap and a national green taxonomy to reduce greenwashing and improve the credibility of ESG practices (Securities Commission Malaysia, 2019; Bank Negara Malaysia, 2022). Together, these initiatives show that Malaysia is moving toward a more structured and internationally aligned ESG regulatory environment.

Third, while empirical research on ESG performance in Malaysia is growing, it remains limited. Existing studies show that Malaysian firms often lag behind global peers in ESG adoption, but strong governance mechanisms, such as active board oversight and CSR committees, can help reduce greenwashing risks (Jamil & Wahyuni, 2025). Evidence also indicates that firms with higher-quality ESG disclosures tend to gain greater investor confidence, whereas companies perceived as engaging in greenwashing may suffer reputational consequences in both financial markets and public perception (Yu et al., 2020; Dorfleitner & Utz, 2023).
Comparing Malaysia with the wider Asian region, several similarities and differences become clear. Like many Asian economies, Malaysia has strengthened its sustainability regulations and faces the dual challenge of rapid economic growth alongside increasing ESG expectations (OECD, 2022). However, unlike larger economies with stronger enforcement systems, such as China and South Korea, Malaysia's ESG landscape remains mostly disclosure-driven and depends heavily on company-reported data with limited external verification (Jamil & Wahyuni, 2024; Hu et al., 2024). In addition, although greenwashing has been studied across Asia through industry-specific and cross-country research (De Freitas Netto et al., 2020; Kathan et al., 2025), Malaysian evidence is still fragmented, relying mainly on case studies and governance-focused analyses.

A key research gap in Malaysia is the lack of large-scale quantitative studies that measure greenwashing across multiple industries using systematic, replicable methods. Most existing work relies on case studies or narrow governance-focused analyses. This study addresses the gap by applying a novel greenwashing score to Malaysia's top public listed companies to provide broader empirical evidence on greenwashing patterns and their determinants.

2.4.4 Empirical Findings

2.4.4.1 ESG Scores: E pillar

Table 2.1 Empirical Findings for E Pillar

Study/ Author (Year)	Independent Variable	Key Focus/ Methodology	Key Findings
Kathan et al. (2025)	ESG Scores (E Pillar)	ESG scores and greenwashing behaviour in large firms globally	High ESG scores, particularly in large firms, may mask greenwashing behaviours.
Treepongkaruna (2024)	ESG Scores (E Pillar)	Correlation between ESG/environment ratings and carbon emissions	High environmental ratings do not correspond to lower carbon emissions, therefore raising doubts about ESG reliability.
Sklavos (2025)	ESG Scores (E Pillar)	Alignment of actual environmental performance vs. ESG disclosures in	Environmental Pillar (E) scores often inflate a firm's

		financial institutions in Europe	sustainability image without necessarily reflecting real environmental responsibility, thus serving as a potential driver of greenwashing risk.

High ESG scores do not always reflect real environmental performance. Kathan et al. (2025) show that even companies with strong ESG ratings may still engage in greenwashing, especially larger firms. Treepongkaruna (2024) also found no clear relationship between Environmental (E) scores and actual emissions, indicating a gap between reported ratings and real environmental outcomes. Sklavos (2025) strengthens this point by introducing a Greenwashing Risk Index that measures the misalignment between disclosure and performance, providing a more reliable way to detect claims that are symbolic rather than supported by genuine environmental action.

H1 (ESG E Pillar): The environmental ESG scores (emissions, resource use, and innovation) are negatively associated with the measured level of greenwashing.

2.4.4.2 E Sub-pillar: Emissions

Table 2.2 Empirical Findings for E Sub-pillar: Emissions

Study/ Author (Year)	Independent Variable	Key Focus/ Methodology	Key Findings

Treepongkaruna et al. (2024)	Emissions	Panel Granger-causality models with firm fixed effects; alternative regressions swapping in the three E sub-scores; rich controls (size, profitability, leverage, capex, R&D, SG&A, cash, dividends; board size/independence)	High E/ESG ratings not linked to lower emissions (null/weak or wrong-signed) → potential decoupling
Cregan et al. (2023)	Emissions Score (from several major raters)	Cross-sectional and panel regressions (industry-specific), within-industry tests of predictive validity, ratings-disagreement analysis	No evidence that emissions scores capture or predict reported emissions; large divergence across providers.
Tanthanongsakkun et al. (2022)	Emissions (From Refinitiv)	Panel designs with firm FE, plus PSM, entropy balancing, IV, and dynamic GMM for causality around governance shocks (staggered	Governance that weakens discipline (e.g., staggered boards) is associated with worse emissions performance;

		boards; takeover threats)	results are consistent when using Refinitiv emissions score as proxy.
Lazar et al. (2024)	Emissions (From Refinitiv)	Panel regressions (sector-level), heat-map significance diagnostics; FE; controls from Compustat/CRSP	Emissions score shows predictive links to downside risk measures; sign/economic strength vary by sector.
Shahrour et al. (2024)	Emission (From Refinitiv)	Panel fixed-effects models split pre- vs. post-Paris Accord; interactions with R&D intensity	R&D strengthens the link between emissions performance (higher emission score) and financial outcomes
Potharla & Turubilli (2024)	Emission Reduction Score (ERS)	Asset-pricing regressions (CAPM, FF-4, FF-5) with robustness adding S & G scores and firm controls	Higher ERS associates with lower synchronicity (more firm-specific information in prices).

Empirical research shows that emissions scores, or broader environmental (E) scores, do not always reflect real CO_2 performance, and different rating providers often given conflicting assessments, which are conditions that increase the risk of a gap between what firms report and what they do (Treepongkaruna et al., 2024; Cregan et al., 2023). Weak governance can worsen this problem: firms with less monitoring or staggered boards tend to have poorer emissions outcomes, even when their Emissions Score appears strong (Tanthanongsakkun et al., 2022). Sector and capability differences also play a role. In some industries, higher Emissions Scores are genuinely linked to lower climate risk, suggesting that the score can reflect real progress when supported by strong R&D capabilities (Lazar et al., 2024; Shahrour et al., 2024). For greenwashing detection, a company that has a high Emissions Score but also uses overly positive narrative tone or has documented environmental violations signals a potential talk–walk mismatch. This mismatch is the pattern that this study's greenwashing score aims to identify. Markets also appear to favour genuine emissions improvements; for example, higher Emission Reduction Scores are linked to lower stock price synchronicity, meaning that a strong emissions score combined with weak market signals can be another warning sign (Potharla & Turubilli, 2024). Overall, Emissions Scores tend to correlate negatively with greenwashing when firms are well governed and supported by external evidence, but the relationship can reverse when high scores occur alongside weak oversight, optimistic disclosures, or violation-related red flags.

H1a (Emissions): Emissions score should be negatively associated with the *measured level* of greenwashing.

2.4.4.3 E Sub-pillar: Resource Use

Table 2.3 Empirical Findings for E Sub-pillar: Resource Use

Study/ Author (Year)	Independent Variable	Key Focus/ Methodology	Key Findings
Treepongkaruna et al. (2024)	Resource Use	Large U.S. panel: firm fixed-effects and panel Granger-	Resource Use (like the broader E pillar) does not reliably predict lower

		causality; swaps in each E sub-score (Emissions, Resource Use, Innovation) with rich financial and governance controls	current/future CO_2; coefficients are generally weak/insignificant, consistent with rating–outcome decoupling.
Lazar et al. (2024)	Resource Use	U.S. firms: sector-level fixed-effects panel regressions of climate Value-at-Risk (VaR) and Expected Shortfall (ES) on E sub-scores	Higher Resource Use scores are associated with lower downside climate risk in some heavy-impact sectors, but effects are heterogeneous across industries.
Shahrour et al. (2024)	Resource Use	Panel fixed-effects models split pre- vs. post-Paris Accord; interactions with R&D intensity	Resource Use score shows stronger links to financial performance where R&D is high, suggesting capabilities help translate resource practices into outcomes.

Studies that include Resource Use in their models show that its effects depend on the context. In large panel analyses that try to predict actual emissions, Resource Use behaves similarly to the overall Environmental (E) pillar, showing limited predictive power and suggesting possible decoupling between reporting and real performance (Treepongkaruna et al., 2024). However, in risk-focused models, Resource Use is sometimes linked to lower climate-related downside risk, especially in environmentally intensive industries (Lazar et al., 2025). Research also shows that Resource Use becomes more meaningful for firm value when companies have strong innovation capabilities, such as higher R&D investment (Shahrour et al., 2024). Based on these mixed findings, this study expects Resource Use to have a negative coefficient if it reflects real efficiency improvements and lower greenwashing, but the effect may be weak or unclear if the score mainly represents disclosure-heavy reporting rather than actual operational performance.

2.4.4.4 E Sub-pillar: Innovation

Table 2.4 Empirical Findings for E Sub-pillar: Innovation

Study/ Author (Year)	Independent Variable	Key Focus/ Methodology	Key Findings
Albitar et al. (2022)	Innovation	Panel regressions with Fixed Effect; interaction with environmental governance	Innovation reduces CO_2, and the effect is stronger when environmental governance is high.
Macchioni et al. (2024)	Innovation	Value-relevance tests linking innovation to equity values	Innovation is positively associated with firms' market values

Cheng et al. (2024)	Granular Green Patent Data (as Innovation proxy)	Panel models using patent-based innovation measures	Green innovation predicts better future environmental and financial outcomes
Ju and Jin (2024)	Green Innovation Index	Firm-level panel, FE; governance moderators	Green innovation improves carbon performance; internal governance shapes strength of the link.
Miao et al. (2024)	Green Technological Innovation	Fixed-effects linear model	Green technological innovation raises emission efficiency (i.e., reduces emissions per output)
Truong (2025)	Green Innovation (Patent Based)	Firm FE models with emissions data	Developing green innovations lowers toxic emissions; effect robust across specs.

Evidence from past empirical studies shows that environmental or green innovation is generally linked to better real environmental outcomes and higher firm value, suggesting lower greenwashing risk. Research finds that firms investing in innovation reduce CO_2 emissions and improve efficiency, especially when supported by strong environmental governance (Albitar et al., 2022; Miao et al., 2024). Patent-based indicators of green innovation also predict improvements in future environmental and financial performance (Cheng et al., 2024), and panel studies show that green innovation enhances carbon performance, with governance strengthening this effect (Ju & Jin, 2024). Market-based findings further demonstrate that innovation is positively related to firm valuation, indicating that investors can distinguish genuine innovation from mere narrative claims (Macchioni et al., 2024). Evidence from manufacturing supports this pattern, showing that green innovation helps reduce toxic emissions (Truong, 2025). Overall, these studies suggest that innovation narrows the gap between what firms say and what they do, supporting a negative relationship between innovation and greenwashing.

H1c (Innovation): The Innovation score should be negatively associated with the *measured level* of greenwashing.

2.4.4.5 Disclosure Quality: Readability

Table 2.5 Empirical Findings for Readability Score

Study/ Author (Year)	Independent Variable	Key Focus/ Methodology	Key Findings
Hu et al. (2024)	Modified Fog Index	Firm-level greenwashing constructs a peer-relative GW score; tests association with readability	Greater readability is associated with less greenwashing especially in high information asymmetry contexts.
Gorovaia and Makrominas (2024)	Report length, positivity, readability proxies in CSR reports	Identifies greenwashing among CSR violators; cross-firm tests	Firms implicated in CSR violations issue more copious but less readable reports, which is consistent with impression management/greenwashing.
Ahn et al. (2023)	Standard readability indices for sustainability reports	Market informativeness: tests firm-specific information in prices	More readable sustainability reports carry more firm-specific information (lower synchronicity), implying

			higher informational quality.
Uddin and Chakraborty (2021)	Readability of sustainability report under GRI/G4	Descriptive or diagnostic analysis of non-financial report readability	Many sustainability reports are hard to read, aligning with concerns about impression management (greenwashing) in non-financial disclosures.
Li (2008)	Fog/Gunning-style 10-K readability (financial reports)	Tests links to current earnings and earnings persistence	Poor readability is associated with properties consistent with obfuscation (greenwashing risk), a foundational result often extended to ESG texts.
Lin et al. (2023)	CSR report readability; institutional ownership	CSR readability, ESG performance, and investor influence	Institutional investors are associated with improved CSR readability and better ESG outcomes, consistent with monitoring effects.
Shimamura et al. (2025)	LLM-assisted readability scoring of sustainability reports	Readability ↔ ESG scores link for U.S. firms	Report readability systematically relates to ESG scores, highlighting how language clarity co-moves with perceived ESG quality.

Overall, the evidence shows that low readability is a warning sign of greenwashing risk. Studies find that firms with less readable ESG disclosures are more likely to engage in greenwashing (Hu et al., 2024), and companies facing CSR violations often issue longer but harder-to-read reports, reflecting classic impression-management behaviour (Gorovaia & Makrominas, 2024). Other research reports that more readable sustainability disclosures provide markets with clearer firm-specific information (Ahn et al., 2023), while reviews of sustainability reports show that readability is generally low, raising concerns about intentional obfuscation (Uddin & Chakraborty, 2021). Foundational work in financial reporting also links poor readability to managerial obfuscation (Li, 2008), and newer studies show that readability affects perceived ESG quality and investor monitoring (Lin et al., 2023; Shimamura et al., 2025). Based on this evidence, this study expects readability to be negatively associated with greenwashing.

H2a (Readability Score): **There is a negative relationship between readability score and the measured level of greenwashing.**

2.4.4.6 Disclosure Quality: Disaggregation

Table 2.6 Empirical Findings for Disaggregation Score

Study/ Author (Year)	Independent Variable	Key Focus/ Methodology	Key Findings
Chen et al. (2015)	Disaggregation Quality (DQ): count of non-missing GAAP line items in annual reports (from Compustat)	Validation tests + regressions relating DQ to information environment (e.g., bid–ask spread, analyst properties) and cost of equity association with readability	Higher DQ → lower information asymmetry and lower cost of equity; DQ captures the “fineness” of reported data.
Johnston et al. (2023)	ITEMS (XBRL-based): count of balance-sheet & income-statement line	Develops an XBRL-native disaggregation metric; compares/validates against DQ and related measures	ITEMS provides a simple, timely disaggregation proxy; conceptually aligned with DQ and not dependent

	items directly from 10-K XBRL filings (extends DQ beyond data vendors)		on aggregator choices.
Hoitash and Hoitash (2014)	ARC / XBRL tag counts: number of GAAP taxonomy tags and custom extensions in 10-K XBRL	Propose reporting complexity/disaggregation measures using XBRL detail; used to study information environment outcomes	XBRL tag-count–based measures capture reporting detail/complexity and are widely used to proxy disclosure granularity
Casey et al. (2023)	Modified DQ (MDQ): refinement of Chen–Miao–Shevlin's DQ	Methodology paper proposing improvements to DQ measurement	Presents MDQ as an updated disaggregation metric, reinforcing the usefulness of line-item granularity as disclosure quality
Du et al. (2022)	Uses DQ construct when examining database discrepancies	Empirical assessment of database differences and their implications for research measures	Notes DQ's definition (non-missing Compustat line items) and highlights measurement sensitivity to data sources

Across accounting research, more disaggregated reporting, whether measured through the original DQ metric based on non-missing Compustat line items or through XBRL tag count, is consistently associated with a better information environment, including lower information asymmetry and lower cost of equity (Chen, Miao, & Shevlin, 2015; Johnston, 2024). XBRL evidence also shows that detailed tagging reflects the level of reporting granularity or complexity, supporting disaggregation as a reliable indicator of disclosure quality (Hoitash & Hoitash, 2018; Hoitash, 2022). Because greenwashing often relies on opacity in both narrative and numerical disclosures, higher disaggregation should reduce the gap between what firms claim and what they actually do by making performance easier to verify. Accordingly, this study expects a negative relationship between disaggregation and greenwashing, meaning that firms providing more detailed line-item information are less likely to engage in greenwashing.

H2b (Disaggregation Score): **There is a negative relationship between disaggregation score and the measured level of greenwashing.**

2.4.4.7 Environmental Violation

Table 2.7 Empirical Findings for Environmental Violation

Study/ Author (Year)	Independent Variable	Key Focus/ Methodology	Key Findings
Zhou et al. (2024)	Environmental Violation Dummy	Impact of environmental administrative penalties on corporate greenwashing via deep learning text analysis	Administrative penalties serve as a deterrent, therefore firms with penalties show different patterns in disclosure tone.

Gorovaia and Makrominas (2024)	Violator status (dummy for firms committing CSR-/environment-related violations)	Binary logit and related tests to identify greenwashing among CSR violators	Firms flagged as violators exhibit greenwashing patterns (e.g., disclosure features consistent with impression management)
Shao et al. (2025)	Environmental penalties (firm-level)	Panel study (China, 2009–2019) on penalties → carbon disclosure	Penalties increase carbon disclosure, consistent with a legitimacy response after sanctions.
Hu and Xu (2025)	Regulation penalties	Panel regressions on penalties → environmental information disclosure (EID)	Penalties significantly enhance EID, reinforcing that enforcement pressure triggers more disclosure.
Karpoff et al. (2005)	EPA violation / enforcement events	Event study of market reaction to environmental violations	Significant negative equity value losses follow violation disclosures, implying reputational costs that can incentivize image repair behaviour.
Huang et al. (2017)	Disclosure of environmental violations (event indicator)	Stock-market response to violation disclosures in China	Shareholders react negatively to violation disclosures, confirming market penalties for poor environmental conduct.

Yun et al. (2023)	Environmental performance fees/charges (proxy for poor performance)	CSR environmental disclosure ↔ performance (panel)	Finds a cover-up phenomenon: higher disclosure coincides with worse environmental performance in underdeveloped regions, which is consistent with symbolic reporting after poor outcomes

Across existing studies, environmental violations and penalties generally lead to reputational and financial costs for firms and often trigger additional disclosure. However, this does not necessarily reflect better environmental performance and instead can create conditions that encourage symbolic communication and greenwashing. Event studies show that markets react negatively to violation announcements, which increases pressure on firms to restore legitimacy through more optimistic reporting (Karpoff et al., 2005; Huang et al., 2017). Panel studies further find that penalties lead to higher levels of environmental and carbon disclosure, reflecting legitimacy-repair efforts rather than assured operational improvements (Shao et al., 2025; Hu et al., 2025). More directly, Zhou et al. (2024) show a strong positive link between administrative penalties and later greenwashing, especially in firms with weak internal controls or facing highly competitive and pollution pressures. These findings indicate that environmental violations or penalties are likely to be positively associated with greenwashing.

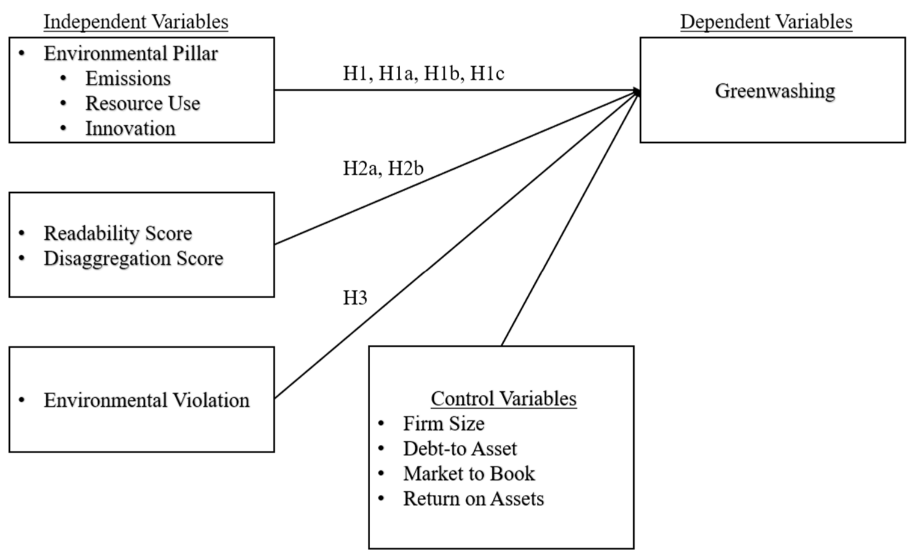
H3 (Environmental Violations): The presence of environmental violations is positively associated with the likelihood of greenwashing.

The literature shows that no single measure, whether ESG environmental scores, disclosure readability, or environmental violations, can reliably detect greenwashing on its own. ESG E-pillar scores often fail to reflect real environmental performance, readability and disaggregation capture only the symbolic side of disclosure, and violations remain underused in empirical research. This highlights the need for a more comprehensive approach. A composite Greenwashing Score model that combines E-pillar ratings, disclosure-quality measures, and violation data can address this gap by capturing both symbolic claims and substantive behaviour. Such a multidimensional measure is especially valuable for identifying greenwashing within the Malaysian corporate context.

2.5 Conceptual Framework & Hypotheses Development

2.5.1 Conceptual Framework

Figure 2.1 Proposed Conceptual Framework



Source: Created for this study

The conceptual framework shows how environmental performance, disclosure quality, and environmental violations are expected to influence greenwashing. First, the environmental score is divided into emissions (H1a), resource use (H1b), and innovation (H1c). Together, these form H1, which predicts that stronger environmental performance should reduce greenwashing. Second, disclosure quality is measured using readability (H2a) and financial disaggregation (H2b). H2 proposes that clearer and more detailed disclosures lower the likelihood of greenwashing by limiting impression-management opportunities. Third, environmental violations (H3) act as an external accountability indicator, with the expectation that firms with more violations are more likely to engage in greenwashing to offset poor operational practices. Overall, the framework links these three dimensions to explain why some firms exhibit higher levels of greenwashing than others.

Chapter 3: Methodology

3.1 Introduction

This chapter provides the detailed operational blueprint for the empirical investigation. It outlines the process of sample selection, defines the data sources, specifies the precise measurement of all variables, and presents the econometric models that will be used to test the hypotheses developed in Chapter 2. The methodology is designed to be transparent, rigorous, and replicable.

3.2 Research Design

Panel regression was employed in this study because it offers distinct advantages over a cross-sectional approach. A cross-sectional model captures data for a single point in time, limiting its ability to account for unobserved heterogeneity across firms. By contrast, panel data combines both cross-sectional and time-series dimensions, allowing the researcher to track variations in firm characteristics and disclosure behaviour over multiple years.

This approach offers three main advantages. First, it accounts for firm-specific factors, such as corporate culture, governance style, or industry norms, that do not change over time but could otherwise distort the results if not controlled for (Hsiao, 2014). Second, panel regression improves the accuracy and reliability of the analysis by using both time-series and cross-sectional data, which increases the number of observations and reduces multicollinearity (Baltagi, 2021). Third, panel models allow the use of fixed-effects or random-effects techniques, making them especially useful for identifying consistent patterns in greenwashing behaviour across firms and over time (Borenstein et al., 2010).

Given that this study investigates how environmental scores, disclosure quality, and violations influence greenwashing across multiple Malaysian firms from 2019 to 2024, a panel approach provides a more accurate and causally reliable framework than a single-year cross-sectional snapshot.

3.3 Sample Selection and Data Sources

The research will employ a quantitative, archival approach using a large panel dataset of publicly listed firms in the Malaysian market.

3.3.1 Target Population

Based on Appendix 1, this study initially selected the top 100 publicly listed companies on Bursa Malaysia, representing the largest and most influential firms in the Malaysian economy. These companies were identified according to their market capitalization as of 5 September 2025. After removing firms with missing or incomplete data, the final sample comprises 82 companies, which form the basis of the empirical analysis.

Financial institutions, such as banks, insurers, and investment holding companies, were excluded for several reasons to ensure consistency with prior ESG and greenwashing research. First, they operate under different regulatory frameworks, particularly Bank Negara Malaysia's Climate Change and Principle-Based Taxonomy (CCPT), which differs from Bursa Malaysia's Sustainability Reporting Guide for non-financial firms and creates inconsistencies in ESG measurement (Jamil & Wahyuni, 2024). Second, ESG metrics for non-financial companies reflect operational impacts like emissions and resource use, whereas financial institutions focus more on governance, investment screening, and portfolio risks. Including both groups would distort greenwashing evaluations. Third, earlier studies often exclude financial firms to avoid bias, as their unique reporting structures may either inflate or suppress indications of greenwashing (Yu et al., 2020; Testa et al., 2023). Restricting the sample to non-financial firms therefore provides a clearer and more comparable assessment of the proposed relationships.

The final sample consists of non-financial firms across a wide range of sectors, including energy, utilities, consumer goods, technology, industrials, healthcare, basic materials, and others.

3.3.2 Sample Period

The study will cover the period 2019 to 2024. This timeframe is being chosen for several reasons.

First, 2019 is used as the pre-policy baseline and marks the beginning of Malaysia's modern sustainability agenda. In that year, the Securities Commission launched the Sustainable and Responsible Investment (SRI) Roadmap (2019–2025) to guide the development of ESG financing in Malaysia, making 2019 an appropriate anchor year for assessing conditions before major reforms (Securities Commission Malaysia, 2019). Around the same time, Bank Negara Malaysia released its initial discussion paper on climate taxonomy in December 2019, which later evolved into the Climate Change and Principle-based Taxonomy (CCPT) framework (Bank Negara Malaysia, 2021).

Second, the main regulatory "shock" that expanded disclosure requirements occurred between 2021 and 2022 would potentially increased incentives for greenwashing. On 30 April 2021, BNM formally issued the CCPT, providing a classification system for climate-related economic activities and outlining how environmental sustainability should be assessed. Bursa Malaysia then strengthened its Sustainability Reporting Framework in September 2022, requiring all Main Market companies to adopt TCFD-aligned climate disclosures by 2025 (Bursa Malaysia, 2022). The accompanying Sustainability Reporting Guide (3rd ed.) introduced a consistent template for reporting text and metrics beginning in FY2022, standardizing what companies must disclose.

Third, the years 2023–2024 mark an important period shaped by regional taxonomy developments that influence how "green" claims are defined and assessed. The ASEAN Taxonomy Version 3, released in March 2024, introduced sector-specific criteria for industries such as construction and transportation (ASEAN Taxonomy Board, 2023; 2024). Meanwhile, Version 2 expanded the Foundation Framework through guiding questions and decision trees to help classify sustainability activities. These regional updates likely shaped how Malaysian companies described and justified their environmental performance in their 2023–2024 disclosures.

Finally, stopping at FY2024 avoids the structural disruption brought about by Malaysia's 2024 decision to adopt ISSB-aligned reporting starting in FY2025 under the National Sustainability Reporting Framework (NSRF). Given that the NSRF was finalized in September 2024 and would go into effect in FY2025, stopping the analysis at that point maintains the uniform pre-ISSB regime and neatly separates the policy-intensification period without introducing a new reporting basis (IFRS Foundation, 2025; Bank Negara Malaysia, 2025).

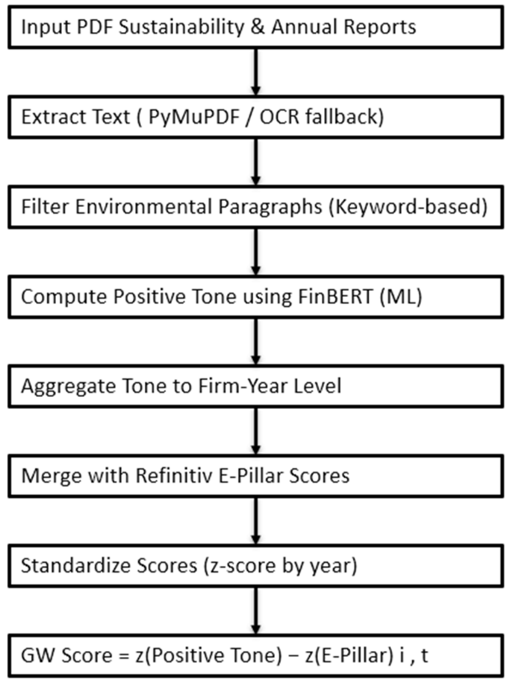
Methodologically, this study uses a firm-year panel covering six fiscal years from FY2019 to FY2024. Based on the Hausman test, the Random Effects (RE) model is chosen as the most suitable estimator, as it allows firm-level variation while if unobserved firm characteristics do not correlate with the independent variables. Year dummy variables are included to control for time-related effects. The years 2022 to 2024 are especially important, as they represent Malaysia's strengthened sustainability reporting environment, making this period useful for examining potential "talk–walk" decoupling under increased ESG disclosure requirements.

3.4 Variable Measurement and Operationalization

3.4.1 Dependent Variables

This thesis's primary methodological contribution is the development of a brand-new composite index to gauge greenwashing. According to Lublóy et al. (2024), this index aims to measure the degree of disconnection between a company's substantive environmental performance and conduct, and its symbolic environmental communications.

Figure 3.1 Construction of Greenwashing Score



According to the workflow in Figure 3.1, the Greenwashing Score is produced using a multi-stage Python workflow that integrates machine-learning sentiment analysis with ESG rating data. The script first sets the file paths and core parameters, including the folder of PDF sustainability reports, the ESG dataset, and the output file. Text is extracted from each PDF using PyMuPDF, with an OCR backup for scanned documents, and then cleaned and filtered so that only environment-related paragraphs remain, based on a comprehensive list of environmental keywords.

Sentiment analysis is performed using FinBERT, a transformer model trained on financial text, which generates a probability that each paragraph expresses a positive tone. Long paragraphs are automatically split into smaller chunks to meet FinBERT's token limits, and results are aggregated to the firm-year level. If FinBERT is unavailable, the script uses VADER as a fallback. This produces the variable PositiveTone, along with a count of environmental words for each firm-year.

ESG data from Refinitiv is then cleaned and standardised, retaining only valid environmental pillar (E-score) observations. After matching firm names across sources, the text-based sentiment data is merged with the E-pillar scores. Z-scores for PositiveTone and E Pillar Score are computed within each year (or industry-year), and the Greenwashing Score is defined as: GWit= z (Positive Tone it) – z (E-pillar Score it).

The Environmental Pillar Score (E-Pillar Score i,t) represents the firm's "perceived walk" as assessed by Refinitiv's standardized environmental rating. A higher score indicates stronger environmental performance and management. In the Greenwashing Score formula, this value is subtracted because a firm with genuinely strong environmental performance should naturally display a positive reporting tone. The difference between tone and score reflects the potential mismatch that the study aims to measure.

A positive score indicates that a firm's environmental "talk" exceeds its environmental "walk," while a negative score suggests stronger performance relative to tone. The final output is a firm-year panel dataset that is saved for subsequent regression analysis.

Table 3.1 Dependent Variable Definition

Variable	Abbreviation	Measurement
Greenwashing Score	GW	GWit= z (Positive Tone it) – z (E-pillar Score it) — where z (·) is the year×industry z-score.

Source: Created for this study.

3.4.2 Independent Variables

3.4.2.1 Environmental Pillar

Environmental Score (E pillar): The primary independent variable for H1. This will be the firm's overall Environmental score from Refinitiv (0-100 score). This variable represents the holistic assessment provided by the rating agency.

The three sub pillars of the environmental dimension are as follows:

Emissions: The independent variable for H1a. This will be the firm's emissions score from Refinitiv (0-100 score). It captures the firm's commitment and effectiveness in reducing environmental emissions across operations or production.

Resource Use: The independent variable for H1b. This will be the firm's resource use score from Refinitiv (0-100 score). It captures the performance and capacity to reduce materials, energy, and water use and to improve eco-efficiency, including supply-chain practices.

Innovation: The independent variable for H1c. This will be the firm's environmental score from Refinitiv (0-100 score). It captures the firm's capacity to lower customers' environmental burdens via modern technologies, processes, or eco-designed products.

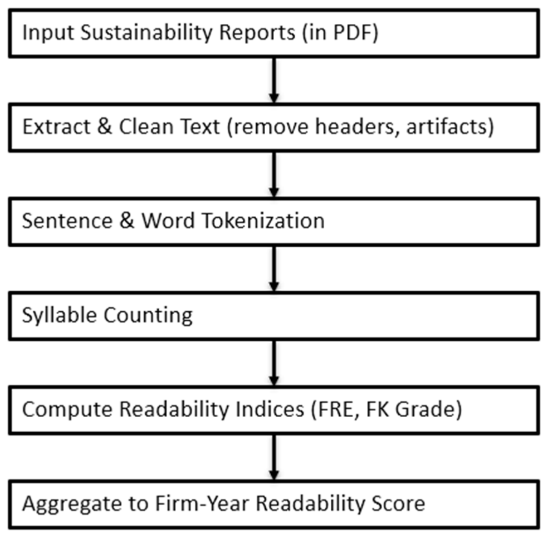
Refinitiv's ESG dataset is widely used in both academic and professional research, making it a reliable and credible source for environmental indicators. Its methodology draws from publicly available, company-reported data (e.g., annual reports, sustainability reports, regulatory filings, and NGO sources), ensuring transparency and replicability. Moreover, the data undergoes annual updates and standardized scoring procedures, which facilitate comparability across firms, industries, and time periods (Refinitiv, 2024). While the reliance on self-reported disclosures may introduce a degree of disclosure bias, Refinitiv remains one of the most comprehensive and validated ESG databases, frequently employed in studies examining ESG performance and greenwashing (Berg et al., 2022; Yu et al., 2020).

3.4.2.2 Disclosure Quality

Disclosure Quality (DQ): To test H2, disclosure quality will be captured by two distinct proxies, reflecting different dimensions of the construct.

3.4.2.2.1 Readability Score

Figure 3.2 Construction of Readability Scores



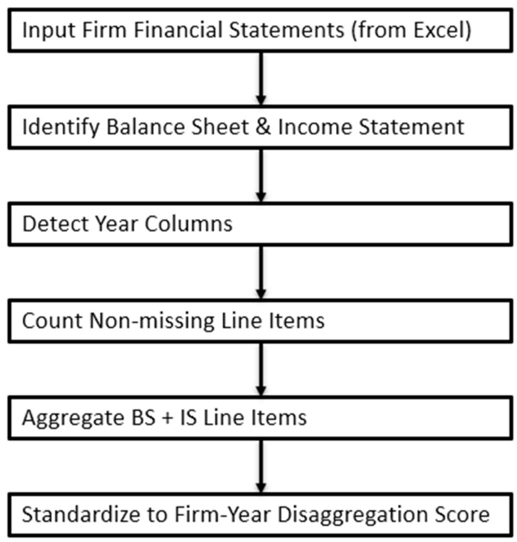
According to the workflow in Figure 3.2, the readability measure in this study is generated using a Python script that extracts, cleans, and analyses text from sustainability reports. Each PDF is processed using PyMuPDF, with an OCR fallback for scanned documents to ensure full text capture. After cleaning the text and breaking it into sentences and words, the script computes standard readability indices through the redability_scores() function, including Flesch Reading Ease**,** Flesch–Kincaid Grade Level**,** Gunning Fog Index**,** SMOG**,** and the Automated Readability Index**.** These metrics are based on sentence length, word count, syllable count, and complex word frequency, which provide an objective measure of disclosure clarity.

The script also allows readability to be calculated using only environment-related paragraphs, identified through keyword filtering, ensuring the measure aligns with the study's focus on environmental reporting. The resulting readability scores are compiled into a firm-year CSV file and used as an independent variable in the regression analysis.

3.4.2.2.2 Disaggregation Score

The second component of disclosure quality, the Disaggregation Score, measures the transparency and detail of a firm's financial reporting. Following Chen et al. (2015), disaggregation is defined as the number of non-missing financial statement line items reported in the balance sheet and income statement. A higher count indicates more detailed reporting and reflects stronger substantive disclosure rather than symbolic communication.

Figure 3.3 Construction of Disaggregation Score



According to Figure 3.3, this measure is automated through a Python script that scans each firm's financial statements which were exported to Microsoft Excel files. These financial statements are taken from the Refinitiv Database. The script identifies the correct statement sheets, locates the header row, and examines each line item, counting it only when the value is numeric and the label is meaningful. Hence, the script excludes items like totals, notes, or other non-informative headings. This ensures the score captures true reporting granularity rather than formatting repetition. For each firm-year, the valid line items across both statements are summed to create the FS_ItemsCount, which is then converted into a z-score.

The resulting Disaggregation Score provides an objective and replicable indicator of disclosure transparency in the greenwashing context. Firms with more detailed financial statements are expected to show less "talk–walk" decoupling because greater disaggregation usually reflects more complete and less selectively curated reporting practices.

3.4.2.3 Environmental Violation

Environmental Violation: This variable serves as the main independent variable for H3. It is coded as a binary indicator, where a value of 1 denotes that the firm is recorded in the Refinitiv database as having faced a significant environmental controversy, regulatory sanction, or fine in year *t*, and 0 otherwise.

Table 3.2 Independent Variables Definition

Variable	Measurement
E pillar	Refinitiv E pillar score (0–100)
Emissions	Refinitiv Emissions sub-score (0–100)
Resource Use	Refinitiv Resource Use sub-score (0–100)
Innovation	Refinitiv Innovation sub-score (0–100)
Readability	Primary: Flesch Reading Ease (FRE) computed for the annual report: FRE = 206.835 – 1.015×(words/sentences) – 84.6×(syllables/word); (higher = clearer).
Disaggregation	Count of non-missing line items in the annual report financial statements (Balance Sheet + Income Statement); sum the unique numeric line items reported (exclude blanks/notes), then convert to a z-score. (Chen–Miao–Shevlin style DQ.)
Violation	Binary dummy = 1 if the firm has ≥1 environmental controversy/penalty in year t (Refinitiv ESG "Environment Controversy" = TRUE), else 0.

Source: Created for this study.

3.4.3 Control Variables

This study controls for firm size, profitability, leverage, and growth to ensure that the estimated effects of environmental performance, disclosure quality, and violations are not confounded by underlying financial characteristics. Firm size (SIZE), measured as the natural log of total assets, is included because larger firms typically disclose more information and have greater capacity to undertake substantive environmental initiatives. Profitability (ROA), calculated as net profit divided by total assets, captures firms' financial strength; more profitable companies may invest in genuine sustainability practices, whereas weaker performers may rely more on symbolic reporting. Leverage (DTA), measured as total debt over total assets, reflects financial pressure. For example, highly leveraged firms may face incentives to use optimistic disclosures when operational performance is constrained. Growth (MTB), defined as the market-to-book ratio, indicates investor expectations; high-growth firms may face stronger pressure to maintain positive narratives, increasing the likelihood of impression management.

These variables align with prior research. Yu et al. (2020) controlled for size, profitability, liquidity, ownership, and board characteristics when examining peer-relative greenwashing, finding that stronger operational performance reduces greenwashing. Treepongkaruna et al. (2024) included firm size, profitability, and leverage when testing the link between Refinitiv E-pillar scores and emissions, highlighting cases where ESG disclosure becomes "cheap talk." Ghitti (2024) similarly incorporated size, profitability, leverage, and market-to-book ratios in greenwashing regressions. Collectively, these studies show that controlling for financial fundamentals is essential to avoid biased results arising from differences in firms' capacity, financial structure, or growth prospects.

Table 3.3 Control Variables Definition

Variable	Abbreviation	Measurement
Firm Size	SIZE	The natural log of total assets
Profitability	ROA	Net Profit/ Total Assets
Leverage	DTA	Total Debt/ Total Assets

Growth	MTB	Market-to-Book ratio = Market Capitalization / Book Value of Equity at fiscal year-end.

Source: Created for this study.

3.5 Formulation of Hypotheses

H1 (E Pillar): Environmental ESG scores (emissions, resource use, and innovation) are negatively associated with greenwashing.

- H1a: Emissions score is negatively associated with greenwashing.
- H1b: Resource use score is negatively associated with greenwashing.
- H1c: Innovation score is negatively associated with greenwashing.

H2a (Readability): Readability score is negatively associated with greenwashing.

H2b (Disaggregation): Disaggregation score is negatively associated with greenwashing.

H3 (Environmental Violations): Environmental violations are positively associated with greenwashing.

3.6 Statistical Model

3.6.1 Panel Multiple Regression Model

This study uses a panel multiple regression model to analyse the determinants of greenwashing among Malaysian listed firms. Panel regression is appropriate because it captures both cross-sectional differences across firms and time-series variation across years, allowing a more comprehensive understanding of disclosure behaviour than a purely cross-sectional or time-series model.

Based on the Hausman Test results, the Random Effects (RE) model is selected as the most suitable estimator. The RE approach assumes that firm-specific unobserved characteristics are not systematically correlated with the explanatory variables, enabling the model to exploit both within-firm and between-firm variation. This makes the RE model more efficient and appropriate for studies such as this, where differences across firms are treated as random rather than fixed attributes.

Year dummies are included in the model to account for time-specific influences such as regulatory changes, reporting standard updates, or macroeconomic conditions. This allows the analysis to isolate whether changes in greenwashing behaviour are associated with firm characteristics rather than broad year-to-year shifts.

Overall, the Random Effects specification provides an efficient and theoretically consistent framework for examining how environmental ratings, disclosure quality, and environmental violations contribute to greenwashing among Malaysian public listed companies.

3.6.2 Model Equation and Variable Definitions

This study's analysis is built on two firm-year panel regression models, Model 1 and Model 2, which assess how environmental performance, disclosure quality, and enforcement indicators influence a firm's greenwashing score (GW). Model 1 uses the overall Environmental I Pillar Score to represent a firm's aggregate environmental performance, whereas Model 2 breaks this score into its three sub-components (Emissions, Resource Use, and Innovation) to determine which specific dimension contributes to greenwashing behaviour. Both models include disclosure variables (readability and disaggregation), environmental violations, and standard firm-level controls such as size, profitability, leverage, and growth. This structure allows the study to evaluate the independent effects of each variable on GW while controlling for unobserved firm traits and year-specific influences.

A POST dummy (0 = FY2019–2021; 1 = FY2022–2024) is included to capture changes in greenwashing behaviour following Malaysia's sustainability reporting upgrade. Interaction terms like (E × POST, Emissions × POST, Resource Use × POST, Innovation × POST) are used to test whether the relationship between environmental performance and greenwashing differs before and after the reform. These interaction effects function as a difference-in-differences–type test, indicating whether the strength or direction of the environmental performance–greenwashing link shifted in the post-upgrade period. Combined, Model 1 and Model 2 provide a comprehensive framework for examining how both aggregate and component-level environmental performance relate to firm-level greenwashing, and whether these relationships remain consistent or change in the later regulatory environment.

Model 1: Main Model Without Disaggregation on E Sub Pillars

Firm Heterogeneity

$Gwi,t = \alpha + \beta 1E\ Pillar\ i,t + \beta 2Readability\ i,t + \beta 3Disaggregation\ i,t + \beta 4Violation\ i,t + \gamma Controlsi,t + \epsilon i,t$

Pre vs. Post Sustainable Reporting Upgrade

$Gwi,t = \alpha + \beta 1E\ Pillar\ i,t + \beta 2Readability\ i,t + \beta 3Disaggregation\ i,t + \beta 4Violation\ i,t + \beta 5Post\ t + \beta 6(E\ Pillar\ i,t \times Post\ t) + \gamma Controls\ i,t + \epsilon i,t$

Model 2: Main Model with Disaggregation on E Sub Pillars

Firm Heterogeneity

$Gwi,t = \alpha + \beta 1Emissionsi,t + \beta 2ResourceUsei,t + \beta 3Innovationi,t + \beta 4Readabilityi,t + \beta 5Disaggregationi,t + \beta 6Violationi,t + \gamma Controlsi,t + \epsilon i,t$

Pre vs. Post Sustainable Reporting Upgrade

$Gwi,t = \alpha + \beta 1Emissionsi,t + \beta 2ResourceUsei,t + \beta 3Innovationi,t + \beta 4Readabilityi,t + \beta 5Disaggregationi,t + \beta 6Violationi,t + \beta 7Post\ t + \beta 8(Emissionsi,t \times Post\ t) + \beta(ResourceUse\ i,t) + \beta(EnvInnovation\ i,t) + \gamma Controls\ i,t + \epsilon i,t$

Whereas:

- Gwi,t = Greenwashing Score for firm i in year t (dependent variable)
- $E\ Pillar\ i,t$ = Environmental pillar score

- *Emissionsi,* = Emissions score
- *ResourceUse i,t* = Resource Use Score
- *Innovation i,t* = Innovation Score
- *Readability i,t* = linguistic readability of firm sustainability report (FRE)
- *Disaggregation i,t* = granularity of reported financial metrics (from Financial Statement/ Income Statement)
- *Violation i,t* = binary indicator of environmental violations
- Controls include firm size (log of total assets), profitability (ROA), leverage (debt-to-assets), and growth (Market-to-book)
- ϵ i,t = error term
- *Post t* = The presence of Malaysia's sustainability reporting upgrade

3.6.3 Software Used

This study uses Eviews and Python to analyse both numerical and textual data. Eviews is applied for panel regression, diagnostic tests, and hypothesis testing, making it suitable for handling large firm-year datasets and comparing fixed and random effects through the Hausman test. Python is used to process unstructured text from annual and sustainability reports, applying NLP techniques such as tokenization, sentiment analysis, and readability scoring through libraries like NLTK, textstat, and VADER. Python also computes disclosure disaggregation by counting financial line items and quantitative metrics. Microsoft Excel supports initial data cleaning and descriptive statistics. All Python scripts and Eviews command files are archived to ensure transparency and allow future researchers to replicate or extend the analysis.

3.7 Data Analysis Plan

3.7.1 Descriptive Analysis

Descriptive analysis refers to a set of statistical techniques used to summarise, organise, and present the main characteristics of a dataset. It commonly includes measures of central tendency (mean, median, mode), measures of variability (standard deviation, range), and indicators of distributional shape such as skewness and kurtosis, together with simple visualisations and tabulations (e.g., frequencies, percentages, histograms, cross-tabulations). The purpose of descriptive analysis is to provide a clear understanding of the sample structure and variable behaviour, assist in preliminary data screening and establish a foundation for subsequent inferential modelling without making causal interpretations. In business and finance research, descriptive statistics typically report groupwise means and standard deviations for continuous variables, as well as counts and percentages for categorical variables, often segmented by year or industry to highlight trends and patterns before conducting regression analysis (Qu, 2007; Field, 2013; Creswell & Creswell, 2018).

For this study, descriptive statistics will be presented in a comprehensive table summarising the key variables of interest. The table will report the number of firm-year observations, mean, median, standard deviation, minimum, and maximum values for the dependent, independent, and control variables. These include the Greenwashing Score; the Environmental Pillar Score and its three sub-pillars (Emissions, Resource Use, and Innovation); the Readability and Disaggregation Scores; and Environmental Violations. The table will also summarise the control variables, including Firm Size (log of total assets), Profitability (ROA), Leverage (debt-to-assets), and Growth (market-to-book ratio). Presenting these statistics offers a transparent overview of the dataset and helps identify any irregularities or data characteristics that should be considered before proceeding with regression analysis.

3.7.2 Pearson Correlation Analysis

Pearson correlation analysis is a statistical technique used to assess the strength and direction of the linear relationship between two continuous variables (Cohen, 2013). The correlation coefficient, denoted as r, ranges from –1 to +1. A value close to +1 indicates a strong positive relationship, meaning both variables increase together, while a value close to –1 reflects a strong negative relationship, where one variable increases as the other decreases. Values near 0 suggest little or no linear association between the variables, indicating that changes in one variable do not systematically correspond to changes in the other (Field, 2013).

Table 3.4 Pearson Correlation Range of Coefficient

Range of Coefficient		Description of Correlation Strength
From	To	
+/- 0.81	+/- 1.00	Very Strong
+/- 0.61	+/- 0.80	Strong
+/- 0.41	+/- 0.60	Moderate
+/- 0.21	+/- 0.40	Weak
+/- 0.00	+/- 0.20	Weak to No Correlation

Source: Hair, Jr., Celsi, Oritinau & Bush (2013)

3.7.3 Diagnostics Test

Before conducting regression analysis, several diagnostic checks are required to ensure the robustness and validity of the models.

First, multicollinearity will be assessed using the Variance Inflation Factor (VIF). Generally, VIF values above 10 indicate serious multicollinearity, though some scholars recommend a more conservative threshold of 5 (Gujarati & Porter, 2009). Any variable that has the VIF of more than 10 will be examined closely, and remedial steps, like variable transformation or removal, will be considered where appropriate.

Second, heteroskedasticity will be tested using the Breusch–Pagan and White tests. A p-value below 0.05 will lead to rejection of the null hypothesis of homoskedasticity, suggesting that the variance of the error terms is not constant. If heteroskedasticity is detected, the study will employ firm-clustered robust standard errors to obtain consistent and unbiased estimates (Wooldridge, 2010).

Third, autocorrelation will be evaluated using the Breusch–Godfrey Serial Correlation LM test for panel data. A p-value < 0.05 indicates that the residuals exhibit serial dependence, violating the assumption of no autocorrelation (Baltagi, 2021). When serial correlation or heteroskedasticity is present, the study will apply Panel-Corrected Standard Errors (PCSE), which provide more reliable inference in the presence of correlated and heteroskedastic error structures (Hoechle, 2007).

Finally, the normality of residuals will be examined using the Jarque–Bera test. A p-value less than 0.05 signals rejection of the null hypothesis of normality, indicating potential non-normal residuals (Jarque & Bera, 1987). Although deviations from normality are less concerning in larger samples due to asymptotic properties, the test remains useful for identifying outliers and possible model misspecifications.

Taken together, these diagnostic procedures ensure that the regression models satisfy key econometric assumptions, and that any violations are addressed using appropriate robust techniques, thereby preserving the credibility and reliability of the empirical results.

3.8 Ethical Considerations

This study relies entirely on secondary data obtained from publicly available sources, such as company websites for annual and sustainability reports, while ESG data, including Environmental Performance and Environmental Violations, are sourced from the Refinitiv ESG database, which is provided through the university's institutional subscription. Because no primary data are collected and no human participants are involved, issues such as informed consent, confidentiality, or respondent protection do not apply.

Despite relying on public information, the study follows established principles of data privacy and responsible use. All firm-level data are drawn from documents already disclosed for regulatory, compliance, or transparency purposes, and are used solely for academic, non-commercial analysis. Proprietary databases such as Refinitiv are accessed under institutional licensing agreements, and any data reproduced in the thesis are presented only in aggregated or statistical form. No confidential, undisclosed, or sensitive information is accessed or shared. By using publicly available and properly licensed datasets and by ensuring accurate and ethical handling of corporate information, the study complies fully with research ethics standards and upholds the integrity of corporate disclosures.

3.9 Summary

Chapter 3 outlined the methodological framework of the study. It explained that while the original sample included the Top 100 non-financial listed firms, the final panel dataset consists of **82 firms** after removing observations with missing data for the 2019–2024 period. The chapter detailed the sample selection, data sources, and variable construction, with the Greenwashing Score as the dependent variable and environmental pillar including sub-pillars, disclosure quality measures, and environmental violations as key independent variables.

Control variables consists of firm size, profitability, leverage, and growth are also defined. The statistical approach used panel regression with firm and time effects, supported by diagnostic tests to ensure robustness. Eviews and Python were employed to analyse both structured financial data and unstructured textual disclosures. Ethical considerations and data replicability were also addressed.
Overall, the methodology provides a rigorous and transparent foundation for examining greenwashing determinants in Malaysia. Chapter 4 will present the empirical results, while Chapters 5 and 6 will discuss the findings, draw conclusions, outline implications, and suggest future research directions.

Chapter 4: Data Analysis

4.1 Introduction

This chapter presents the empirical analysis undertaken to examine the determinants of greenwashing among Malaysian public listed companies. Consistent with quantitative research practice, the chapter begins with a descriptive analysis that summarises the central tendencies, dispersion, and distributional characteristics of the dataset, thereby enabling a clearer understanding of the variables prior to model estimation (Loeb et al., 2017). This is followed by a correlation analysis, which provides preliminary insights into the linear relationships among the variables and assists in identifying potential multicollinearity concerns that may influence the validity of subsequent regression results (Gujarati & Porter, 2009).

The chapter then proceeds with a series of diagnostic tests designed to determine the appropriate panel estimator and to verify key econometric assumptions. These diagnostics include tests for heteroskedasticity, serial correlation, and correct model specification. Conducting these assessments is critical for ensuring reliable estimation and valid statistical inference within panel data settings (Baltagi, 2021). After establishing the suitability of the econometric approach, the chapter presents regression results for both the baseline models and the extended specifications that incorporate interaction terms between environmental indicators and the post-2022 sustainability reporting period. These models collectively allow for a comprehensive evaluation of how environmental performance, disclosure quality, and environmental violations influence the extent of greenwashing behaviour among Malaysian firms across time.

In short, this chapter provides the full empirical foundation for answering the study's research questions and testing the hypotheses developed earlier.

4.2 Descriptive Analysis

Descriptive analysis is employed to transform raw data into meaningful statistical summaries that offer clearer insight into the characteristics and patterns within the dataset. This approach allows the study to provide an organised overview of the variables, illustrate their distributions, and establish the foundational understanding needed for subsequent empirical analysis (Loeb et al., 2017). Accordingly, the descriptive statistics for all variables used in this study are reported in Table 3, providing a comprehensive snapshot of the sample's central tendencies, variability, and overall data behaviour.

Table 4.1 Descriptive Analysis

Variables	Mean	Median	Std. Deviation	Min.	Max.	Observation
GW	-0.0384	-0.0834	1.3175	-3.2734	5.6530	369
E	49.49	51.46	20.85	0.75	94.72	369
Emissions	58.51	60.79	25.21	0.31	99.72	369
Resource Use	54.81	58.33	23.89	0.7093	95.33	369
Innovation	28.27	18.12	28.23	0.0200	98.09	369
Readability	25.82	26.48	6.4385	0.4308	42.23	369
Disaggregation	0.1968	0.2635	0.9196	-3.1848	2.5097	369
Violation	0.0705	0.0000	0.2562	1	0.0000	369
SIZE	8.9994	9.0408	1.4125	5.5356	12.2353	369
DTA	0.4711	0.4584	0.2191	0.0537	1.4776	369
MTB	4.5653	2.0125	4.2954	-0.8285	58.46	369
ROA	0.0699	0.0463	2.5218	-0.2572	0.7882	369

Table 4.1 presents the descriptive statistics for all variables used in the analysis with the total of 369 firm-year observations. The Greenwashing Score (GW) has a mean of –0.0384 and a standard deviation of 1.3175, indicating considerable variation in the degree of talk–walk decoupling across firms. The minimum value of –3.2734 and maximum of 5.6530 show that while some firms demonstrate strong alignment between their environmental disclosures and actual performance, others exhibit notably high levels of greenwashing.

For the environmental indicators, the E Pillar shows a mean score of 49.49, with its sub-pillars averaging 58.51 for Emissions, 54.81 for Resource Use, and a substantially lower 28.27 for Innovation. The relatively large standard deviations across these measures indicate marked heterogeneity in environmental performance among Malaysian firms. The Innovation sub-pillar reports a minimum value near zero, reflecting that many firms undertake minimal or no innovation activities.

The disclosure-related variables also exhibit considerable variation. Readability has a mean of 25.82 with moderate dispersion, suggesting meaningful differences in the linguistic complexity of sustainability disclosures. Disaggregation displays a mean near zero but a high standard deviation (0.9196), indicating wide differences in the granularity of line-item reporting across firms. Environmental Violations show a mean value of 0.0705, meaning roughly 7% of firm-year observations involve recorded violations; the median of zero confirms that most firms do not experience violations within the sample period.

Regarding the control variables, SIZE has a mean of 8.9994, consistent with the sample comprising large publicly listed firms. The Debt-to-Asset ratio (DTA) has a mean of 0.4711, suggesting moderate leverage levels on average. The Market-to-Book ratio (MTB) shows substantial dispersion, indicating significant variation in market valuation across firms. Finally, ROA records a mean of 0.0699 but a large standard deviation (2.5218), highlighting considerable differences in profitability, including the presence of loss-making firms in certain years.

4.3 Correlation Analysis

The Pearson correlation analysis was conducted to examine the linear relationships among the variables used in this study, providing an initial assessment of how each independent variable relates to the dependent variable before proceeding to regression analysis. This step is essential in quantitative research as it helps identify the strength and direction of pairwise associations, detect potential multicollinearity issues, and ensure that the variables behave in a manner consistent with theoretical expectations. By analysing these preliminary correlations, the study gains clearer insight into the underlying data patterns and establishes a foundation for interpreting the subsequent panel regression results.

Table 4.2 Pearson Correlation Matrix

	GW	E	EMISS	RESUSE	INO	READ	DISAGG	VIO	Size	DTA	MTB	ROA
GW	1.0000											
E	-0.5417	1.0000										
EMISS	-0.3965	0.7650	1.0000									
RESUSE	-0.5190	0.7672	0.5473	1.0000								
INO	-0.2740	0.4686	0.2120	0.3059	1.0000							
READ	-0.0046	-0.1155	-0.1480	-0.1106	-0.0704	1.0000						
DISAG	-0.1287	0.2867	0.1994	0.1962	0.2160	-0.0722	1.0000					
VIOLAT	-0.0752	-0.0060	0.0016	0.0077	0.0541	-0.0041	0.1393	1.0000				
SIZE	-0.2099	0.4338	0.3769	0.2899	0.3502	-0.3218	0.6096	0.1613	1.0000			
DTA	-0.1287	0.1165	0.0651	0.0622	0.0488	-0.0451	0.2298	0.0709	0.2320	1.0000		
MTB	-0.1787	0.0101	-0.0070	0.0318	0.0110	-0.0227	-0.1506	-0.0224	-0.3405	0.2514	1.0000	
ROA	-0.0769	-0.0706	-0.0362	-0.0474	-0.1130	0.0829	-0.1889	0.1121	-0.3728	-0.2216	0.4646	1.0000

Table 4.2 reports the Pearson correlation coefficients for all variables used in the analysis. The Greenwashing Score (GW) exhibits moderate negative correlations with the Environmental Pillar (r = –0.5417), Emissions Score (r = –0.3965), and Resource Use (r = –0.5190). These results suggest that firms with stronger environmental performance generally display lower levels of greenwashing, aligning with expectations that substantive environmental practices reduce the extent of talk–walk decoupling. Innovation also shows a weaker negative correlation with GW (r = –0.2740), indicating that innovation-related efforts are only modestly associated with reductions in greenwashing behaviour.

The environmental sub-pillars (E, Emissions, and Resource Use) show strong positive correlations with one another, ranging from 0.5473 to 0.7672. This pattern is expected, as these measures collectively capture related dimensions of environmental performance. Although the correlations are moderately high, they remain below the conventional multicollinearity threshold of 0.80, suggesting that these variables can be included together in regression models without inflating standard errors.

The disclosure quality variables exhibit weak associations with GW. Readability is effectively uncorrelated with GW (r = –0.0046), while Disaggregation shows a small negative correlation (r = –0.1287). These findings indicate that disclosure clarity and reporting granularity do not have a meaningful bivariate relationship with greenwashing, consistent with the regression results in which these variables were statistically insignificant.

Environmental Violation displays a weak negative correlation with GW (r = –0.0752), reflecting the limited distribution of violations across firms and indicating minimal direct linear association.

Among the control variables, Firm Size demonstrates a moderate positive correlation with the environmental indicators (e.g., r = 0.4338 with E), suggesting that larger firms tend to achieve higher environmental scores, possibly due to greater resources and reporting capacity. Size also shows a moderate negative correlation with ROA (r = –0.3728), indicating that the larger firms in the sample tend to be less profitable on average. Debt-to-Asset (DTA) and Market-to-Book (MTB) reveal only weak correlations with most variables, suggesting limited direct association. ROA shows a moderate positive correlation with MTB (r = 0.4646), consistent with the expectation that more profitable firms receive higher market valuations.

Crucially, none of the correlations among the independent variables exceed 0.80, and most fall below 0.60, confirming that multicollinearity is not a significant concern in the dataset.

4.4 Diagnostic Tests

To ensure the reliability and validity of the empirical results, a series of diagnostic tests were conducted on the final regression specifications Model 1 and Model 2, which include the full set of variables such as the environmental pillar or its sub-pillars, disclosure quality measures, environmental violations, and the pre–post pandemic dummy with interaction terms. Performing diagnostics only on the final specification is consistent with standard econometric practice, as it ensures that decisions regarding model selection, estimator appropriateness, and inference accuracy are based on the complete empirical framework rather than on partial or preliminary models.

4.4.1 Hausman Test

The Hausman test was conducted to determine whether the fixed effects (FE) or random effects (RE) estimator is more suitable for the final model. The null hypothesis states that the random effects estimator is consistent and efficient; therefore, failing to reject the null supports the use of RE. Conversely, if the null is rejected, it indicates that the regressors are correlated with unobserved firm-specific effects, making the fixed effects (FE) estimator the appropriate choice.

Figure 4.1 Hausman Test Result for Model 1

Correlated Random Effects - Hausman Test
Equation: Untitled
Test cross-section random effects

Test Summary	Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.
Cross-section random	8.117750	8	0.4221

Cross-section random effects test comparisons:

Variable	Fixed	Random	Var(Diff.)	Prob.
E	-0.923742	-0.965826	0.023902	0.7855
READ	-0.135268	-0.190417	0.004745	0.4234
DISAGGREGATION	0.108853	0.102341	0.009545	0.9469
VIOLATION	-0.272196	-0.300237	0.002740	0.5922
SIZE	0.058057	-0.090488	0.097732	0.6347
DTA	-1.072364	-0.430084	0.541853	0.3829
MTB	0.007913	-0.011227	0.000143	0.1091
ROA	-1.593277	-1.338839	0.095282	0.4098

Based on Figure 4.1, (χ^2 = 8.1178, p-value = 0.4221), the null hypothesis is not rejected. Since the p-value is greater than 0.05, the results indicate that the random effects estimator is appropriate for the final model.

Figure 4.2 Hausman Test Result for Model 2

Correlated Random Effects - Hausman Test
Equation: Untitled
Test cross-section random effects

Test Summary	Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.
Cross-section random	12.941110	10	0.2270

Cross-section random effects test comparisons:

Variable	Fixed	Random	Var(Diff.)	Prob.
EMISS	-0.331222	-0.266201	0.005676	0.3881
RESUSE	-0.450561	-0.577607	0.008258	0.1621
ENVINO	0.002749	-0.041728	0.000783	0.1120
READ	-0.146972	-0.226493	0.004920	0.2569
DISAGGREGATION	0.069298	0.067230	0.009848	0.9834
VIOLATION	-0.250134	-0.242922	0.003069	0.8964
SIZE	0.153347	-0.103534	0.104448	0.4267
DTA	-1.031478	-0.488350	0.560628	0.4682
MTB	0.008822	-0.011586	0.000146	0.0913
ROA	-1.569143	-1.322099	0.096145	0.4256

Based on Figure 4.2, ($\chi^2 = 12.9411$, p-value = 0.2270), the null hypothesis is not rejected. Since the p-value is greater than 0.05, the results indicate that the random effects estimator is appropriate for the final model.

4.4.2 Test for Heteroskedasticity

Panel datasets often exhibit heteroskedasticity because firms differ in size, disclosure practices, and reporting behaviour (Landstrom, 2019). To test for heteroskedasticity in the final model, the Breusch–Pagan test and the White test (for RE specifications) were employed. The null hypothesis for both tests assumes homoscedasticity, whereas rejection of the null indicates the presence of heteroskedasticity in the error structure.

Figure 4.3 Breusch-Pagan-Godfrey Test Result for Model 1

Heteroskedasticity Test: Breusch-Pagan-Godfrey
Null hypothesis: Homoskedasticity

F-statistic	1.146721	Prob. F(8,360)	0.3311
Obs*R-squared	9.169447	Prob. Chi-Square(8)	0.3282
Scaled explained SS	19.54416	Prob. Chi-Square(8)	0.0122

For Model 1, the BPG test result ($\chi^2 = 0.4782$, p-value = 0.4838) indicates that heteroskedasticity is present in the model.

Figure 4.4 Breusch-Pagan-Godfrey Test Result for Model 2

Heteroskedasticity Test: Breusch-Pagan-Godfrey
Null hypothesis: Homoskedasticity

F-statistic	0.990580	Prob. F(10,358)	0.4511
Obs*R-squared	9.935266	Prob. Chi-Square(10)	0.4462
Scaled explained SS	18.65950	Prob. Chi-Square(10)	0.0448

For Model 2, the BPG test result, $\chi^2 = 18.6595$, p-value = 0.0448, indicates that heteroskedasticity is present in the model.

Figure 4.5 White Test Result for Model 1

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	1.247569	Prob. F(43,325)	0.1473
Obs*R-squared	52.27894	Prob. Chi-Square(43)	0.1569
Scaled explained SS	111.4296	Prob. Chi-Square(43)	0.0000

For Model 1, the White test result (χ^2 = 111.4296, p-value = 0.0000) indicates that heteroskedasticity is **present** in the model.

Figure 4.6 White Test Result for Model 2

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	1.311730	Prob. F(64,304)	0.0700
Obs*R-squared	79.84986	Prob. Chi-Square(64)	0.0873
Scaled explained SS	149.9667	Prob. Chi-Square(64)	0.0000

For Model 2, the White test result, χ^2 = 149.9667, p-value = 0.0000, indicates that heteroskedasticity is **present** in the model.

4.4.3 Test for Serial Correlation

According to Landstrom (2019), serial correlation in the error terms violates the assumption of independently distributed disturbances and can result in biased standard errors and unreliable inference. To assess whether autocorrelation is present in the panel dataset, the Breusch–Godfrey Lagrange Multiplier (LM) Serial Correlation Test for panel data was applied. This test is suitable for random-effects models and is available in EViews, making it appropriate for the final model used in this study. The null hypothesis states that there is no serial correlation in the residuals, while rejection of the null indicates the presence of serial correlation.

Figure 4.7 Serial Correlation LM Test Result for Model 1

Breusch-Godfrey Serial Correlation LM Test:
Null hypothesis: No serial correlation at up to 2 lags

F-statistic	37.98696	Prob. F(2,358)	0.0000
Obs*R-squared	64.59922	Prob. Chi-Square(2)	0.0000

From Figure 4.7, the test result, F-statistic = 37.9870, p-value = 0.0000) indicates that the null hypothesis is rejected. A p-value below 0.05 provides evidence of first-order serial correlation, while a p-value above 0.05 suggests that serial correlation is not present in the panel residuals.

Figure 4.8 Serial Correlation LM Test Result for Model 2

Breusch-Godfrey Serial Correlation LM Test:
Null hypothesis: No serial correlation at up to 2 lags

F-statistic	40.38220	Prob. F(2,356)	0.0000
Obs*R-squared	68.23373	Prob. Chi-Square(2)	0.0000

From Figure 4.8, the test result, F-statistic = 40.3822, p-value = 0.0000 indicates that the null hypothesis is rejected. A p-value below 0.05 provides evidence of first-order serial correlation, while a p-value above 0.05 suggests that serial correlation is not present in the panel residuals.

Diagnostic tests confirmed the presence of both heteroskedasticity and serial correlation in the final regression specifications. These violations undermine the reliability of standard error estimates in conventional panel estimators. To ensure robust inference, the study adopts Panel-Corrected Standard Errors (PCSE) following the cross-section SUR covariance structure. PCSE is widely recommended for panels exhibiting heteroskedasticity and contemporaneous correlation across units, as it provides more efficient and consistent standard errors compared to traditional approaches (Beck & Katz, 1995). By applying PCSE, the study mitigates the distortions caused by these error-structure issues and strengthens the validity of the estimated coefficients.

4.4.5 Multicollinearity Test (Variance Inflation Factor – VIF)

Multicollinearity was assessed using Variance Inflation Factor (VIF) values computed for all independent variables in the final model. In accordance with established guidelines, VIF values exceeding 10 indicate severe multicollinearity, whereas values between 5 and 10 reflect moderate but generally acceptable levels of correlation among predictors (Gujarati & Porter, 2009; Hair et al., 2010).

Table 4.3 Multicollinearity Test Result

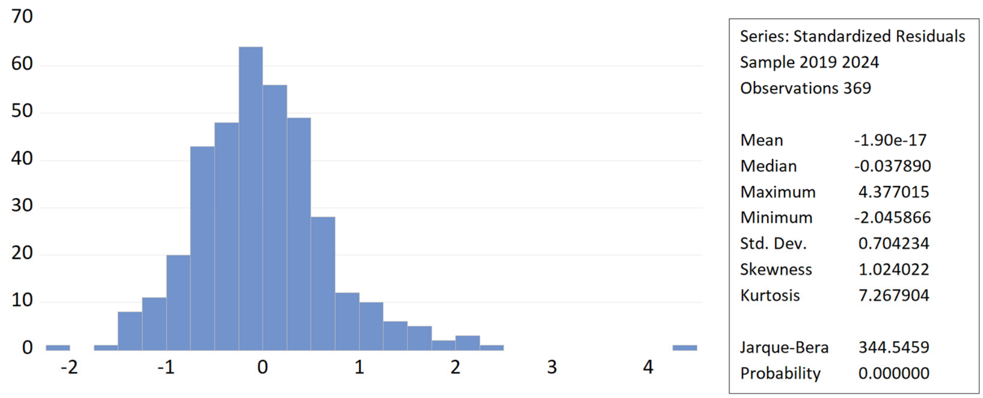
Variables	VIF
E	5.4425
Emissions	2.7223
Resource Use	2.5239
Innovation	1.5066
Readability	1.2056
Disaggregation	1.6936
Environmental Violation	1.0960
Size	2.8351
Debt-to-asset	1.4418
Market-to-book	1.8477
ROA	1.6432

The VIF results indicate that all variables fall below the accepted threshold (maximum VIF = 10), demonstrating that multicollinearity is not a concern in the final model. This suggests that the regression coefficients are stable and not inflated due to excessive correlation among the predictors, supporting the reliability of the estimated relationships.

4.4.6 Normality Test (Jarque Bera Test)

The Jarque–Bera (JB) test was applied to evaluate whether the residuals of the regression model follow a normal distribution, which is an underlying assumption of ordinary least squares estimation. The test assesses deviations from normality by comparing the skewness and kurtosis of the residuals with those expected under a normal distribution (Jarque & Bera, 1980). A statistically significant JB statistic indicates that the residuals are not normally distributed, signalling a violation of the normality assumption and reinforcing the need for robust standard error adjustments to maintain valid statistical inference.

Figure 4.9 Jarque Bera Test Results



The normality of residuals was examined using both the histogram of standardized residuals and the Jarque–Bera (JB) test. The histogram displays a distribution that is broadly bell-shaped but shows clear positive skewness (1.0240) and elevated kurtosis (7.2679), indicating heavier tails than would be expected under normality. This visual assessment is supported by the JB test result, which reports a statistic of 344.5459 with a p-value of 0.0000. Since the p-value falls below the 0.05 significance level, the null hypothesis of normally distributed residuals is rejected (Jarque & Bera, 1980), confirming significant deviations from normality. Although non-normal residuals do not bias coefficient estimates in large samples, they compromise the reliability of conventional standard errors, thereby justifying the use of robust estimation techniques to ensure valid inference.

4.5 Model Results

4.5.1 Interpretation of Model Results

Table 4.4 Regression Results for Model 1 and Model 2

Variables	Model 1	Pre vs. Post	Model 2	Pre vs. Post
E Pillar	-0.9658 (0.0000)***	-1.0264 (0.0000)***		
E Pillar× Post		0.1520 (0.4010)		
Emissions			-0.2662 (0.0001) ***	-0.1842 (0.0203) **
Emissions× Post				-0.2690 (0.0091) ***

Resource Use			-0.5776 (0.0000) ***	-0.5445 (0.0014) ***
Resource Use× Post				0.2522 (0.0368) **
Innovation			-0.0417 (0.3020)	0.0489 (0.3530)
Innovation× Post				-0.0212 (0.7151)
Readability Score	-0.1904 (0.2465)	-0.2326 (0.1995)	-0.2264 (0.1749)	-0.2156 (0.2680)
Disaggregation Score	0.1023 (0.3571)	0.0861 (0.4324)	0.0672 (0.5342)	0.0373 (0.7799)
Environmental Violation	-0.3002 (0.1908)	-0.2903 (0.2030)	-0.2429 (0.2932)	-0.2131 (0.3174)
Post		-0.6821 (0.2995)		-0.0445 (0.9271)
Size	-0.0905 (0.1985)	-0.1097 (0.1141)	-0.1035 (0.1268)	0.3398 (0.2193)
Debt-to-asset	-0.4301 (0.1484)	-0.3749 (0.1924)	-0.4884 (0.1075)	-1.1823 (0.1860)
Market-to-book	-0.0112 (0.1347)	-0.0154 (0.0439) **	-0.0116 (0.0817)	0.0119 (0.3005)
Return-on-asset	-1.3388 (0.0451) **	-1.2980 (0.0558) *	-1.3221 (0.0546) **	-1.8498 (0.0120) **
Constant	5.3637 (0.0000) ***	5.9601 (0.0000) ***	4.3209 (0.0000) ***	5.7581 (0.0000)***
Specification and Diagnostic Tests				
R-squared	0.3338	0.3433	0.3383	0.3424
No. of Observation	369	369	369	369
Estimator	RE – PCSE	RE – PCSE	RE – PCSE	RE – PCSE

Notes: ***, **, * denote 1%, 5%, and 10% significance levels.

"POST is added as a time dummy (1 = post-pandemic, 0 = pre-pandemic) to capture any overall shift in greenwashing after COVID-19. The interaction term (E × POST) tests whether the effect of environmental performance on greenwashing differs between the pre- and post-pandemic periods."

Table 4.4 reports the regression results for Model 1 and Model 2, both estimated using the Random Effects model with Panel-Corrected Standard Errors (RE–PCSE). This estimator was chosen because diagnostic tests confirmed the presence of serial correlation, as evidenced by the significant Breusch–Godfrey LM test (p = 0.0000). The models are based on firm-year observations from 2019 to 2024 and include all specified control variables, along with the POST dummy variable, which captures changes in greenwashing behaviour following Malaysia's 2022 sustainability reporting enhancement.

In Model 1, the Environmental Pillar Score exhibits a highly significant negative association with the Greenwashing Score ($\beta = -0.9658$, $p < 0.001$). This suggests that firms with stronger environmental performance demonstrate less talk–walk decoupling, aligning with the expectation that substantive environmental actions reduce the need for symbolic disclosure. Conversely, the interaction term between the Environmental Pillar and the POST variable is statistically insignificant ($\beta = 0.1520$, $p = 0.4010$). This indicates that Malaysia's 2022 sustainability reporting upgrade did not materially alter the relationship between environmental performance and greenwashing behaviour.

Model 2, which decomposes the Environmental Pillar into its three underlying sub-components, offers more exquisite insights. The Emissions Score shows a significantly negative relationship with the Greenwashing Score both before the regulatory change ($\beta = -0.2662$, $p = 0.0001$) and after 2022 through its interaction term ($\beta = -0.2690$, $p = 0.0091$). This consistent pattern indicates that stronger emissions performance is associated with lower greenwashing across both policy environments, reinforcing the view that emissions metrics capture substantive environmental effort. Resource Use likewise demonstrates a strong negative association with greenwashing ($\beta = -0.5776$, $p < 0.001$). However, unlike Emissions, its interaction with the POST variable becomes positive and significant ($\beta = 0.2522$, $p = 0.0368$), suggesting that following the reporting reforms, higher Resource Use scores may be linked with increased greenwashing, possibly reflecting emerging decoupling or strategic disclosure responses. Finally, Innovation remains statistically insignificant in both its main effect ($\beta = -0.0417$, $p = 0.3020$) and interaction term ($\beta = -0.0212$, $p = 0.7151$), indicating that environmental innovation does not play a meaningful role in explaining firm-level greenwashing behaviour within the Malaysian context.

The disclosure quality variables, including readability and disaggregation, are statistically insignificant across all model specifications. Readability coefficients range from -0.1904 to -0.2264 (all $p > 0.17$), while disaggregation coefficients range from 0.1023 to 0.0373 (all $p > 0.35$). These results indicate that differences in narrative clarity or financial statement granularity do not meaningfully influence greenwashing behaviour once environmental performance and firm characteristics are considered. This outcome may reflect Malaysia's disclosure-driven ESG environment, where firms adhere to mandated reporting templates but do not substantially differentiate themselves through enhanced readability or detailed reporting.

Environmental Violations also exhibit negative but consistently insignificant coefficients across all models, despite theoretical expectations that firms facing regulatory sanctions would engage in greater symbolic disclosure. This insignificance may stem from limited variation in violation observations, or the binary nature of the variable, which may not fully capture the severity or frequency of misconduct.

Among the control variables, only Return-on-Assets (ROA) demonstrates a consistently significant negative association with greenwashing (p-values between 0.012 and 0.055). This indicates that more profitable firms demonstrate closer alignment between their environmental disclosures and actual environmental performance. Other controls variables like firm size, leverage, and market-to-book ratio do not show statistically meaningful relationships with greenwashing.

4.5.2 Justification for the Absence of Adjusted R^2 in Random Effects Estimation

Random Effects (RE) models do not report an Adjusted R^2 because the estimator is based on Generalized Least Squares (GLS) rather than Ordinary Least Squares (OLS). Adjusted R^2 is an OLS-specific goodness-of-fit measure that penalizes the addition of regressors to correct for model complexity (Wooldridge, 2010). In RE estimation, however, the error term is decomposed into between-entity and within-entity components, fundamentally altering the variance structure used in model estimation (Baltagi, 2021). This GLS transformation breaks the assumptions underpinning the standard Adjusted R^2 formula, making it statistically inappropriate for RE models.

For this reason, econometric software such as Stata and Eviews typically report only a "pseudo R^2" or "overall R^2," which represents the squared correlation between the GLS-weighted fitted values and the actual dependent variable (Greene, 2018). This statistic cannot be adjusted using the conventional OLS formula because both the denominator and penalty components rely on assumptions that do not hold under GLS estimation (Hsiao, 2014). As a result, Adjusted R^2 is intentionally omitted for RE models, and researchers instead rely on panel-specific diagnostics to evaluate model validity and explanatory power (Barrett, 2012).

In summary, Adjusted R^2 is not reported for Random Effects models because the statistical foundations required to compute it are incompatible with GLS-based RE estimation. Therefore, the overall (pseudo) R^2 serves only as a descriptive fit metric, while diagnostic tests provide the appropriate basis for assessing model adequacy.

4.6 Hypotheses Testing

The results from both models show that environmental performance is the strongest predictor of greenwashing, while disclosure quality and environmental violations play a limited role. This pattern generally supports the conceptual framework and aligns with past studies showing that firms with stronger environmental outcomes rely less on symbolic reporting.

For environmental performance, the hypotheses H1 which is the Environmental Pillar has a strong negative effect on greenwashing is supported, consistent with Delmas & Burbano (2011) and Clarkson et al. (2008). For the sub-pillars, H1a (Emissions) and H1b (Resource Use) are both supported, and these effects remain significant across all periods. This agrees with research showing that emissions control and resource efficiency reflect substantive environmental commitment (Testa et al., 2023). Resource Use is the strongest predictor. However, H1c (Innovation) is not supported. This aligns with studies suggesting that innovation-related disclosures are often long-term or symbolic rather than performance-driven (Marquis & Qian, 2013).

For disclosure quality, both hypotheses H2a and H2b are rejected. Readability and disaggregation do not significantly influence greenwashing. This matches Malaysian studies noting that sustainability texts are highly standardised (Hu et al., 2024; Kishan & Azhar, 2025) and contrasts with Western findings where clarity signals transparency (Li, 2008).

The hypothesis H3 is not supported. Violations do not predict higher greenwashing, contradicting evidence from other countries showing that violations expose symbolic reporting (Kim & Lyon, 2014). The insignificance may reflect limited or underreported violation data in Malaysia.

All interaction terms are insignificant, meaning the 2022 reporting upgrade did not change how environmental performance or disclosure quality relate to greenwashing. This is consistent with research arguing that disclosure reforms alone do not shift firm behaviour without strong enforcement (Lyon & Montgomery, 2015).

In short, only substantive environmental performance, especially emissions and resource use, reduces greenwashing. Disclosure quality variables and environmental violations have no explanatory power, and the reporting reform did not alter greenwashing behaviour during the period studied.

4.7 Summary of Key Findings

From the previous sections, the results indicate that environmental performance like Emissions and Resource Use are the strongest predictors of greenwashing. In Model 1, the Environmental Pillar score significantly reduces greenwashing, although this relationship remains the same before and after the 2022 sustainability reporting upgrade. Model 2 offers more detailed insights, indicating that Emissions and Resource Use consistently and significantly lower greenwashing for both regulatory periods, whereas Innovation does not show any meaningful effect.

Disclosure quality variables like Readability and Disaggregation and Environmental Violations are insignificant, showing that narrative clarity, reporting detail, and recorded violations do not explain greenwashing behaviour. Similar to environmental performance, none of the interaction terms with the Post dummy are significant, denoting that the 2022 reporting reforms did not change how environmental performance influences greenwashing. Among the control variables, only ROA reduces greenwashing, while firm size, leverage (DTB) , and growth (MTB) show minimal or no explanatory value.

Chapter 5: Discussion and Conclusion

5.1 Introduction

This chapter summarises and interprets the key findings of the study, which examined the determinants of greenwashing among Malaysia's top non-financial public-listed companies from 2019 to 2024. Using panel regression with Random Effects and Panel-Corrected Standard Errors (PCSE), the analysis explored how environmental performance, disclosure quality, and environmental violations shape firms' talk–walk decoupling, and whether these relationships shifted following the 2022 sustainability reporting upgrade. The results show that environmental performance, particularly emissions and resource use, are the most consistent and significant factors reducing greenwashing, while disclosure characteristics and the regulatory reform exhibit limited influence.

5.2 Discussion of Key Findings

This study examined how environmental performance, disclosure quality, and environmental violations influence greenwashing among large Malaysian listed firms. The results provide mixed support for the hypotheses, with environmental performance emerging as the only consistent predictor of reduced greenwashing, while disclosure-related variables and violation indicators show no meaningful association. These findings reflect Malaysia's disclosure-driven ESG environment and contribute new evidence to a limited empirical literature.

The Environmental Pillar score demonstrates a significant negative relationship with greenwashing, showing that firms with better overall environmental performance tend to result in sustainability disclosures that more closely reflect their actual environmental outcomes. This finding aligns with literature review that firms with weaker environmental practices are more likely to rely on symbolic reporting to gain legitimacy (Delmas & Burbano, 2011; Marquis & Qian, 2014). However, the lack of a significant moderating effect after the 2022 sustainability reporting upgrade suggests that Malaysia's strengthened disclosure standards did not immediately alter firms' underlying reporting behaviour. This outcome diverges from signalling theory predictions, which posit that heightened scrutiny should discourage opportunistic disclosure practices (Connelly et al., 2011). Instead, the results imply that firms may have adapted to the revised reporting framework in a structural or compliance-oriented manner, without corresponding shifts in substantive behaviour within the examined timeframe.

The Emissions score remains a robust negative predictor of greenwashing in both pre- and post-2022 models. Firms with better emissions management communicate more consistently with their actual performance. This aligns with prior research showing that genuine emissions reductions typically occur in firms with strong environmental processes and monitoring systems (Shahrour et al., 2024; Albitar et al., 2022). The result also contrasts with studies reporting weak predictive power of emissions ratings due to cross-agency inconsistency (Cregan et al., 2023; Treepongkaruna et al., 2024). In Malaysia's context, Refinitiv's structured methodology may more accurately capture emissions performance relative to firms' symbolic disclosure strategies.

Resource Use is the strongest environmental predictor of reduced greenwashing. This suggests that efficient resource management, such as reduced water or energy intensity, reflects substantive operational improvements rather than symbolic commitments. Past literature similarly finds that better resource efficiency is associated with lower climate-related risk (Lazar et al., 2024) and stronger alignment between internal sustainability practices and external disclosures. These results imply that Malaysian firms with tangible operational efficiency gains are less likely to misrepresent their environmental performance, reinforcing the argument that substantive environmental outcomes reduce the need for symbolic reporting.

However, innovation shows no significant relationship with greenwashing, contrary to findings in prior studies where green innovation predicts lower emissions and stronger sustainability performance (Cheng et al., 2024; Miao et al., 2024). Three explanations are plausible. First, innovation ratings may reflect disclosure rather than actual outcomes, particularly in emerging markets where verification mechanisms are weaker and firms face fewer requirements to demonstrate the real environmental impact of their innovative activities (Marquis & Qian, 2013). Second, the effects of environmental innovation tend to materialise over a longer horizon, meaning that the 2019–2024 observation period may be too short to capture measurable improvements arising from new technologies, processes, or R&D investments (Rennings, 2000). Third, innovation disclosures by Malaysian firms may be largely aspirational (Zainal et al., 2013). For example, companies highlight plans, pilot projects, or intentions without translating into substantive operational changes that would influence environmental performance. As a result, innovation does not show a significant relationship with greenwashing in this context. This aligns with scholarly arguments that innovation disclosures are susceptible to symbolic inflation (Pizzetti et al., 2019).

Readability does not significantly affect greenwashing, diverging from studies suggesting clearer ESG text indicates higher transparency and lower impression management (Hu et al., 2024). According to the overview of the Malaysian reporting system by Ahmad & Companies Commission of Malaysia (2019), the null result can be explained by Malaysia's template-driven reporting, where firms comply with standardised formats, reducing variation in readability across sustainability reports. When readability is homogeneous, it loses explanatory power. Similarly, disaggregation shows no significant association with greenwashing. Although financial disaggregation typically signals higher reporting transparency (Chen et al., 2015), the measure may not capture sustainability-specific granularity. Firms may provide detailed financial disclosures while still selectively presenting environmental information. This supports the idea that financial reporting quality does not directly translate into ESG reporting integrity.

From the results, environmental violations do not significantly predict greenwashing. This finding contradicts expectations from prior literature, which suggests that firms facing environmental penalties typically have stronger incentives to engage in symbolic reporting to repair legitimacy and mitigate reputational damage (Boxenbaum & Jonsson, 2008). There are two reasons environmental violations of Malaysian firms do not significantly explain the greenwashing. Firstly, violations are rare among the samples, therefore reducing statistical variation (Ma et al., 2021). Secondly, publicly available records often lag actual misconduct, resulting in delayed or incomplete reporting that reduces their usefulness in explaining firms' disclosure behaviour in Malaysia (Dechezleprêtre & Sato, 2017). These factors weaken the relationship between violations and disclosure behaviour. Similar patterns are reported in other emerging-market studies, where environmental violations are infrequent and unevenly recorded, limiting statistical variation and reducing the ability of empirical models to detect a significant relationship between violation records and firms' disclosure practices (Zhang et al., 2022).

The interaction terms for E Pillar, Emissions, Resource Use, and Innovation with the Post dummy are all insignificant. This indicates that the introduction of the 2022 sustainability reporting upgrade did not significantly alter the determinants of greenwashing. This finding is consistent with research suggesting that disclosure reforms alone do not immediately change corporate behaviour (Lyon & Montgomery, 2015; Li, Du and He, 2024). The insignificance of the Post variable suggests that Malaysia's early-stage regulatory strengthening has not yet translated into observable shifts in ESG reporting credibility.

Taken together, the results highlight that only substantive environmental performance, particularly emissions and resource use, are being served as the reliable predictors of reduced greenwashing, while disclosure-based variables and violations remain weak indicators. This reinforces the argument that meaningful environmental outcomes, rather than narrative or stylistic features of sustainability reports, are the clearest signals of genuine corporate environmental commitment.

5.3 Implications of the Findings

5.3.1 Theoretical Implications

This study contributes to the theoretical understanding of greenwashing, signalling, and legitimacy in several important ways. First, by operationalising greenwashing as the gap between standardised positive disclosure tone and standardised environmental performance, the study advances greenwashing measurement and strengthens the notion of greenwashing as a form of "talk–walk" decoupling, where firms promote favourable sustainability narratives that do not correspond to their actual practices (Bromley & Powell, 2012). Second, the findings highlight the differing importance of environmental sub-pillars, showing that emissions and resource use are stronger predictors of greenwashing than the aggregate Environmental Pillar score. This supports emerging calls in ESG research for more granular assessment of environmental components rather than reliance on broad composite indicators (Albitar et al., 2022). Third, the insignificance of readability and disaggregation challenges the assumption that stylistic or structural disclosure features correspond to substantive sustainability performance. This suggests that in reporting environments defined by templated formats or boilerplate language, disclosure quality metrics may not effectively differentiate symbolic reporting from genuine environmental effort. Finally, the study adds to the growing literature on ESG practices in emerging markets by demonstrating that regulatory upgrades alone do not produce immediate behavioural change without enforcement strengthening. This is consistent with evidence that firms operating in transitioning regulatory systems often adopt symbolic compliance strategies rather than undertaking substantive environmental improvements when oversight remains weak (Marquis & Qian, 2014; Zhang et al., 2022).

5.3.2 Practical and Policy Implications

The findings of this study offer several practical implications for regulators, investors, and companies. For regulators such as Bursa Malaysia and the Securities Commission Malaysia, the results indicate that enhancing assurance, verification, and enforcement mechanisms is likely to be more effective than merely expanding disclosure templates. This aligns with prior research showing that disclosure-based reforms do not translate into substantive behavioural changes without strong monitoring and enforcement structures (Lyon & Montgomery, 2015; Li et al., 2024). Regulatory oversight should therefore place greater emphasis on environmental performance indicators like Emissions and Resource Use, which the study identifies as the strongest predictors of reduced greenwashing.

For investors and analysts, the findings suggest that environmental performance measures offer more meaningful insights into genuine sustainability efforts than stylistic features such as tone, readability, or disclosure length. Consistent with earlier evidence that firms with stronger environmental outcomes engage less in symbolic reporting (Delmas & Burbano, 2011), the study cautions investors against relying solely on polished or positively framed sustainability narratives, which may not accurately reflect operational performance.

For companies, the results emphasise that improving operational sustainability initiatives related to emissions reduction and resource-efficiency are more effective for reducing greenwashing risk than refining the communicative aspects of reporting. This is in line with arguments that substantive environmental actions provide greater legitimacy and stakeholder value than symbolic disclosure enhancements (Marquis & Qian, 2014). Firms aiming to build stakeholder trust should therefore prioritise measurable environmental improvements rather than primarily focusing on narrative strategies.

5.4 Limitations of the Findings

Several limitations should be considered when interpreting the findings of this study. First, the construction of the Greenwashing Score is primarily designed to capture active greenwashing, where firms exaggerate or overstate their environmental performance through overly positive disclosure tone. By defining greenwashing as the standardized difference between positive disclosure tone (“talk”) and standardized environmental performance (“walk”), the measure is effective in identifying firms that communicate strong environmental claims while delivering weak actual performance. However, this approach may under-detect passive greenwashing, where firms with poor environmental performance provide limited or minimal disclosure. In such cases, low disclosure tone combined with low environmental performance may appear as alignment rather than concealment. As a result, some low-performing but low-communication firms may be classified as non-greenwashing, even though they may be strategically withholding information. This suggests that the Greenwashing Score is better suited to detecting exaggerated claims than omission-based greenwashing (Lyon & Montgomery, 2015; Pizzetti et al., 2019).

Second, the text-based measures used in the study, such as sentiment scores, readability indices, and keyword-based environmental filtering, depend heavily on how firms write their sustainability reports. In Malaysia, sustainability disclosures often follow standardized templates or boilerplate language, which reduces linguistic variation and limits the ability of text analytics to distinguish between genuine and symbolic reporting (Hu et al., 2024). Similarly, the disaggregation measure applied in this study follows the established approach of Chen et al. (2015) by focusing on the granularity of financial statement line items. While financial disaggregation is widely recognised as a proxy for disclosure transparency and information environment quality, it may not fully capture transparency in environmental or sustainability-specific disclosures. Firms may produce highly detailed financial reports while remaining vague or selective in their environmental narratives.

Third, environmental violations do not exhibit a significant relationship with greenwashing, partly due to data sparsity. Descriptive statistics show that the average violation rate is low, indicating that most firms in the sample have no recorded environmental violations. When violations are rare events, statistical models face difficulties in identifying systematic relationships, even if such relationships exist in theory. Furthermore, publicly available violation data often suffer from reporting delays, meaning that recorded violations may not align temporally with firms' disclosure behaviour. Together, the rarity of violations and the lag in reporting reduce the empirical power of violation variables to explain greenwashing dynamics.

Fourth, the environmental innovation variable is measured contemporaneously, which may limit its ability to capture the true effects of innovation on greenwashing. Environmental innovation typically involves long gestation periods, where investments in research and development, cleaner technologies, or process improvements only translate into measurable environmental performance gains after several years (Rennings, 2000). By examining innovation and greenwashing within the same fiscal year, the study may not fully capture these delayed effects. This timing mismatch may help explain why innovation does not show a significant relationship with greenwashing, despite strong evidence in the broader literature linking innovation to long-term environmental improvements.

Finally, although the Random Effects model accounts for some unobservable firm-specific characteristics, the possibility of endogeneity cannot be fully ruled out. Unobserved factors such as internal governance culture, stakeholder pressure, or sustainability capabilities may influence both environmental disclosure practices and actual environmental performance. These factors are not directly measured in the model and may affect the estimated relationships.

5.5 Recommendations for Future Research

Based on the limitations identified in this study, several directions for future research are recommended.

First, future studies should refine greenwashing measurement to capture both active and passive forms of greenwashing. While the current Greenwashing Score is effective in identifying firms that exaggerate environmental claims, alternative measures could incorporate indicators of disclosure omission, silence, or selective non-reporting. For example, future research could compare environmental disclosure intensity or completeness against objective performance benchmarks to better detect firms that conceal poor environmental performance through limited communication.

Second, researchers should enhance text-based disclosure analysis by adopting more advanced and context-sensitive natural language processing techniques. The use of transformer-based models, domain-specific ESG dictionaries, or contextual embeddings may help overcome the limitations posed by boilerplate sustainability reporting. Additionally, incorporating multilingual text analysis would be particularly valuable in Malaysia, where sustainability reports may be published in both English and Malay, potentially affecting sentiment and readability measurements.

Third, future research would benefit from developing richer and more informative environmental violation datasets. Beyond binary violation indicators, future studies should consider incorporating violation severity, such as the size of penalties, the nature of offences, and whether corrective actions were required. Combining regulatory data with media reports, NGO databases, and enforcement announcements could improve coverage and reduce reporting delays, thereby strengthening the ability to examine how firms adjust disclosure behaviour in response to environmental misconduct.

Fourth, future studies should explicitly account for time-lag effects, particularly when examining environmental innovation. Incorporating lagged innovation variables over multiple years would allow researchers to better capture the delayed impact of innovation on environmental performance and greenwashing behaviour. This approach would align more closely with the long-term nature of technological and process-based environmental improvements.

Finally, to address potential endogeneity concerns, future research could employ more advanced econometric techniques such as instrumental variable approaches, dynamic panel models or quasi-experimental designs. Besides, including additional firm-level variables like board sustainability expertise, external assurance of sustainability reports, ownership structure, or stakeholder engagement may also help isolate causal mechanisms and deepen understanding of the drivers of greenwashing.

Taken together, these recommendations highlight opportunities to improve greenwashing measurement, data quality, and causal inference. Addressing these areas would contribute to a more comprehensive understanding of how firms balance symbolic and substantive environmental actions, particularly in emerging-market contexts undergoing rapid ESG regulatory development.

5.6 Concluding Remarks

This study set out to examine the determinants of greenwashing among Malaysia's largest non-financial public-listed companies, focusing on whether environmental performance, disclosure quality, and environmental violations influence firms' "talk–walk" decoupling behaviour. By integrating text analytics, environmental ratings, and violation records within a panel regression framework, the study provides one of the first large-sample, multi-dimensional assessments of greenwashing in Malaysia. The findings show that substantive environmental performance, particularly emissions management and resource efficiency are the strongest predictor of reduced greenwashing, supporting arguments that firms with more robust environmental practices rely less on symbolic sustainability reporting (Delmas & Burbano, 2011; Marquis & Qian, 2014). In contrast, disclosure-related measures such as readability and financial disaggregation do not significantly influence greenwashing, reflecting the standardised and often boilerplate nature of Malaysian ESG reporting (Ahmad & CCM, 2019; Hu et al., 2024). Environmental violations also exhibit no significant explanatory power, likely due to limited, delayed, or inconsistently reported misconduct records, which reduces their usefulness for empirical analysis in emerging-market contexts (Dechezleprêtre & Sato, 2017; Zhang et al., 2022).

The absence of significant moderation effects following the 2022 sustainability reporting upgrade further indicates that regulatory strengthening has not yet resulted in measurable behavioural change. This aligns with evidence that disclosure-based reforms rarely shift reporting behaviour in the short term unless accompanied by rigorous verification and enforcement mechanisms (Lyon & Montgomery, 2015; Li et al., 2024). Collectively, these findings reinforce the conclusion that genuine environmental performance remains the most reliable indicator of authentic sustainability commitment.

Despite several limitations, the study makes important theoretical and practical contributions. Theoretically, it advances a replicable greenwashing measure grounded in talk–walk decoupling (Bromley & Powell, 2012). Empirically, it offers insight into ESG behaviour in an emerging market where disclosure standards are evolving but enforcement remains uneven. Practically, the results guide regulators, investors, and companies by identifying environmental indicators that most effectively distinguish symbolic reporting from substantive environmental performance. Future research incorporating longer time periods, richer violation datasets, and more advanced analytical tools will be essential for understanding how ESG practices evolve as Malaysia's sustainability reporting regime matures. Overall, the study highlights the continued need for stronger assurance, enforcement, and performance-based evaluation to ensure that ESG disclosures in Malaysia progress beyond symbolic compliance toward genuine environmental stewardship.

REFERENCES

Ahmad, F. & Companies Commission of Malaysia. (2019). AN OVERVIEW OF MALAYSIAN BUSINESS REPORTING SYSTEM. In *SSM ANNUAL DIALOGUE 2019*.

Ahn, M., Jung, D., Kim, J., Lee, W., & Sunwoo, H. (2023). Do more readable sustainability reports provide more value-relevant information to shareholders? Finance Research Letters, 57, 104154. https://doi.org/10.1016/j.frl.2023.104154

Albitar, K., Borgi, H., Khan, M., & Zahra, A. (2022). Business environmental innovation and CO2 emissions: The moderating role of environmental governance. Business Strategy and the Environment, 32(4), 1996–2007. https://doi.org/10.1002/bse.3232

Almaqtari, F. A., Rehman, S., Nigam, S., & Khan, M. (2024). The impact of board structure, IT governance, and fintech on green finance and sustainability: an Integrated model. *Strategic Change*, *34*(2), 337–357. https://doi.org/10.1002/jsc.2623

Amel-Zadeh, A., & Serafeim, G. (2018). Why and How Investors Use ESG Information: Evidence from a Global Survey. Financial Analysts Journal, 74(3), 87–103. https://doi.org/10.2469/faj.v74.n3.2

ASEAN Taxonomy Board. (2023). *Asian Taxonomy for Sustainability Finance: Version 2*. https://asean.org/wp-content/uploads/2023/03/ASEAN-Taxonomy-Version-2.pdf

ASEAN Taxonomy Board. (2024). *ASEAN Taxonomy Board Release ASEAN Taxonomy for Sustainable Finance – Version 3: For Transportation and Construction Sectors*. https://www.sfinstitute.asia/wp-content/uploads/2024/07/ATB_Media_Statement_ATV3.pdf

Baltagi, B. H. (2021). Econometric Analysis of Panel Data. In *Springer texts in business and economics*. https://doi.org/10.1007/978-3-030-53953-5

Bank Negara Malaysia. (2022). Climate Change and Principle-based Taxonomy (CCPT). https://www.bnm.gov.my/documents/20124/938039/CCPT.pdf

Bank Negara Malaysia. (2025). *Navigating the Financial Sector's Transition to the National Sustainability Reporting Framework.*

https://www.bnm.gov.my/documents/20124/17523783/fsr24h2_en_box3.pdf

Barrett, G. (2012). Microeconometrics Using Stata, by A. Colin Cameron and Pravin K. Trivedi (Stata Press, College Station, Texas, USA, 2009), pp. xl + 692. *Economic Record*, *88*(283), 595–596. https://doi.org/10.1111/1475-4932.12006

Beck, N., & Katz, J. N. (1995). What To Do (and Not to Do) with Time-Series Cross-Section Data. *American Political Science Review*, *89*(3), 634–647. https://doi.org/10.2307/2082979

Berg, F., Kölbel, J. F., & Rigobon, R. (2022). Aggregate confusion: the divergence of ESG ratings. Review of Finance, 26(6), 1315–1344. https://doi.org/10.1093/rof/rfac033

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111. https://doi.org/10.1002/jrsm.12

Boxenbaum, E., & Jonsson, S. (2008). Isomorphism, Diffusion and Decoupling. In *The SAGE Handbook of Organizational Institutionalism* (pp. 78–98). https://doi.org/10.4135/9781849200387.n3

Bromley, P., & Powell, W. W. (2012). From smoke and mirrors to Walking the talk: Decoupling in the contemporary world. *Academy of Management Annals*, *6*(1), 483–530. https://doi.org/10.1080/19416520.2012.684462

Bursa Malaysia. (2022). Sustainability Reporting Guide (2nd edition).

Bursa Malaysia. (2022). Sustainability Reporting Guide (3rd ed.) & Toolkits. https://www.bursamalaysia.com/regulation/listing_requirements/sustainability_reporting

Calamai, T., Balalau, O., Guenedal, T. L., & Suchanek, F. M. (2025). Corporate Greenwashing Detection in Text – a survey. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2502.07541

Casey, R., Gao, F., Kirschenheiter, M., Li, S., & Pandit, S. (2023). Measuring disaggregation quality. Journal of Accounting Auditing & Finance, 0148558X2211486. https://doi.org/10.1177/0148558x221148626

Chen, S., Miao, B., & Shevlin, T. (2015). A new measure of disclosure quality: the level of disaggregation of accounting data in annual reports. Journal of

Accounting Research, 53(5), 1017–1054. https://doi.org/10.1111/1475-679x.12094

Cheng, Q., Lin, A., & Yang, M. (2024). Green innovation and firms' financial and environmental performance: The roles of pollution prevention versus control. Journal of Accounting and Economics, 101706. https://doi.org/10.1016/j.jacceco.2024.101706

Choi, Y. R., & Shepherd, D. A. (2005). Stakeholder perceptions of age and other dimensions of newness. Journal of Management, 31(4), 573–596. https://doi.org/10.1177/0149206304272294

Clarkson, P. M., Li, Y., Richardson, G. D., & Vasvari, F. P. (2007). Revisiting the relation between environmental performance and environmental disclosure: An empirical analysis. Accounting Organizations and Society, 33(4–5), 303–327. https://doi.org/10.1016/j.aos.2007.05.003

Clarkson, P. M., Overell, M. B., & Chapple, L. (2011). Environmental Reporting and its Relation to Corporate Environmental Performance. Abacus, 47(1), 27–60. https://doi.org/10.1111/j.1467-6281.2011.00330.x

Cohen, N. (2013). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. In *Routledge eBooks*. https://doi.org/10.4324/9780203774441

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2010). Signaling Theory: A Review and assessment. Journal of Management, 37(1), 39–67. https://doi.org/10.1177/0149206310388419

Cregan, C., Kelly, J. A., & Clinch, J. P. (2023). Are environmental, social and governance (ESG) ratings reliable indicators of emissions outcomes? A case study of the airline industry. Corporate Social Responsibility and Environmental Management, 31(2), 909–928. https://doi.org/10.1002/csr.2608

Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th ed.). Sage Publications.

De Freitas Netto, S. V., Sobral, M. F. F., Ribeiro, A. R. B., & Da Luz Soares, G. R. (2020). Concepts and forms of greenwashing: a systematic review. Environmental Sciences Europe, 32(1). https://doi.org/10.1186/s12302-020-0300-3

Dechezleprêtre, A., & Sato, M. (2017). The impacts of environmental regulations on competitiveness. *Review of Environmental Economics and Policy*, *11*(2), 183–206. https://doi.org/10.1093/reep/rex013

Delmas, M. A., & Burbano, V. C. (2011). The drivers of greenwashing. California Management Review, 54(1), 64–87. https://doi.org/10.1525/cmr.2011.54.1.64

Dorfleitner, G., & Utz, S. (2023). Green, green, it's green they say: a conceptual framework for measuring greenwashing on firm level. Review of Managerial Science, 18(12), 3463–3486. https://doi.org/10.1007/s11846-023-00718-w

Du, K., Huddart, S., & Jiang, X. D. (2022). Lost in standardization: Effects of financial statement database discrepancies on inference. Journal of Accounting and Economics, 76(1), 101573. https://doi.org/10.1016/j.jacceco.2022.101573

Dyck, A., Lins, K. V., Roth, L., & Wagner, H. F. (2018). Do institutional investors drive corporate social responsibility? International evidence. Journal of Financial Economics, 131(3), 693–714. https://doi.org/10.1016/j.jfineco.2018.08.013

Esposito, P., Doronzo, E., Riso, V., & Tufo, M. (2025). Sustainability in energy companies under the lens of cultural pressures: When do we talk of greenwashing? Corporate Social Responsibility and Environmental Management. https://doi.org/10.1002/csr.3111

Fahmi, F. M., Azmi, N. F. R., & Mat, T. Z. T. (n.d.). Corporate Characteristics and Sustainability Reporting: From the Lens of the Legitimacy Theory. *Asia-Pacific Management Accounting Journal*, *17*(2), 101–130. https://doi.org/10.24191/apmaj.v17i2-04

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage Publications Ltd. https://vlb-content.vorarlberg.at/fhbscan1/330900091084.pdf

Flammer, C., Hong, B., & Minor, D. (2019). Corporate governance and the rise of integrating corporate social responsibility criteria in executive compensation: Effectiveness and implications for firm outcomes. Strategic Management Journal, 40(7), 1097–1122. https://doi.org/10.1002/smj.3018

Font, X., Walmsley, A., Cogotti, S., McCombes, L., & Häusler, N. (2012). Corporate social responsibility: The disclosure–performance gap. Tourism Management, 33(6), 1544–1553. https://doi.org/10.1016/j.tourman.2012.02.012

Forliano, C., Battisti, E., De Bernardi, P., & Kliestik, T. (2025). Mapping the greenwashing research landscape: a theoretical and field analysis. Review of Managerial Science. https://doi.org/10.1007/s11846-025-00856-3

Ghitti, M., Gianfrate, G., & Palma, L. (2023). The agency of greenwashing. *Journal of Management & Governance*, *28*(3), 905–941. https://doi.org/10.1007/s10997-023-09683-8

Gorovaia, N., & Makrominas, M. (2024). Identifying greenwashing in corporate-social responsibility reports using natural-language processing. European Financial Management, 31(1), 427–462. https://doi.org/10.1111/eufm.12509

Greene, W. H. (2018). *Econometric analysis* (8th ed.). Pearson.

Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics* (5th ed.). McGraw-Hill.

Guo, R., Tao, L., Li, C. B., & Wang, T. (2015). A Path analysis of Greenwashing in a trust crisis among Chinese energy companies: The role of brand legitimacy and brand loyalty. Journal of Business Ethics, 140(3), 523–536. https://doi.org/10.1007/s10551-015-2672-7

Hair, Jr. J., Wolfinbarger, M. C., Ortinau, D. J., & Bush, R. P. (2013). *Essential of Marketing. New York: Mc Graw-Hill*

Hikal, H. M. M., Abubakr, A. A. M., Musa, A. M. H., Abdelraheem, A. A. E., & Adam, M. I. A. B. (2025). Sustainability auditing and reporting in Malaysia: Strengthening transparency, accountability, and corporate responsibility. *International Journal of Innovative Research and Scientific Studies*, *8*(4), 1068-1078. https://www.researchgate.net/profile/Abubkr-Abdelraheem/publication/393017970_Sustainability_auditing_and_reporting_in_Malaysia_Strengthening_transparency_accountability_and_corporate_responsibility/links/685d0adce8fa0f5c2827d9b9/Sustainability-auditing-and-reporting-in-Malaysia-Strengthening-transparency-accountability-and-corporate-responsibility.pdf

Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. *Stata Journal*, 7(3), 281–312. https://doi.org/10.1177/1536867X0700700301

Hoitash, R., & Hoitash, U. (2014). Measuring Accounting Disclosure Complexity with XBRL. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2433677

Hsiao, C. (2014). *Analysis of panel data* (3rd ed.). Cambridge University Press.

Hu, B., & Xu, Q. (2025). Environmental regulation penalties and corporate environmental information disclosure. International Review of Economics & Finance, 104344. https://doi.org/10.1016/j.iref.2025.104344

Hu, P., Li, X., Li, N., Wang, Y., & Wang, D. D. (2024). Peeking into corporate greenwashing through the readability of ESG disclosures. Sustainability, 16(6), 2571. https://doi.org/10.3390/su16062571

Hu, P., Li, X., Li, N., Wang, Y., & Wang, D. D. (2024). Peeking into corporate greenwashing through the readability of ESG disclosures. Sustainability, 16(6), 2571. https://doi.org/10.3390/su16062571

Hu, S., Chen, P., & Zhang, C. (2025). How does green finance reform affect corporate ESG greenwashing behavior? International Review of Financial Analysis, 104037. https://doi.org/10.1016/j.irfa.2025.104037

Hu, X., Hua, R., Liu, Q., & Wang, C. (2023). The green fog: Environmental rating disagreement and corporate greenwashing. Pacific-Basin Finance Journal, 78, 101952. https://doi.org/10.1016/j.pacfin.2023.101952

Huang, H., Wu, D., & J, G. (2017). Chinese shareholders' reaction to the disclosure of environmental violations: a CSR perspective. International Journal of Corporate Social Responsibility, 2(1). https://doi.org/10.1186/s40991-017-0022-z

Huang, Y., Xiong, N., & Liu, C. (2024). Renewable energy technology innovation and ESG greenwashing: Evidence from supervised machine learning methods using patent text. Journal of Environmental Management, 370, 122833. https://doi.org/10.1016/j.jenvman.2024.122833

Hummel, K., & Schlick, C. (2016). The relationship between sustainability performance and sustainability disclosure: Reconciling voluntary disclosure theory and legitimacy theory. *Journal of Accounting and Public Policy, 35*(5), 455–476. https://doi.org/10.1016/j.jaccpubpol.2016.06.001

Hutton, A. P., Marcus, A. J., & Tehranian, H. (2009). Opaque financial reports, R2, and crash risk. Journal of Financial Economics, 94(1), 67–86. https://doi.org/10.1016/j.jfineco.2008.10.003

IFRS Sustainability. (2025). *SUSTAINABILITY DISCLOSURE STANDARDS (ISSB STANDARDS)— APPLICATION AROUND THE WORLD JURISDICTIONAL PROFILE: Malaysia*. https://www.ifrs.org/content/dam/ifrs/publications/sustainability-jurisdictions/pdf-profiles/malaysia-ifrs-profile.pdf

Jamil, N. N., & Wahyuni, E. T. (2025). Greenwashing and board effectiveness: Moderating role of CSR committee from Malaysia evidence. Edelweiss Applied Science and Technology, 9(5), 1508–1521. https://doi.org/10.55214/25768484.v9i5.7187

Jamil, N., & Wahyuni, D. (2024). Corporate greenwashing in Malaysia: Evidence from oil & gas. GBMR Journal. http://gbmrjournal.com/pdf/v16n3s/V16N3s-9.pdf

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, *6*(3), 255–259. https://doi.org/10.1016/0165-1765(80)90024-5

Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55(2), 163–172. https://doi.org/10.2307/1403192

Jay Westerveld (1986). *Essay on hotel environmental practices*.

Jensen, M. C., & Meckling, W. H. (1979). Theory of the firm: Managerial behavior, agency costs, and ownership structure. In Rochester studies in economics and policy issues (pp. 163–231). https://doi.org/10.1007/978-94-009-9257-3_8

Johnston, J. A., Reichelt, K. J., & Sapkota, P. (2023). Measuring Financial Statement disaggregation using XBRL. Journal of Information Systems, 38(1), 119–147. https://doi.org/10.2308/isys-2021-004

Ju, W., & Jin, S. (2024). The impact of green innovation on the carbon performance of Chinese manufacturing enterprises: Moderating role of internal governance. Heliyon, 10(10), e31272. https://doi.org/10.1016/j.heliyon.2024.e31272

Karpoff, J. M., Lott, J. R., Jr, & Wehrly, E. W. (2005). The Reputational Penalties for Environmental Violations: Empirical evidence. The Journal of Law and Economics, 48(2), 653–675. https://doi.org/10.1086/430806

Kathan, M. C., Utz, S., Dorfleitner, G., Eckberg, J., & Chmel, L. (2025). What you see is not what you get: ESG scores and greenwashing risk. Finance Research Letters, 106710. https://doi.org/10.1016/j.frl.2024.106710

Keilmann, J., Koch, T., 2023. When Environmental claims are empty promises: how greenwashing affects corporate reputation and credibility. Environ. Commun. 1–19. https://doi.org/10.1080/17524032.2023.2267782.

Kim, E., & Lyon, T. P. (2014). Greenwash vs. Brownwash: Exaggeration and Undue Modesty in Corporate Sustainability Disclosure. Organization Science, 26(3), 705–723. https://doi.org/10.1287/orsc.2014.0949

Kishan, K., & Azhar, Z. (2025). Greenwashing in sustainability reporting: evidence from Malaysia. *Journal of Financial Reporting & Accounting*. https://doi.org/10.1108/jfra-01-2025-0060

Kölbel, J. F., Heeb, F., Paetzold, F., & Busch, T. (2020). Can sustainable investing save the world? Reviewing the mechanisms of investor impact. Organization & Environment, 33(4), 554–574. https://doi.org/10.1177/1086026620919202

Landstrom, J. (2019). Regression analysis and panel data. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3487658

Lazar, E., Pan, J., & Wang, S. (2024). Measuring climate-related and environmental risks for equities. Journal of Environmental Management, 373, 123393. https://doi.org/10.1016/j.jenvman.2024.123393

Lee, M. T., & Raschke, R. L. (2022). Stakeholder legitimacy in firm greening and financial performance: What about greenwashing temptations?. Journal of Business Research, 155, 113393. https://doi.org/10.1016/j.jbusres.2022.113393

Lenowitz, J. A. (2022). Legitimacy types and procedures. In Oxford University Press eBooks (pp. 234–255). https://doi.org/10.1093/oso/9780198852346.003.0008

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. Journal of Accounting and Economics, 45(2–3), 221–247. https://doi.org/10.1016/j.jacceco.2008.02.003

Li, W., Du, H., & He, F. (2024). Mandatory corporate ESG disclosure and default risk – Evidence from China. *Pacific-Basin Finance Journal*, *89*, 102578. https://doi.org/10.1016/j.pacfin.2024.102578

Lian, Y., Ye, T., Zhang, Y., & Zhang, L. (2023). How does corporate ESG performance affect bond credit spreads: Empirical evidence from China. International Review of Economics & Finance, 85, 352–371. https://doi.org/10.1016/j.iref.2023.01.024

Lin, P. T., Jin, Y., Gao, F., Yang, R., & Lin, Q. (2023). Institutional investors, CSR report readability and the moderating role of ESG performance. SAGE Open, 13(4). https://doi.org/10.1177/21582440231208514

Liu, C., & Wang, X. (2022). Media and institutional investors focus on the impact on corporate sustainability performance. Sustainability, 14(21), 13878. https://doi.org/10.3390/su142113878

Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S., & Reber, S. (2017). Descriptive Analysis in Education: A Guide for Researchers. NCEE 2017-4023. *National Center for Education Evaluation and Regional Assistance*. http://files.eric.ed.gov/fulltext/ED573325.pdf

Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S., & Reber, S. (2017). Descriptive Analysis in Education: A Guide for Researchers. NCEE 2017-4023. *National Center for Education Evaluation and Regional Assistance*. http://files.eric.ed.gov/fulltext/ED573325.pdf

Lublóy, Á., Keresztúri, J. L., & Berlinger, E. (2024). Quantifying firm-level greenwashing: A systematic literature review. Journal of Environmental Management, 373, 123399. https://doi.org/10.1016/j.jenvman.2024.123399

Lyon, T. P., & Maxwell, J. W. (2011). Greenwash: Corporate Environmental Disclosure under Threat of Audit. *Journal of Economics & Management Strategy*, *20*(1), 3–41. https://doi.org/10.1111/j.1530-9134.2010.00282.x

Lyon, T. P., & Montgomery, A. W. (2015). The means and end of greenwash. Organization & Environment, 28(2), 223–249. https://doi.org/10.1177/1086026615575332

Ma, X., Fan, D., Zhou, Y., & Yang, C. (2021). The impact of inspection on the sustainable production strategy: Environmental violation and abatement in emerging markets. *Transportation Research Part E Logistics and Transportation Review*, *150*, 102294. https://doi.org/10.1016/j.tre.2021.102294

Macchioni, R., Fiondella, C., & Prisco, M. (2024). The value relevance of environmental innovation: Evidence from European Union context. Journal of Cleaner Production, 446, 141246. https://doi.org/10.1016/j.jclepro.2024.141246

Mahoney, L. S., Thorne, L., Cecil, L., & LaGore, W. (2012). A research notes on standalone corporate social responsibility reports: Signaling or greenwashing? Critical Perspectives on Accounting, 24(4–5), 350–359. https://doi.org/10.1016/j.cpa.2012.09.008

Marquis, C., & Qian, C. (2013). Corporate Social Responsibility reporting in China: symbol or substance? Organization Science, 25(1), 127–148. https://doi.org/10.1287/orsc.2013.0837

Miao, C., Chen, Z., & Zhang, A. (2024). Green technology innovation and carbon emission efficiency: The moderating role of environmental uncertainty. The Science of the Total Environment, 938, 173551. https://doi.org/10.1016/j.scitotenv.2024.173551

Motz, M., Uzun, S., Hariharan, A., & Weinhardt, C. (2025). Unveiling Green Facades: Detecting Greenwashing Tendencies in Corporate Sustainability Reports. Proceedings of the . . . Annual Hawaii International Conference on System Sciences/Proceedings of the Annual Hawaii International Conference on System Sciences. https://doi.org/10.24251/hicss.2025.113

National Sustainability Reporting Framework (NSRF). (2024). Securities Commission Malaysia. https://www.sc.com.my/api/documentms/download.ashx?id=e98c3900-7b35-4cf5-a07d-fd17acf8734e

Ngu, S. B., & Amran, A. (2021). Materiality Disclosure in Sustainability Reporting: Evidence from Malaysia. *Asian Journal of Business and Accounting*, *14*(1), 225–252. https://doi.org/10.22452/ajba.vol14no1.9

Nyilasy, G., Gangadharbatla, H., & Paladino, A. (2013). Perceived greenwashing: The interactive effects of green advertising and corporate environmental performance on consumer reactions. Journal of Business Ethics, 125(4), 693–707. https://doi.org/10.1007/s10551-013-1944-3

OECD. (2022). Trends in ESG investing and quality infrastructure investment in Asia-Pacific. https://doi.org/10.1787/86d154c1-en

Parguel, B., Benoît-Moreau, F., & Larceneux, F. (2011). How sustainability ratings might deter 'Greenwashing': A closer look at ethical corporate communication. *Journal of Business Ethics*, *102*(1), 15–28. https://doi.org/10.1007/s10551-011-0901-2

Patten, D. M. (2002). The relation between environmental performance and environmental disclosure: a research note. Accounting Organizations and Society, 27(8), 763–773. https://doi.org/10.1016/s0361-3682(02)00028-4

Pizzetti, M., Gatti, L., & Seele, P. (2019). Firms Talk, Suppliers walk: Analyzing the locus of greenwashing in the blame game and introducing 'Vicarious Greenwashing.' *Journal of Business Ethics*, *170*(1), 21–38. https://doi.org/10.1007/s10551-019-04406-2

Potharla, S., & Turubilli, S. K. (2024). Emission reduction scores and market behaviour: Analysing stock synchronicity in India. The Indonesian Journal of Accounting Research, 27(03). https://doi.org/10.33312/ijar.762

PwC. (2024). Spotlight on sustainability: National Sustainability Reporting Framework. https://www.pwc.com/my/en/assets/publications/2024/pwc-my-national-sustainability-reporting-framework.pdf

Qu, X. (2007). Multivariate data analysis. *Technometrics*, *49*(1), 103–104. https://doi.org/10.1198/tech.2007.s455

Quoquab, F. (2021). Do they mean what they say? Investigating the greenwash behaviour in the Malaysian property development sector. *Journal of Cleaner Production*, 290, 125–150. https://doi.org/10.1016/j.jclepro.2020.125150

Ramus, C. A., & Montiel, I. (2005). When are corporate environmental policies a form of greenwashing? *Business & Society*, *44*(4), 377–414. https://doi.org/10.1177/0007650305278120

Rennings, K. (2000). Redefining innovation — eco-innovation research and the contribution from ecological economics. *Ecological Economics*, *32*(2), 319–332. https://doi.org/10.1016/s0921-8009(99)00112-3

Revon Media (2023). *Sarawak Energy's comprehensive tree-planting initiatives*. Revon Media. https://revonmedia.com/2023/01/30/sarawak-energys-comprehensive-tree-planting-initiatives/

Scaltrito, N. D. (2015). Assessing disclosure quality: a methodological issue. Journal of Modern Accounting and Auditing, 11(9). https://doi.org/10.17265/1548-6583/2015.09.004

SEC (2022). *SEC Climate-Related Disclosure Proposal*. https://www.sec.gov/newsroom/press-releases/2024-31

Securities Commission Malaysia. (2019). Sustainable and Responsible Investment (SRI) Roadmap for the Malaysian Capital Market. https://www.sc.com.my/resources/publications-and-research/sri-roadmap

Seele, P., & Gatti, L. (2015). Greenwashing revisited: In search of a Typology and Accusation-Based Definition incorporating legitimacy strategies. Business Strategy and the Environment, 26(2), 239–252. https://doi.org/10.1002/bse.1912

Shahrour, M. H., Rohani, A., Wojewodzki, M., & Tran, D. V. (2024). Carbon Performance and Financial Performance: How R&D Makes a Difference Pre- and Post-Paris Accord. International Journal of Finance & Economics. https://doi.org/10.1002/ijfe.3109

Shahudin, F., Aziz, S., & Ahmad, F. (2015). Modelling the greenwashing behavior among Malaysian firms: the roles of organizational and individual drivers. *International Journal of Business*, *7*, 2289–1552. https://ijbel.com/wp-content/uploads/2015/09/KLIBEL7_Bus-13.pdf

Shao, J., Li, W., Huang, L., & Tian, Y. (2025). Environmental penalties and corporate carbon disclosure in China: divergent effects of resource availability and the role of social media pressure. Frontiers in Environmental Science, 12. https://doi.org/10.3389/fenvs.2024.1426046

Shimamura, T., Tanaka, Y., & Managi, S. (2025). Evaluating the impact of report readability on ESG scores: A generative AI approach. International Review of Financial Analysis, 104027. https://doi.org/10.1016/j.irfa.2025.104027

Sklavos, G., Zournatzidou, G., Ragazou, K., & Sariannidis, N. (2025). Unmasking Greenwashing in Finance: A PROMETHEE II-Based Evaluation of ESG disclosure and green Accounting alignment. Risks, 13(7), 134. https://doi.org/10.3390/risks13070134

Smith, M., Yahya, K., & Amiruddin, A. M. (2007). Environmental disclosure and performance reporting in Malaysia. *Asian Review of Accounting*, *15*(2), 185–199. https://doi.org/10.1108/13217340710823387

Stewart, R. (2025). ESG and the Changing Language of Corporate Social Responsibility.

Suchman, M. C. (1995). Managing Legitimacy: strategic and institutional approaches. Academy of Management Review, 20(3), 571–610. https://doi.org/10.5465/amr.1995.9508080331

Tanthanongsakkun, S., Treepongkaruna, S., & Jiraporn, P. (2022). Carbon emissions, corporate governance, and staggered boards. Business Strategy and the Environment, 32(1), 769–780. https://doi.org/10.1002/bse.3174

Testa, F., Daddi, T., Iraldo, F., Gusmerotti, N. M., & Frey, M. (2023). Mapping the greenwashing research landscape: A theoretical and field analysis. Business Strategy and the Environment, 32(1), 372–389. https://doi.org/10.1002/bse.3116

Testa, F., Miroshnychenko, I., Barontini, R., & Frey, M. (2018). Does it pay to be a greenwasher or a brownwasher? *Business Strategy and the Environment, 27*(8), 1104–1116. https://doi.org/10.1002/bse.2058

The Harvard Law School Forum on Corporate Governance. (2022). ESG Ratings: A Compass without Direction. https://corpgov.law.harvard.edu/2022/08/24/esg-ratings-a-compass-without-direction/

Treepongkaruna, S., Yong, H. H. A., Thomsen, S., & Kyaw, K. (2024). Greenwashing, carbon emission, and ESG. Business Strategy and the Environment, 33(8), 8526–8539. https://doi.org/10.1002/bse.3929

Truong, Y. (2025). Green innovations and toxic emissions of manufacturing firms: Do green innovative firms pollute less? Corporate Social Responsibility and Environmental Management. https://doi.org/10.1002/csr.70065

Tsang, A., Frost, T., & Cao, H. (2022). Environmental, Social, and Governance (ESG) disclosure: A literature review. The British Accounting Review, 55(1), 101149. https://doi.org/10.1016/j.bar.2022.101149

Uddin, N., & Chakraborty, V. (2021). An investigation of the readability of sustainability reports. Journal of Emerging Technologies in Accounting, 19(1), 69–78. https://doi.org/10.2308/jeta-18-10-01-18

Uyar, A., Karaman, A. S., & Kilic, M. (2020). Is corporate social responsibility reporting a tool of signaling or greenwashing? Evidence from the worldwide logistics sector. Journal of Cleaner Production, 253, 119997. https://doi.org/10.1016/j.jclepro.2020.119997

Vigneau, L., & Adams, C. A. (2023). The failure of transparency as self-regulation. Sustainability Accounting Management and Policy Journal, 14(4), 852–876. https://doi.org/10.1108/sampj-01-2022-0051

Wang, M., Zhang, Y., & Gong, X. (2025). The Impacts of Social Credit Environment Improvement on Corporate ESG Greenwashing: Evidence from China. International Review of Economics & Finance, 104409. https://doi.org/10.1016/j.iref.2025.104409

Wan-Hussin, W. N., Qasem, A., Aripin, N., & Ariffin, M. S. M. (2021). Corporate Responsibility Disclosure, Information Environment and Analysts' Recommendations: Evidence from Malaysia. *Sustainability*, *13*(6), 3568. https://doi.org/10.3390/su13063568

Webb, B. (2023). Greenhushing is on the rise. What do fashion brands need to know? Vogue Business. https://www.voguebusiness.com/sustainability/greenhushing-is-on-the-rise-what-do-fashion-brands-need-to-know

Westerveld, J. (1986). *Essay on hotel environmental practices*.

Why sustainable business needs better ESG ratings | MIT Sloan. (2021). MIT Sloan. https://mitsloan.mit.edu/ideas-made-to-matter/why-sustainable-business-needs-better-esg-ratings?

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2^{nd} ed.). MIT Press.

Wu, Y., Zhang, K., & Xie, J. (2020). Bad greenwashing, good greenwashing: corporate social responsibility and information transparency. Management Science, 66(7), 3095–3112. https://doi.org/10.1287/mnsc.2019.3340

Wu, Z., Long, S., & Wang, Y. (2025). ESG greenwashing and organizational resilience: Exploring sustainable development paths from the perspective of unsystematic risk. Journal of Environmental Management, 390, 126395. https://doi.org/10.1016/j.jenvman.2025.126395

Xu, X., Li, Z., & Liu, F. (2025). Greenwashing in ESG information disclosure: An intertemporal signaling game approach. International Journal of Production Economics, 109674. https://doi.org/10.1016/j.ijpe.2025.109674

Yu, E. P., Van Luu, B., & Chen, C. H. (2020). Greenwashing in environmental, social and governance disclosures. Research in International Business and Finance, 52, 101192. https://doi.org/10.1016/j.ribaf.2020.101192

Yun, F., Lan, T., & Chen, Y. (2023). Cover-up or true? Does CSR disclosure really contribute to corporate environmental performance? Frontiers in Environmental Science, 11. https://doi.org/10.3389/fenvs.2023.1139088

Zainal, D., Zulkifli, N., & Saleh, Z. (2013). Corporate Social Responsibility Reporting in Malaysia: A research note. (2013). In *Journal of Accounting Perspectives* (Vol. 6, pp. 21–36). https://ajap.um.edu.my/index.php/AJAP/article/download/3684/1633/9498

Zervoudi, E., Moschos, N., & Christopoulos, A. (2025). From the Corporate Social Responsibility (CSR) and the Environmental, Social and Governance (ESG) Criteria to the Greenwashing Phenomenon: A Comprehensive Literature Review About the Causes, Consequences and Solutions of the Phenomenon with Specific Case Studies. https://doi.org/10.3390/su17052222

Zhang, K., Pan, Z., Janardhanan, M., & Patel, I. (2022). Relationship analysis between greenwashing and environmental performance. *Environment Development and Sustainability*, *25*(8), 7927–7957. https://doi.org/10.1007/s10668-022-02381-9

Zhou, K., Qu, Z., Liang, J., Tao, Y., & Zhu, M. (2024). Threat or shield: Environmental administrative penalties and corporate greenwashing. Finance Research Letters, 61, 105031. https://doi.org/10.1016/j.frl.2024.105031

APPENDICES

Appendix 1: List of the Sampled Firms

Rank	Name	Sector
1	Tenaga National Berhad	Utilities
2	IHH Healthcare Berhad	Healthcare
3	PRESS METAL Aluminium Holdings	Basic Materials
4	CelcomDigi Berhad	Technology
5	Petronas Gas Berhad	Utilities
6	YTL Power International Berhad	Utilities
7	SD Guthrie Berhad	Consumer Non-Cyclicals
8	Petronas Chemicals Group Berhad	Basic Materials
9	Sunway Berhad	Industrials
10	MISC Berhad	Industrials
11	Gamuda Berhad	Industrials
12	YTL Corporation Berhad	Utilities
13	Maxis Berhad	Technology
14	Telekom Malaysia Berhad	Technology
15	IOI Corporation Berhad	Consumer Non-Cyclicals
16	Axiata Group Berhad	Technology
17	Kuala Lumpur Kepong Berhad	Basic Materials
18	Nestle Malaysia Berhad	Consumer Non-Cyclicals
19	Petronas Dagangan Berhad	Energy
20	99 Speed Mart Retail Holding	Consumer Non-Cyclicals
21	Westports Holdings Berhad	Industrials
22	QL Resources Berhad	Consumer Non-Cyclicals
23	MR DIY Group (M) Berhad	Consumer Cyclicals
24	Sime Darby Berhad	Consumer Non-Cyclicals

25	PPB Group Berhad	Consumer Non-Cyclicals
26	United Plantations Berhad	Consumer Non-Cyclicals
27	Genting Malaysia Berhad	Consumer Cyclicals
28	IOI Properties Group Berhad	Others
29	KPJ Healthcare Berhad	Healthcare
30	Genting Berhad	Consumer Cyclicals
31	Dialog Group Berhad	Energy
32	IJM Corporation Berhad	Industrials
33	Fraser & Neave Holdings Berhad	Consumer Non-Cyclicals
34	Sime Darby Property Berhad	Others
35	Time Dotcom Berhad	Technology
36	Sunway Construction Berhad	Industrials
37	Malayan Cement Berhad	Basic Materials
38	Chin Hin Group Berhad	Consumer Cyclicals
39	Inari Amertron Berhad	Technology
40	Yinson Holdings Berhad	Energy
41	Eco-Shop Marketing Berhad	Consumer Non-Cyclicals
42	Batu Kawan Berhad	Basic Materials
43	Vitrox Corporation Berhad	Technology
44	Frontken Corporation Berhad	Industrials
45	Zetrix AI Berhad	Technology
46	Hap Seng Consolidated Berhad	Consumer Non-Cyclicals
47	Eco World Development Group Berhad	Others
48	Heineken Malaysia Berhad	Consumer Non-Cyclicals
49	Malaysian Pacific Industries Berhad	Technology
50	GAS Malaysia Berhad	Utilities
51	GreatTech Technology Berhad	Industrials

52	Carlsberg Brewery Malaysia Berhad	Consumer Non-Cyclicals
53	Tanco Holdings Berhad	Others
54	S P Setia Berhad	Others
55	Top Glove Corporation Berhad	Healthcare
56	Scientex Berhad	Basic Materials
57	Malakoff Corp Berhad	Utilities
58	UOA Development Berhad	Others
59	ITMAX System Berhad	Industrials
60	Hong Leong Industries Berhad	Consumer Non-Cyclicals
61	Genting Plantations Berhad	Consumer Non-Cyclicals
62	IGB Berhad	Others
63	Unisem (M) Berhad	Technology
64	Oriental Holdings Berhad	Consumer Cyclicals
65	OSK Holdings Berhad	Others
66	Kelington Group Berhad	Industrials
67	Hartalega Holdings Berhad	Healthcare
68	Farm Fresh Berhad	Consumer Non-Cyclicals
69	UEM Sunrise Berhad	Others
70	UWC Berhad	Industrials
71	Mega First Corporation Berhad	Utilities
72	Hextar Global Berhad	Basic Materials
73	Capital A Berhad	Industrials
74	Johor Plantations Group Berhad	Consumer Non-Cyclicals
75	UMS Integration	Others
76	Sarawak Oil Palms Berhad	Consumer Non-Cyclicals
77	Kossan Rubber Industries Berhad	Healthcare
78	Tropicana Corporation Berhad	Others

79	Kerjaya Prospek Group Berhad	Industrials
80	Mah Sing Group Berhad	Others
81	Nationagate Holdings Berhad	Technology
82	Matrix Concepts Holdings Berhad	Others
83	Pentamaster Corporation Berhad	Industrials
84	Guan Chong Berhad	Consumer Non-Cyclicals
85	Sam Engineering & Equipment Berhad	Industrials
86	DXN Holdings Berhad	Consumer Non-Cyclicals
87	Far East Holdings Berhad	Consumer Non-Cyclicals
88	Ranhill Utilities Berhad	Utilities
89	Hume Cement Industires Berhad	Basic Materials
90	Bintulu Port Holdings Berhad	Industrials
91	Hextar Technologies Solution Berhad	Basic Materials
92	WCE Holdings Berhad	Industrials
93	7-Eleven Malaysia Holdings Berhad	Consumer Non-Cyclicals
94	Leong Hup International Berhad	Consumer Non-Cyclicals
95	Oriental Kopi Holdings Berhad	Consumer Cyclicals
96	Malaysian Resources Corporation Berhad	Industrials
97	VS Industry Berhad	Technology
98	Kim Loong Resources Berhad	Consumer Non-Cyclicals
99	Solarvest Holdings Berhad	Utilities
100	KSL Holdings Berhad	Others

Appendix 2: Line of Codes for Greenwashing Scores

GW Scores

```
!pip install pandas numpy pymupdf

# For best tone accuracy (FinBERT):

!pip install torch torchvision torchaudio transformers scipy

# If your PDFs are scanned:

!pip install pdf2image pytesseract

# (and install the Tesseract OCR app; add it to PATH on Windows)


# ======================== CONFIG (EDIT THESE) ========================

reports_folder = r"E:\data\reports"     # folder with PDFs like "Nestle_2020.pdf"

esg_csv        = r"E:\data\ESG_scores.csv" # CSV with columns: firm_id, year, E_Pill
ar_Score (and optional industry)

output_csv     = r"E:\data\gw_scores.csv"


# Minimum env text per firm-year to be considered valid (0 disables)

MIN_ENV_WORDS  = 500


# Z-score level: 'year' or 'industry-year' (requires 'industry' column in esg_csv)

ZSCORE_LEVEL   = 'year'

# ==================================================================


import os, re, math, numpy as np, pandas as pd

from pathlib import Path


# -------------------- 1) TEXT EXTRACTION FROM PDF (WITH OCR AS BACKUP) ---------
-----------

def extract_text_from_pdf(pdf_path: str) -> str:
```

```
    text = ""
    try:
        import fitz  # PyMuPDF
        doc = fitz.open(pdf_path)
        for page in doc:
            text += page.get_text("text") + "\n"
        doc.close()
    except Exception:
        text = ""

    # If suspiciously short, try OCR (scanned PDFs)
    if len(text.strip()) < 200:
        try:
            from pdf2image import convert_from_path
            import pytesseract
            pages = convert_from_path(pdf_path, dpi=300)
            ocr_pages = [pytesseract.image_to_string(img, lang="eng") for img in pag
es]
            text = "\n".join(ocr_pages)
        except Exception:
            pass
    return text


def clean_light(text: str) -> str:
    text = re.sub(r'Page\s*\d+(\s*of\s*\d+)?', ' ', text, flags=re.I)
    text = re.sub(r'\s+', ' ', text).strip()
```

```
    return text


def split_paragraphs(text: str):

    # split on line breaks / paragraph-like breaks; keep non-empty

    paras = re.split(r'\n+|\r+|\.\s{2,}', text)

    return [p.strip() for p in paras if p.strip()]


# -------------------- 2) KEEP ENVIRONMENTAL PARAGRAPHS --------------------

ENV_KWS = [

    # Core environmental terms

    'sustainab','environment','emission','carbon','energy','water','waste','climat
e',

    'footprint','biodivers','hazard','pollut','renewable','scope','ghg','tcf','net z
ero',

    'greenhouse','recycling','recycle','efficiency','ecosystem','solar','wind','hydr
o',

    'geothermal','biomass','fossil fuel','deforestation','afforestation','reforestat
ion',

    'plastic','packaging','air quality','soil','agriculture','organic','chemical','t
oxic',

    'hazardous','safety','resource use','material','circular economy','life cycl
e','offset',

    'carbon neutrality','zero emission','carbon credit','environmental impact','ecol
ogy',

    'biodiversity','forest','tree planting','renewable energy','green energy','water
usage',

    'energy conservation','energy saving','waste management','wastewater','e-waste',

    'landfill','sustainable development','environmental management','iso 14001','es
g',

    'sustainability report','responsible sourcing','supply chain','net-zero','carbon
neutral',
```

```
    'environmental footprint','climate change','global warming','nature-based','ecos
ystem services',


    # Governance & policy related (environmental disclosures)

    'paris agreement','unfccc','sdg','sdgs','sdg13','sdg12','sustainable development
goals',

    'task force on climate','tcfd','environmental policy','environmental risk','envi
ronmental governance',

    'environmental target','environmental performance','sustainability framework','s
cience based target',


    # Corporate initiatives / disclosures

    'csr','corporate responsibility','social responsibility','sustainable finance',

    'green bond','sri sukuk','low-carbon','transition plan','sustainability strateg
y',

    'environmental innovation','environmental initiative','green initiative','eco-fr
iendly',

    'environmental compliance','environmental audit','environmental monitoring','sus
tainable production',

    'environmental conservation','nature conservation','preservation','rehabilitatio
n',

    'climate risk','climate adaptation','climate resilience','carbon offset','carbon
trading',


    # Newer reporting / metrics

    'cdp','gria','gria report','gri','global reporting initiative','sasb','integrate
d report',

    'integrated reporting','bursa sustainability','bnm taxonomy','principles-based t
axonomy',

    'sc srr','sri roadmap','green taxonomy'

]
```

```
def keep_env_paragraphs(paragraphs, min_words=25):
    kept = []
    for p in paragraphs:
        lw = p.lower()
        if len(p.split()) >= min_words or any(k in lw for k in ENV_KWS):
            kept.append(p)
    return kept

# ──────────────────── 3) FILENAME → firm_id, year ─────────────────────────
def parse_firm_year(filename: str):
    base = os.path.splitext(os.path.basename(filename))[0]
    m = re.search(r'(20\d{2})', base)           # look for 2019..2029 etc.
    year = int(m.group(1)) if m else None
    firm = re.sub(r'[_\-\s]*20\d{2}.*$', '', base)
    firm = (firm.lower().replace('_',' ').replace('-',' ').strip())
    firm = re.sub(r'\s+berhad$', '', firm).strip()
    firm = re.sub(r'\sbhd$', '', firm).strip()
    firm = re.sub(r'\(m\)$', '', firm).strip()
    return firm, year

# ──────────────────── 4) SENTIMENT ENGINE (FinBERT → VADER) ───────────────
def load_sentiment_engine():
    """
    Returns a dict: {'engine': 'finbert'|'vader', 'score_paragraph': fn}
    score_paragraph(paragraph_text) -> (pos_prob in [0,1], paragraph_word_count)
    """
```

```
    try:

        import torch

        from transformers import AutoTokenizer, AutoModelForSequenceClassification

        from scipy.special import softmax

        DEVICE   = "cuda" if torch.cuda.is_available() else "cpu"

        MODEL_ID = "yiyanghkust/finbert-tone"

        tokenizer = AutoTokenizer.from_pretrained(MODEL_ID)

        model     = AutoModelForSequenceClassification.from_pretrained(MODEL_ID).to
(DEVICE)

        model.eval()


        def chunk_tokens(text, max_len=512, stride=300):

            tokens = tokenizer.encode(text, add_special_tokens=False)

            if len(tokens) <= max_len:

                return [text]

            chunks, start = [], 0

            while start < len(tokens):

                end = min(start + max_len, len(tokens))

                piece = tokenizer.decode(tokens[start:end])

                chunks.append(piece)

                if end == len(tokens): break

                start = max(end - stride, 0)

            return chunks


        def finbert_posprob(text):

            inputs = tokenizer(text, return_tensors="pt", truncation=True, max_lengt
h=512).to(DEVICE)
```

```
        with torch.no_grad():
            logits = model(**inputs).logits
        probs = softmax(logits.detach().cpu().numpy())[0]  # [neg, neu, pos]
        token_count = int(inputs["input_ids"].shape[1])
        return float(probs[2]), token_count

    def score_paragraph(paragraph_text):
        chunks = chunk_tokens(paragraph_text, max_len=512, stride=300)
        numer, denom = 0.0, 0
        for ch in chunks:
            ppos, tl = finbert_posprob(ch)
            numer += ppos * tl
            denom += tl
        para_words = sum(len(c.split()) for c in chunks)
        return (numer/denom) if denom>0 else np.nan, para_words

    return {'engine':'finbert','score_paragraph':score_paragraph}

except Exception:
    # VADER fallback (no GPU/torch needed)
    import nltk
    try:
        nltk.data.find('sentiment/vader_lexicon.zip')
    except LookupError:
        nltk.download('vader_lexicon')
    from nltk.sentiment.vader import SentimentIntensityAnalyzer
```

```
    sia = SentimentIntensityAnalyzer()


    def score_paragraph_vader(paragraph_text):

        comp = sia.polarity_scores(paragraph_text)['compound']      # [-1,1]

        posprob = (comp + 1.0) / 2.0                                # [0,1]

        return posprob, len(paragraph_text.split())


    return {'engine':'vader','score_paragraph':score_paragraph_vader}


# -------------------- 5) POSITIVETONE AGGREGATION -------------------------

def compute_positive_tone(df_paras, score_paragraph_fn):

    """

    Input df_paras: columns ['firm_id','year','paragraph_text']

    Output DataFrame: firm_id, year, PositiveTone, total_env_words

    """

    rows = []

    for (fid, yr), g in df_paras.groupby(['firm_id','year']):

        para_scores, para_weights, total_words = [], [], 0

        for txt in g['paragraph_text']:

            s, w = score_paragraph_fn(txt)

            if not (s is None or (isinstance(s, float) and math.isnan(s))):

                para_scores.append(s)

                para_weights.append(max(w,1))

                total_words += w

        if len(para_scores) == 0:

            rows.append((fid, yr, np.nan, total_words))
```

```
        else:

            pos_doc = np.average(para_scores, weights=para_weights)

            rows.append((fid, yr, float(pos_doc), total_words))

    return pd.DataFrame(rows, columns=['firm_id','year','PositiveTone','total_env_wo
rds'])


# -------------------- 6) LOAD ESG CSV (accepts exact columns) ---------------

def load_esg_csv(esg_csv_path):

    df = pd.read_csv(esg_csv_path)

    # if exact columns are present, use them

    if set(['firm_id','year','E_Pillar_Score']).issubset(df.columns):

        pass

    else:

        # try a light auto-map (extend as needed)

        def _find_col(cols, pats):

            cols_l = [c.lower() for c in cols]

            for pat in pats:

                for i,c in enumerate(cols_l):

                    if re.search(pat, c):

                        return cols[i]

            return None

        firm_col = _find_col(df.columns, [r'^company\s*name', r'\bfirm_id\b', r'\bfi
rm\b', r'\bentity\b'])

        year_col = _find_col(df.columns,  [r'\bfiscal.*year\b', r'\byear\b'])

        e_col    = _find_col(df.columns,  [r'\bE[_\s\-]*pillar.*score\b', r'\benviro
nment.*pillar.*score\b', r'\bE_pillar_score\b'])

        if not (firm_col and year_col and e_col):
```

```
            raise ValueError(f"Cannot detect columns; CSV has: {list(df.columns)}")

        df = df.rename(columns={firm_col:'firm_id', year_col:'year', e_col:'E_Pillar
_Score'})


    # normalize firm names similar to filename parsing

    df['firm_id'] = (df['firm_id'].astype(str).str.lower()

                     .str.replace(r'[_\-]+',' ', regex=True)

                     .str.replace(r'\s+berhad$','', regex=True)

                     .str.replace(r'\sbhd$','', regex=True)

                     .str.replace(r'\(m\)$','', regex=True)

                     .str.strip())

    df['year'] = pd.to_numeric(df['year'], errors='coerce').astype('Int64')

    df = df.dropna(subset=['year','E_Pillar_Score']).copy()

    df['year'] = df['year'].astype(int)

    df['E_Pillar_Score'] = pd.to_numeric(df['E_Pillar_Score'], errors='coerce')

    df = df.dropna(subset=['E_Pillar_Score']).copy()

    return df


# --------------------- 7) Z-SCORES & GWSCORE ----------------------------

def zscore_series(s: pd.Series) -> pd.Series:

    mu = s.mean()

    sd = s.std(ddof=0)

    if (sd is None) or (sd == 0) or np.isnan(sd):

        return s*0

    return (s - mu) / sd


def add_zscores(df, level='year'):
```

```
    if level == 'industry-year' and 'industry' in df.columns:

        df['z_PosTone'] = df.groupby(['industry','year'])['PositiveTone'].transform
(zscore_series)

        df['z_EPillar'] = df.groupby(['industry','year'])['E_Pillar_Score'].transfor
m(zscore_series)

    else:

        df['z_PosTone'] = df.groupby('year')['PositiveTone'].transform(zscore_serie
s)

        df['z_EPillar'] = df.groupby('year')['E_Pillar_Score'].transform(zscore_seri
es)

    df['GWScore'] = df['z_PosTone'] - df['z_EPillar']

    return df


# ------------------- 8) MAIN PIPELINE ------------------------------

def main():

    # A) Load ESG

    df_ESG = load_esg_csv(esg_csv)

    print("Loaded ESG CSV:", df_ESG.shape, list(df_ESG.columns))


    # B) Build paragraph table from PDFs

    pdfs = [str(p) for p in Path(reports_folder).glob("*.pdf")]

    if not pdfs:

        raise FileNotFoundError(f"No PDFs found in: {reports_folder}")

    rows = []

    for pdf in pdfs:

        firm, year = parse_firm_year(pdf)

        if not firm or not year:

            print("Skipping (cannot parse firm/year):", pdf)
```

```
            continue

        text = extract_text_from_pdf(pdf)

        text = clean_light(text)

        paras = split_paragraphs(text)

        env_paras = keep_env_paragraphs(paras, min_words=25)

        for p in env_paras:

            rows.append((firm, year, p))

    df_paras = pd.DataFrame(rows, columns=['firm_id','year','paragraph_text'])

    if df_paras.empty:

        raise RuntimeError("No environment-related paragraphs extracted. Loosen filt
ers or check OCR.")

    print("Paragraph rows:", df_paras.shape)


    # C) Sentiment engine

    eng = load_sentiment_engine()

    print("Sentiment engine:", eng['engine'])


    # D) PositiveTone per firm-year (+ coverage)

    df_Pos = compute_positive_tone(df_paras, eng['score_paragraph'])

    print("Firm-years (before coverage filter):", df_Pos.shape[0])

    if MIN_ENV_WORDS and MIN_ENV_WORDS > 0:

        before = df_Pos.shape[0]

        df_Pos.loc[df_Pos['total_env_words'] < MIN_ENV_WORDS, 'PositiveTone'] = np.n
an

        flagged = df_Pos['PositiveTone'].isna().sum()

        print(f"Applied MIN_ENV_WORDS={MIN_ENV_WORDS}. Flagged NA: {flagged}/{befor
e}")
```

```
    # E) Merge + z-scores + GWScore

    keep_cols = ['firm_id','year','E_Pillar_Score'] + (['industry'] if 'industry' in
df_ESG.columns else [])

    df = pd.merge(df_Pos[['firm_id','year','PositiveTone']],

                  df_ESG[keep_cols],

                  on=['firm_id','year'],

                  how='inner')

    df = df.dropna(subset=['PositiveTone','E_Pillar_Score']).copy()

    level = 'industry-year' if (ZSCORE_LEVEL=='industry-year' and 'industry' in df.c
olumns) else 'year'

    df = add_zscores(df, level=level)



    # F) Save

    df.sort_values(['firm_id','year']).to_csv(output_csv, index=False)

    print(f"Saved {df.shape[0]} firm-year rows → {output_csv}")

    if not df.empty:

        print(df.head(10).to_string(index=False))



if __name__ == "__main__":

    # If you plan to use NLTK anywhere else (e.g., sentence splitting in other cod
e),

    # ensure punkt resources are present; otherwise ignore.

    try:

        import nltk

        try:

            nltk.data.find('tokenizers/punkt')

        except LookupError:
```

```
        nltk.download('punkt')

        # Some newer NLTK versions require punkt_tab as well:

        try: nltk.download('punkt_tab')

        except Exception: pass

except Exception:

    pass


main()
```

Appendix 3: Line of Codes for Readability Scores

Readability Scores

```
# === Part A: config & imports ===

# (edit these two paths)

reports_folder = r"E:\data\reports"   # PDFs like Firm_2022.pdf
output_csv     = r"E:\data\readability_scores.csv"

# Optional: SMOG needs >=3 sentences
MIN_SENTENCES = 3

import os, re, math
import numpy as np
import pandas as pd
from pathlib import Path

pip install pymupdf
pip install pdf2image pytesseract

# === Part B: text extraction from PDFs ===

def extract_text_from_pdf(pdf_path: str) -> str:
    text = ""
    # 1) Try PyMuPDF (fast, works for most digital PDFs)
    try:
        import fitz  # PyMuPDF
        doc = fitz.open(pdf_path)
        for page in doc:
            text += page.get_text("text") + "\n"
        doc.close()
    except Exception:
        text = ""

    # 2) If very short → probably scanned; try OCR
    if len(text.strip()) < 200:
        try:
            from pdf2image import convert_from_path
            import pytesseract
            pages = convert_from_path(pdf_path, dpi=300)
            chunks = [pytesseract.image_to_string(img, lang="eng") for img in pages]
            text = "\n".join(chunks)
        except Exception:
            # If OCR unavailable, we'll return whatever we got (may be empty)
            pass
    return text

# === Part C: cleaning & tokenization ===
# Keep technical tokens like CO2, %, numbers. Remove page artifacts.

def clean_text(text: str) -> str:
```

```
    text = re.sub(r'Page\s*\d+(\s*of\s*\d+)?', ' ', text, flags=re.I)
    text = re.sub(r'\s+', ' ', text).strip()
    return text

# Sentence split: use NLTK if available; else regex fallback
try:
    import nltk
    try:
        nltk.data.find('tokenizers/punkt')
    except LookupError:
        nltk.download('punkt')
    from nltk.tokenize import sent_tokenize
    def split_sentences(text: str):
        return [s.strip() for s in sent_tokenize(text) if s.strip()]
except Exception:
    def split_sentences(text: str):
        parts = re.split(r'(?<=[\.\?\!])\s+(?=[A-Z(])', text)
        return [s.strip() for s in parts if s.strip()]

WORD_RE = re.compile(r"[A-Za-z]+(?:'[A-Za-z]+)?")  # words like don't
def split_words(text: str):
    return WORD_RE.findall(text)

# === Part D: syllable count ===

def _syllables_pyphen(word: str):
    try:
        import pyphen
        dic = pyphen.Pyphen(lang='en_US')
        parts = dic.inserted(word.lower())
        if parts:
            return max(1, parts.count('-') + 1)
        return 1
    except Exception:
        return None

def _syllables_heuristic(word: str):
    w = re.sub(r'[^a-zA-Z]', '', word).lower()
    if not w:
        return 0
    # remove trailing silent 'e'
    if len(w) > 2 and w.endswith('e'):
        w = w[:-1]
    groups = re.findall(r'[aeiouy]+', w)
    return max(1, len(groups))

def count_syllables(word: str):
    s = _syllables_pyphen(word)
    return s if s is not None else _syllables_heuristic(word)

# === Part E: readability metrics ===
```

```python
def readability_scores(text: str, min_sentences: int = 3) -> dict:
    sents = split_sentences(text)
    words = []
    for s in sents:
        words.extend(split_words(s))

    n_sent = len(sents)
    n_words = len(words)
    n_chars = sum(len(w) for w in words)

    if n_words == 0 or n_sent == 0:
        return {
            'sentences': n_sent, 'words': n_words, 'characters': n_chars,
            'avg_words_per_sentence': np.nan,
            'avg_chars_per_word': np.nan,
            'pct_complex_words': np.nan,
            'flesch_reading_ease': np.nan,
            'flesch_kincaid_grade': np.nan,
            'gunning_fog': np.nan,
            'smog_index': np.nan,
            'automated_readability_index': np.nan
        }

    syllables = [count_syllables(w) for w in words]
    total_syll = sum(syllables)
    complex_words = sum(1 for s in syllables if s >= 3)  # "polysyllables"

    words_per_sent = n_words / n_sent
    chars_per_word = n_chars / n_words
    syll_per_word  = total_syll / n_words
    pct_complex    = (complex_words / n_words) * 100.0

    # Flesch Reading Ease (higher = easier)
    fre  = 206.835 - 1.015 * words_per_sent - 84.6 * syll_per_word
    # Flesch-Kincaid Grade Level
    fkgl = 0.39 * words_per_sent + 11.8 * syll_per_word - 15.59
    # Gunning Fog Index
    fog  = 0.4 * (words_per_sent + 100.0 * (complex_words / n_words))
    # SMOG (needs enough sentences)
    if n_sent >= min_sentences and complex_words > 0:
        smog = 1.0430 * math.sqrt(complex_words * (30.0 / n_sent)) + 3.1291
    else:
        smog = np.nan
    # Automated Readability Index
    ari  = 4.71 * (n_chars / n_words) + 0.5 * words_per_sent - 21.43

    return {
        'sentences': n_sent,
        'words': n_words,
        'characters': n_chars,
        'avg_words_per_sentence': words_per_sent,
        'avg_chars_per_word': chars_per_word,
```

```
        'pct_complex_words': pct_complex,
        'flesch_reading_ease': fre,
        'flesch_kincaid_grade': fkgl,
        'gunning_fog': fog,
        'smog_index': smog,
        'automated_readability_index': ari
    }

# === Part F (optional): filter environmental paragraphs ===

ENV_KWS = [
    'environment','emission','carbon','energy','water','waste','climate',
    'biodivers','scope','ghg','net zero','tcf d','pollut','renewable'
]

def keep_env_only(text: str) -> str:
    # split into rough sentences, keep those matching any keyword
    parts = re.split(r'(?<=[\.\?\!])\s+', text)
    kept = [p for p in parts if any(k in p.lower() for k in ENV_KWS)]
    return ' '.join(kept) if kept else text  # fall back to full text if none matche
d

# === Part G: filename parsing (Firm_YYYY.pdf) ===

def parse_firm_year(filename: str):
    base = os.path.splitext(os.path.basename(filename))[0]
    m = re.search(r'(20\d{2})', base)  # first year like 2019..2029
    year = int(m.group(1)) if m else None
    firm = re.sub(r'[_\-\s]*20\d{2}.*$', '', base)
    firm = (firm.lower().replace('_',' ').replace('-',' ').strip())
    firm = re.sub(r'\s+berhad$', '', firm).strip()
    firm = re.sub(r'\sbhd$', '', firm).strip()
    firm = re.sub(r'\(m\)$', '', firm).strip()
    return firm, year

# === Part H: run over folder and compute metrics ===

def process_reports(folder: str, env_only: bool = False) -> pd.DataFrame:
    pdfs = [str(p) for p in Path(folder).glob("*.pdf")]
    if not pdfs:
        raise FileNotFoundError(f"No PDFs found in: {folder}")

    rows = []
    for pdf in pdfs:
        firm, year = parse_firm_year(pdf)
        if not firm or not year:
            # skip files without parsable firm/year
            continue

        raw = extract_text_from_pdf(pdf)
        txt = clean_text(raw)
        if env_only:
```

```
            txt = keep_env_only(txt)  # focus on "E" sections

        scores = readability_scores(txt, min_sentences=MIN_SENTENCES)
        rows.append({'firm_id': firm, 'year': year, **scores})

    return pd.DataFrame(rows)

# === Part I: execute and save ===

if __name__ == "__main__":
    df = process_reports(reports_folder, env_only=False)  # set True for E-only read
ability
    df.sort_values(['firm_id','year']).to_csv(output_csv, index=False)
    print(f"Saved {df.shape[0]} rows → {output_csv}")
    if not df.empty:
        display(df.head(10))
```

Appendix 4: Line of Codes for Disaggregation Scores

Disaggregation Score

```
# ========================= CONFIG (EDIT) =========================
FOLDER = r"E:\data\financialstatement"     # folder of *.xlsx like AXIA.KL.xlsx
# optional CSV with columns: firm_id, industry
OUTPUT_CSV = r"E:\data\disclosure_disaggregation_dq.csv"
# ================================================================

import os, re, pandas as pd, numpy as np
from pathlib import Path

# ---------- utilities ----------
NULL_LIKE = {"", "-", "—", "–", "na", "n/a", "nil", "none"}
EXCLUDE_LABEL_REGEX = r"""(?ix)
^\s*(total|sum|subtotal|grand\s*total|balance\s*carried|balance\s*brought|brought\s*forward|
notes?|company|group|consolidated|for\s+the\s+year|current\s+year|previous\s+year|
ytd|quarter|unaudited|audited)\b
"""

BS_HINTS = ["balance sheet", "financial position", "assets", "liabilities", "equity"]
IS_HINTS = ["income statement", "profit or loss", "statement of profit", "revenue", "sales", "operating profit"]
INDUSTRY_CSV = None

def _clean_firm(s: str) -> str:
    s = str(s).replace("_"," ").replace("-"," ").strip().lower()
    s = re.sub(r"\.[a-z]{2,4}$", "", s)  # .KL, .MY
    s = re.sub(r"\s+berhad$", "", s)
    s = re.sub(r"\sbhd$", "", s)
    s = re.sub(r"\(m\)$", "", s)
    s = re.sub(r"\s+", " ", s)
    return s

def firm_from_filename(p: str) -> str:
    return _clean_firm(os.path.splitext(os.path.basename(p))[0])

def _is_numericish_scalar(v) -> bool:
    if isinstance(v, (int, float, np.integer, np.floating)):
        return not (pd.isna(v))
    if v is None or (isinstance(v, float) and pd.isna(v)):
        return False
    s = str(v).strip().replace(",", "")
    if s.lower() in NULL_LIKE: return False
    s = re.sub(r"\s+", "", s)
    return bool(re.fullmatch(r"\(?-?\d+(\.\d+)?\)?", s))

def _any_numericish(values) -> bool:
    """Return True if ANY value in an iterable is numericish."""
    for v in (values if isinstance(values, (list, tuple)) else [values]):
        if _is_numericish_scalar(v):
            return True
    return False
```

```
def _header_year(x) -> int|None:
    if pd.isna(x): return None
    m = re.search(r"(20\d{2})", str(x))
    return int(m.group(1)) if m else None

def _likely_label_col_idx(df: pd.DataFrame) -> int|None:
    # return first object-typed column index; else 0
    for idx, c in enumerate(df.columns):
        if df[c].dtype == object:
            return idx
    return 0 if df.shape[1] else None

def _is_excluded_label(lbl: str) -> bool:
    if not isinstance(lbl, str): return True
    lab = re.sub(r"\s+", " ", str(lbl).replace("\xa0"," ")).strip().lower()
    if not lab or len(lab) < 2: return True
    if re.search(EXCLUDE_LABEL_REGEX, lab): return True
    if re.fullmatch(r"[\d\W]+", lab): return True
    return False

def _sheet_name_looks_like(stmt: str, sheet_name: str, sample_labels: list[str]) ->
bool:
    s = (sheet_name or "").lower()
    hints = BS_HINTS if stmt == "BS" else IS_HINTS
    if any(h in s for h in hints):
        return True
    joined = " ".join(sample_labels).lower()
    return any(h in joined for h in hints)

# ---------- header detection (by POSITION) ----------
def _find_year_header_row_positions(raw: pd.DataFrame, scan_top=20):
    """
    Find header row and RETURN:
      - header_row_idx (int)
      - year_pos_map: dict[int_col_position -> int_year]
      - dup_year_cols: dict[int_year -> list[int_col_positions]]
    """
    R = min(scan_top, raw.shape[0])
    best_r, best_map = -1, {}
    for r in range(R):
        mapping = {}
        for c in range(raw.shape[1]):
            y = _header_year(raw.iat[r, c])
            if y: mapping[c] = y
        if len(mapping) >= 2:
            best_r, best_map = r, mapping
            break
    if best_r < 0:
        return -1, {}, {}
    # build duplicate index mapping
    dup = {}
    for c_pos, y in best_map.items():
        dup.setdefault(y, []).append(c_pos)
    return best_r, best_map, dup
```

```
def _make_headered_by_position(raw: pd.DataFrame):
    """
    Promote a row to header if needed, then return:
      df (with headers set),
      year_pos_map (pos->year) using the NEW header row positions,
      dup_year_cols (year->list[pos]).
    """
    # 1) Try as-is (assume first row already header). Build position map from header labels.
    df0 = raw.copy()
    df0.columns = [str(c).replace("\n"," ").strip() for c in df0.columns]
    pos_map0 = {}
    for j, c in enumerate(df0.columns):
        y = _header_year(c)
        if y: pos_map0[j] = y
    if len(pos_map0) >= 2:
        # dup map
        dup0 = {}
        for pos, y in pos_map0.items():
            dup0.setdefault(y, []).append(pos)
        return df0, pos_map0, dup0

    # 2) Promote a row that has ≥2 year labels
    hdr_idx, pos_map, dup = _find_year_header_row_positions(raw)
    if hdr_idx >= 0:
        cols = [str(x).strip() for x in list(raw.iloc[hdr_idx, :])]
        body = raw.iloc[hdr_idx+1:, :].copy()
        body.columns = cols
        # Re-build position map from promoted header row (use positions!)
        new_pos_map = {}
        for j, c in enumerate(body.columns):
            y = _header_year(c)
            if y: new_pos_map[j] = y
        if len(new_pos_map) >= 2:
            dup_new = {}
            for pos, y in new_pos_map.items():
                dup_new.setdefault(y, []).append(pos)
            return body, new_pos_map, dup_new

    # 3) Fail
    return raw, {}, {}

# ---------- counting ----------
def _count_items_by_position(df: pd.DataFrame, year_pos_map: dict[int,int], dup_year_cols: dict[int,list[int]]) -> dict[int,int]:
    """
    Count unique line items with numeric values for each YEAR using COLUMN POSITIONS.
    If a year has duplicate columns, we treat it as numeric if ANY of the dup columns is numeric.
    """
    if not year_pos_map:
        return {}

```

```
    label_col_idx = _likely_label_col_idx(df)
    if label_col_idx is None:
        return {}

    counts = {}
    n_rows, n_cols = df.shape
    for i in range(n_rows):
        raw_label = df.iat[i, label_col_idx]
        if not isinstance(raw_label, str):
            continue
        lab = re.sub(r"\s+", " ", str(raw_label).replace("\xa0"," ")).strip()
        if _is_excluded_label(lab):
            continue

        # evaluate per year
        for year, col_positions in dup_year_cols.items():
            # grab all values across duplicate columns (if any)
            vals = []
            for pos in col_positions:
                if pos < n_cols:
                    vals.append(df.iat[i, pos])
            if _any_numericish(vals):
                counts.setdefault(year, set()).add(lab.lower())

    return {yr: len(s) for yr, s in counts.items()}

# ---------- IO helpers ----------
def _read_any(xls: pd.ExcelFile, sheet: str) -> pd.DataFrame|None:
    try:
        df0 = pd.read_excel(xls, sheet_name=sheet, engine="openpyxl")
        if df0.shape[0] and df0.shape[1]:
            return df0
    except Exception:
        pass
    try:
        raw = pd.read_excel(xls, sheet_name=sheet, header=None, engine="openpyxl")
        return raw
    except Exception:
        return None

# ---------- per-file ----------
def extract_counts_from_file(path: str) -> pd.DataFrame:
    firm = firm_from_filename(path)
    try:
        xls = pd.ExcelFile(path, engine="openpyxl")
    except Exception as e:
        print(f"[ERR] open {os.path.basename(path)}: {e}")
        return pd.DataFrame(columns=["firm_id","year","stmt","items"])

    rows = []
    for sheet in xls.sheet_names:
        raw = _read_any(xls, sheet)
        if raw is None or raw.empty:
            continue

```

```
        df, year_pos_map, dup_year_cols = _make_headered_by_position(raw)
        if len(year_pos_map) < 2:
            continue

        # Decide BS or IS using sheet name / sample labels
        label_idx = _likely_label_col_idx(df)
        sample_labels = []
        if label_idx is not None and df.shape[0]:
            colname = df.columns[label_idx]
            sample_labels = [str(x) for x in df[colname].dropna().astype(str).head(20).tolist()]

        if _sheet_name_looks_like("BS", sheet, sample_labels):
            stmt = "BS"
        elif _sheet_name_looks_like("IS", sheet, sample_labels):
            stmt = "IS"
        else:
            continue

        year_counts = _count_items_by_position(df, year_pos_map, dup_year_cols)
        for yr, cnt in year_counts.items():
            rows.append([firm, int(yr), stmt, int(cnt)])

    return pd.DataFrame(rows, columns=["firm_id","year","stmt","items"])

# ---------- z-score ----------
def zscore(s: pd.Series) -> pd.Series:
    mu, sd = s.mean(), s.std(ddof=0)
    if sd == 0 or np.isnan(sd): return s*0
    return (s - mu) / sd

# ---------- main ----------
def main():
    files = [str(p) for p in Path(FOLDER).glob("*.xlsx") if not p.name.startswith("~$")]
    if not files:
        raise FileNotFoundError(f"No .xlsx files in {FOLDER}")

    parts = []
    for f in files:
        df = extract_counts_from_file(f)
        if df.empty:
            print(f"[SKIP] {os.path.basename(f)} → no FS found or no year headers.")
        else:
            parts.append(df)
            print(f"[OK]   {os.path.basename(f)} → {len(df)} stmt-year rows")

    if not parts:
        # quick sheet listing to help tweak hints
        for f in files:
            try:
                xls = pd.ExcelFile(f, engine="openpyxl")
                print(f"> {os.path.basename(f)} sheets: {xls.sheet_names}")
            except Exception as e:
```

```
                print(f"> {os.path.basename(f)} open error: {e}")
        raise RuntimeError("No statements parsed in any file. Add phrases to BS_HINTS / IS_HINTS to match your sheet names.")

    stmt_counts = pd.concat(parts, ignore_index=True)

    # Sum BS + IS per firm-year
    agg = (stmt_counts.groupby(["firm_id","year"], as_index=False)["items"]
           .sum().rename(columns={"items":"FS_ItemsCount"}))

    # Attach industry (optional)
    if INDUSTRY_CSV and Path(INDUSTRY_CSV).exists():
        ind = pd.read_csv(INDUSTRY_CSV)
        ind["firm_id"] = ind["firm_id"].astype(str).apply(_clean_firm)
        agg["firm_id"] = agg["firm_id"].astype(str).apply(_clean_firm)
        agg = agg.merge(ind[["firm_id","industry"]], on="firm_id", how="left")
    else:
        agg["industry"] = pd.NA

    # z-scores
    agg["z_FS_ItemsCount_year"] = agg.groupby("year")["FS_ItemsCount"].transform(zscore)
    if agg["industry"].notna().any():
        agg["z_FS_ItemsCount_industry_year"] = agg.groupby(["industry","year"])["FS_ItemsCount"].transform(zscore)
    else:
        agg["z_FS_ItemsCount_industry_year"] = pd.NA

    agg = agg.sort_values(["firm_id","year"])
    agg.to_csv(OUTPUT_CSV, index=False)
    print(f"\nSaved → {OUTPUT_CSV}")
    print(agg.head(12).to_string(index=False))

if __name__ == "__main__":
    main()
```