UNIVERSITY OF MACEDONIA

SCHOOL OF BUSSINESS ADMINISTRATION

Master Program in BUSINESS ANALYTICS AND DATA SCIENCE

Master Dissertation

Exploratory Analysis and Development of ESG Corporate Ratings

Elisavet Sotiriadou
bad20028

Supervisors:
Dr. Georgiou Andreas Professor Department of Business Administration

Examiners:
Dr. Moschidis Odisseas Professor Department of Business Administration
Dr. Kaparis Konstantinos Assistant Professor Department of Business Administration

Thessaloniki, March 2023

Table of Contents

List of Tables

List of Figures

Abstract

Environmental, social, and governance (ESG) risks have become increasingly important for businesses, investors, and stakeholders. Effective ESG risk management requires the development of robust ESG rating models that accurately capture the different dimensions of ESG risks. However, the existing ESG models have several limitations, including the lack of consensus on materiality criteria and the absence of global reporting standards. This study categorizes the existing ESG models into market-based and academic research-based models, highlighting their main differences. Furthermore, after analysing the latest updates regarding ESG regulations and global reporting standards, the study emphasizes the need for consensus on materiality criteria for each sector and regulatory cooperation among institutions in different countries or sectors to achieve ESG goals. Overall, the global ESG measurement systems are still incomplete, leaving room for further improvement. The study aims to develop an ESG risk rating model, which explores the conditional independence of individual risks using three multivariable methods Logit/Probit and Panel Data Regression. The statistical data collected from global businesses includes cases with both the presence and absence of ESG events and a wide range of predictive variables that refer to financial, administrative, and managerial indicators, both common and idiosyncratic. The results indicate that larger and financially stronger companies, particularly in essential industries, are more prone to experiencing ESG incidents.

Keywords

ESG score, ESG rating, ESG risk rating, financial performance, logistic regression, regression model, panel data regression

Chapter 1: Introduction

The ESG (Environmental, Social and Governance) risk analysis as well as the corresponding management practices, are gaining increasing ground in both operational decisions and asset management strategies. Furthermore, they have gained interest of government policies at national and international level. This development is emphatically reflected on the sustainable development goals of the United Nations (United Nations Sustainable Development Goals, SDGs). The ESG risk assessment is a rapidly emerging practice that aims to support adequately informed business, investment, and financial decisions.

The business discussions regarding business operations are often referring to environmental and social risks or other governance issues. As all those can increase the external risk of companies, ESG ratings measure how a company combines E, S and G, into business operations, as well as the company business's impact and sustainability. Thus, companies are seeking alternative ways to reduce this risk by improving their risk management capabilities and disclosure their ESG risk rating. For example, an ESG-conscious organization can enhance employee productivity and motivation. Irrespective of the size of the company or organization, incorporating ESG risk factors in the decision-making process amounts to effective risk management to avoid financial or reputational damage.

Logit and probit models are the statistical models used in this study to predict the probability of an event taking place, as the dependent variable. Furthermore, the panel data regression is used to analyze the two-dimension time-series data. A few ESG incidents were collected to be used as data sample for this study. The ESG incidents are considered binary data and can take only one of two values: 'non-default' or 'default'. Specifically, disclosure data in Bloomberg articles and other public papers were the source for collecting the incidents. Moreover, European but only listed companies have been chosen, basically from nations in European Union and UK. The data sample refers to 50 firms and the ESG events are collected from the timeframe of 2015 to 2021, while financial indicators used as the explanatory variables are chosen based on the previous year which means the timeline of 2014 to 2020. This is because the true purpose is to predict the next year's ESG probabilities by applying the current year's data. Based on models results a sensitivity analysis takes place and an out-of-sample forecasting using data from 2015 to 2020 and trying to estimate the probability of ESG incident during the year of 2021.

The structure of this Theses is as follows. Section 2 presents the related literature review analyzing the market-based models and academic results based on those models and their main differences. Section 3 presents a review of the related regulations including all the latest updates regarding ESG. Section 4 is the analysis of the of the empirical results regarding the model development, regression analysis, sensitivity analysis and out-of-sample forecasting. Finally, Section 5 provides the conclusion of this thesis by summarizing the analyzed results, discussing any model or method limitations, and providing suggestions for further research.

Chapter 2: Literature Review

ESG stands for Environmental, Social, and Governance. It is a set of criteria used to evaluate a company's performance in these three areas, which are believed to have a significant impact on the sustainability and long-term success of the business. Gradually gaining attention, the environmental, social and governance issues became rather critical. The official definitions of each component are:

Environmental factors include a company's impact on natural resources and the environment, such as its carbon footprint, energy usage, waste management, and water consumption.

Social factors include a company's impact on people, communities, and society, such as its labor practices, diversity and inclusion policies, product safety, and human rights record.

Governance factors include a company's internal policies and procedures that affect its decision-making and transparency, such as its board composition, executive compensation, risk management, and ethical standards (Peterdy, 2023).

Another related concept is Corporate Social Responsibility (CSR). ESG and CSR are related concepts but have distinct differences. ESG considers factors such as a company's carbon emissions, labor practices, and executive compensation. On the other hand, CSR is a voluntary initiative that companies undertake to take responsibility for their impact on society and the environment. CSR activities can include philanthropy, ethical labor practices, and sustainable business practices. (UNIDO, n.d.)

ESG criteria are used by investors, ratings agencies, and other stakeholders to assess a company's overall sustainability performance and potential risks and opportunities. The economic aspect is encompassed within the ESG concept, but it is not explicitly stated. ESG is categorized as a non-financial metric, however, it is relied upon by investors, academics, and regulators. The ESG adoption, will be one of the most important drivers of business development and entrepreneurship, like society, is impossible to stand still in the face of the continuous changes. ESG adoption refers to the process of integrating environmental, social, and governance (ESG) factors into the decision-making processes of companies and investors. It involves assessing the impact of a company's operations on the environment and society, as well as evaluating the company's corporate governance practices. Companies and investors adopt ESG practices to identify and manage risks, improve performance, and demonstrate their commitment to sustainability and responsible business practices. ESG adoption is becoming increasingly important as stakeholders demand greater transparency and accountability from companies, and as environmental and social issues gain greater prominence in the public discourse (Ground, 2022).

ESG ratings methodologies can vary greatly between rating agencies, with different approaches used to assess a company's environmental, social, and governance performance. For example, some rating agencies may focus more heavily on a company's impact on the environment, while others may place more emphasis on social factors such as labor practices or human rights.

Additionally, different agencies may use different sources of data or analytical tools, leading to varying assessments of the same company.

Investors are advised to comprehend the various definitions and procedures associated with ESG ratings prior to incorporating them in their investment decisions. This helps them avoid being influenced by misleading information. However, the lack of transparency surrounding data sources, weighting, and methodologies often makes it challenging for investors to assess the reliability of ESG ratings and ensure that companies' ESG performance is accurately evaluated.

A new reality is gradually being created for businesses. It is clear, that any adoption effort should be done with care, to identify, prioritize and publicize the ESG issues that are consistent with the needs of each business. So, companies should act rapidly and in an organized manner to incorporate into strategic management decisions. Sustainable development is now a priority at the global level. Companies should try to turn it into a competitive advantage in the context of their inclusion in the new production model. (Ground, 2022).

In Greece, a growing number of companies are recognizing the importance of sustainability and are taking steps to incorporate ESG considerations into their business operations. Many companies are also working to meet international ESG reporting standards, such as the Global Reporting Initiative (GRI) and the Sustainability Accounting Standards Board (SASB) (Serafeim, 2022).

In addition, there has been an increase in the number of Greek companies that have issued green bonds to finance environmentally friendly projects, something that confirms that they are taking environmental considerations seriously. The Athens Stock Exchange has also introduced the FTSE4Good Index, which includes companies that meet specific ESG criteria (Serafeim, 2022).

As of July 2021, 35 listed companies in Greece, as reported by the Athens Stock Exchange, had adopted ESG complied standards. By May 2022, the number had increased to 49, with the objective of encouraging more companies to follow suit. It is estimated that by 2023-2025, the number of listed companies adopting ESG standards will reach 60. Overall, although there is still much progress to be made, Greek companies increasingly recognize the importance of ESG considerations and are taking steps to incorporate them into their business practices (Naftemporiki, 2022).

The procedures for selecting ESG data and conducting assessments can vary significantly, resulting in divergent evaluations of companies' ESG performance. The lack of global reference standards and agreement on what should be considered important for each sector is the main cause for the complexity surrounding the term of ESG. Those discrepancies between evaluators can be attributed as a problem of defining a widely accepted conceptual framework. This overall means that there are differences in definitions of ESG constructs. On the other hand, methodological differences could be attributed as a problem of comparability (Hirai & Brady, 2021).

Afterward, the ESG regulations and standards are scrutinized, along with an evaluation of market-based models and an overview of academic studies related to these models, typically created by rating agencies.

2.1 Regulations

While the European Union's influence on regulations spans diverse sectors worldwide, it may be deemed excessive as it establishes stringent guidelines in particular domains. The recently introduced EU regulations will necessitate an unprecedented degree of ESG reporting and encompass a wide range of companies that were not previously obligated to disclose non-financial information. This includes both public and private entities, regardless of their origin, which meet specific EU-presence criteria (Schulte, 2021). The ESG concept was formally proposed in 2004 even though there were references about it even earlier (Parrino, 2019). Thus, it would be beneficial to provide a concise overview of the origins and key aspects of ESG In the current section.

Over the past five years, the number of ESG regulations and standards has grown considerably, almost doubling on a global scale. There are currently more than 600 ESG reporting provisions, each with varying interpretations of sustainability. With the increasing mainstream adoption of ESG, several frameworks have been developed to address it. The "Group of Five" standard-setting organizations lead various reporting frameworks, which include the Carbon Disclosure Project (CDP), Climate Disclosure Standards Board (CDSB), Global Reporting Initiative (GRI), International Integrated Reporting Council (IIRC), and Sustainability Accounting Standards Board (SASB) (EY, 2021). Out of many available frameworks, the GRI, CDP, SASB, TCFD, and WDI are currently the most widely used, as reported by GreenBiz (Atkins, 2020).

The ESGframework is helpful because it refers to the values of sustainability reporting and it emphasizes the importance of governance. This leads an organization to integrate sustainability non-financial reports with leadership priorities and administration. Actually, it helps the better understanding of the sharedresponsibility that corporations have in global governance, especially in human rights, anti-corruption and international development (Davis, 2013) (Ruggie, 2013) (Wettstein, 2009) (Wolf et al., 2010).

Established in 2000, the Carbon Disclosure Project (CDP) is a non-profit organization that manages a global disclosure system for investors, companies, cities, states, and regions to monitor their environmental impacts. The main objective of the CDP is to combat climate change by transforming capital markets through prioritizing environmental reporting and risk management. The CDP's framework has resulted in unprecedented engagement on environmental issues worldwide over the past 20 years (Atkins, 2020). CDP established its environmental disclosure program in 2002. It is true that since then it has grown significantly with records of over 9.600 companies in 800 cities and 120 states and regions in 2021 (Niemoller, 2021).

The initiative of Principles for Responsible Investment (PRI) is developed by a group of the world’s largest institutional investors. In early 2005, the then UN Secretary-General, Kofi Annan, invited them to join a process to develop the Principles for Responsible Investment. PRI is also supported by the United Nations. The six principles of PRI as presented in Table 1, require the promotion, incorporation, and disclosure of ESG performance. If investors are interested in fulfilling their obligation these issues must be considered as the PRI reflects the view that ESG issues can influence the performance of investment portfolios (Chesebrough, 2022).

Table 1: The six PRI Principles

PRI Principles	Description
Principle 1	Incorporation of ESG issues into investment analysis and decision-making processes
Principle 2	Active owners and incorporation of ESG issues into ownership policies and practices
Principle 3	Seeking of appropriate disclosure on ESG issues by investments
Principle 4	Promotion of acceptance and implementation of the principles within the investment industry
Principle 5	Effectiveness enhancement in implementing the principles
Principle 6	Reporting activities and progress towards implementing the principles

Source: PRI, n.d.

The PRI is a voluntary framework, launched in 2006 by UNEP Finance Initiative and the UN Global Compact. Based on PRI, all investors can incorporate ESG considerations into their decision-making and investment practices. This is an incentive for change in corporate operations and investment practices. As ESG-focused regulations become more commonplace in a business environment, companies that promote social value creation will be in a better position. In 2018, the number of signatories has increased to over 2.300. As of June 2019, there were 2.450 signatories. In even more recent investor surveys, ESG integration has been highlighted as a key motivator for seeking to maximize financial returns and manage risk, so this number keeps increasing (Atkins, 2020).

Originally headquartered in Boston, Massachusetts, the Global Reporting Initiative (GRI) was established in 1997 by the United Nations Environment Program (UNEP) and CERES, a non-profit organization based in the United States. The GRI began to emphasize ESG issues in 2009, shortly after the concept of ESG began to gain traction. GRI's primary objective is to encourage companies to adopt responsible environmental business practices by creating an accountability framework that utilizes Key Performance Indicators (KPIs) to demonstrate their performance. GRI's KPIs (i.e., Greenhouse gas emissions, Water usage, Energy usage, Waste generation and management, Biodiversity etc.) are designed to provide a comprehensive and standardized approach to measuring and reporting on a company's ESG performance. By implementing GRI's KPIs, companies can demonstrate their commitment to sustainability and provide stakeholders with transparent and reliable information about their ESG practices. (Atkins, 2020). Many investors, businesses, and governments use GRI’s ESG framework today so it’s more relevant than ever before with almost four-in-five largest global companies to report with GRI (GRI, 2022).

It is important to note that there is improvement, but people still need to get to know further about ESG. Thus, the GRI Standards were updated in 2016, replacing the previous GRI G4 Guidelines with GRI Standards Version 6, and are updated regularly to ensure they remain relevant and effective. According to the GRI Perspective and its latest updates, sustainability can be divided into two main directions. Based on that, there are organizations that publish standards and others that issue frameworks or guiding principles. Standards are the agreed-upon level of quality requirements, so the reported entities should meet those requirements. On the other hand, there are Frameworks that provide the 'framework' for contextualizing information and are usually put into practice. Both derive authority. Some of them are being endorsed by most stakeholders, and other are required by the law (GRI, 2022).

The Climate Disclosure Standards Board (CDSB) was a global partnership between businesses and environmental NGOs with a shared mission of aligning corporate reporting practices to recognize the value of natural and social capital on par with financial capital. In 2009, the CDSB released the initial version of the Climate Change Reporting Framework (CCRF) during the World Business Summit on Climate Change, which was subsequently revised to reflect evolving reporting requirements and policy changes. Initially, that was focused on the risks and the opportunities that climate change presents to an organisation's strategy, and its financial performance. The following years and after two public consultations, the updated CDSB Framework was released in 2015. The updates refer to reporting environmental and climate change information. The CDSB Framework was updated in spring of 2018, in order to mainly align with the recommendations of the Task Force on Climate-related Financial Disclosures (TCFD) and other key mainstream reporting requirements. Following market changes, user demands and according to the demanding environment requirements the CDSB Framework expanded its focus on social information. So, it was released the last update in 2022 that includes reporting for environmental but also social information (CDSB, 2022). As of January 2022, endorsement of the initiative has been demonstrated by more than 374 companies across 10 sectors and 32 countries. (CDSB Framework, 2022).

The Sustainability Accounting Standards Board (SASB) started working on developing standards in 2011. The main topic was to surface both sustainability and financial fundamentals. The main goal is to identify the subset of sustainability issues most relevant to financial performance in each of 77 industries. There is an important aspect of SASB that sets it apart from the other frameworks. SASB Standards refer to metrics by industry, and they are in use from entities, such as WHO, CDP, EPA, OSHA, and other industry organizations such as ICAO, IPIECA, EPRI and GRESB. The records refer more than 200 entities (SASB Standards, n.d.).

The Taskforce on Climate-related Financial Disclosures (TCFD) started working on further investigation and on how to improve reporting of climate-related financial information in the global financial system, at the end of 2015 (Atkins, 2020). Recommendations were structured around four thematic areas that represent core elements of how companies operate. Those include governance, strategy, risk management, metrics, and targets. The TCFD update of 2017 is a way for companies to report their climate-related financial risks and opportunities through their existing reporting processes. As of November 2022, over 4000 organizations from more than 100 countries have demonstrated their support (TCFD, n.d.).

Sustainable Development Goals (SDG) and ESG refers to quantifiable and measurable factors that represent sustainable practices. The SDG is emerging in 2015. There were four key pillars proposed, that were based on and coincide with many elements of the SDGs (Boffo & Patalano, 2020). From the principals of government, planet, people, and prosperity, it is clear that SDG is focused metrics around societal impact. SDGs are global goals set out by the United Nations being recognised as a framework for responsible investment in addition to ESG which is a rating system used by companies to build their environmental and social credentials (Safety4Sea, 2022).

The Workforce Disclosure initiative (WDI) was another framework created in 2016 by NGO ShareAction and supported by an international investor coalition. Before WDI framework, the CDP had collected data on how companies manage workers across their operations and supply chains. One of the main goals of this action was a try to provide institution investors with meaningful information. The WDI framework was supervised by the non-profit ShareAction in the UK and its main topic was “responsible investment”. As of 2019, the WDI had 137 signatories and 118 companies using the framework. In 2021, 173 global companies demonstrating their commitment to transparency (Atkins, 2020).

Until recently, it was difficult for ESG to penetrate the mainstream of investment strategies and grow, even though many people were in favour of it. In part, this was due to investor misconceptions that sustainable investing offered limited choices and compromised important financial objectives. Today, it is clear that due to the rapid growth, investment strategies and diversification of ESG funds, the industry is in a transitional phase. (Boffo & Patalano, 2020)

At the same time, the definition of ESG terms and practices is not yet clear. There appear to be different core approaches to ESG investing, and meanings may differ by stakeholder group, especially across borders.

The Organisation for Economic Co-operation and Development (OECD) has been involved in developing guidance on responsible investing in a number of ways, but most notable is a consultation for supervisory guidance on integrating ESG factors into investment and risk management. The Organisation for Economic Co-operation and Development (OECD) has also developed ESG frameworks and standards. In 2018, the OECD published its Corporate Governance Principles, which included recommendations on ESG issues such as climate change, resource efficiency, and social responsibility. The OECD has also developed guidelines for multinational enterprises, which provide recommendations on responsible business conduct including ESG considerations. In addition, the OECD regularly publishes reports on ESG issues such as sustainable finance, green growth, and sustainable development. (OECD, 2021).

According to OECD index valuation approaches begin by considering relevant criteria within each of the E, S, and G factors to further articulate the driving forces. Environmental factors may include natural resource use, carbon emissions, energy efficiency, pollution or waste, and sustainability initiatives. Social factors include issues related to the workforce (health, diversity, training) and broader societal issues such as privacy, human rights, and community engagement. A poor environmental record can leave a company vulnerable to legal action or regulatory sanctions; poor treatment of workers can lead to high absenteeism, lower productivity, and weak customer relationships; and weak governance can incentivize unethical behavior related to pay, accounting and disclosure irregularities, and fraud (OECD, 2021).

Table 2 displaying the ESG factors according to ESG Rating providers, OECD:

Table 2: ESG Factors

Environmental Factors	Social Factors	Governance Factors
Environmental Opportunities	Diversity	Corporate Ethics
Pollution and Waste	Supply Chain	Management Compensation
Energy efficiency	Human Rights	Shareholder Rights
Carbon Emissions	Workforce	Board Independence
Natural Resource Use		Board Diversity

Recently, this information is being presented in a more standardised way extends the scope of mandatory ESG reporting to all large companies and SMEs listed on regulated markets. The Corporate Sustainability Reporting Directive (CSRD) requires companies to report more detailed on the impact of corporate activities. The first set of standards is about to be adopted by the Commission by June 2023. Based on common criteria in line with EU climate targets companies should report impact on the environment, human rights, and social standards. The CSRD is said that will amend the existing Non-Financial Reporting Directive (NFRD) (Safety4Sea, 2022).

According to the EU Sustainable Finance Disclosure Regulation (SFDR), a disclosure for ESG risks, policies and results should take place for all asset managers. Its purpose is to share all information about the impact of investments (J.P. Morgan, n.d.). On the other hand, and in terms of sustainability it makes easier the comparison of financial products by providing more transparency (KPMG, n.d.).

Furthermore, The International Sustainability Standards Board (ISSB) is a proposed standard-setting board that would be responsible for developing and issuing global sustainability standards for companies and organizations (IFRS Foundation, n.d.)

The ISSB is being developed by the International Financial Reporting Standards (IFRS) Foundation, which is currently seeking input and feedback on the proposal. The ISSB would work to establish a comprehensive set of sustainability reporting standards that would be consistent, transparent, and comparable across all industries and regions (IFRS Foundation, 2020)

The ISSB proposal aims to develop a framework that encompasses various environmental, social, and governance (ESG) aspects, including climate change, biodiversity, human rights, labor practices, and corruption. This framework intends to assist companies in identifying and managing sustainability risks and opportunities, while providing stakeholders, including investors, with reliable and consistent information on a company's ESG performance (Cohn, 2021).

In collaboration with other sustainability reporting-focused organizations and initiatives like the Global Reporting Initiative (GRI) and the Sustainability Accounting Standards Board (SASB), the ISSB seeks to establish a globally unified sustainability reporting standard. The primary objective is to help companies operate sustainably and responsibly while providing investors with relevant information for making informed decisions (Eccles, 2021).

There is room for improvement in regulatory cooperation among members of institutions in different countries, as formal agreements can be effective in achieving goals. This is a macro trend that is gaining momentum. The data shows that companies embracing ESG criteria are performing better and safer for all stakeholders, investors, employees, customers, and the community. The frameworks and answers are still evolving. Although the final global ESG measurement systems are not complete (Tavanti, 2015).

Table 3 is displaying just a few of the many ESG frameworks and their key information:

Table 3: ESG Global Frameworks by Year of Release

Name of Framework	Organization	Release Year	Primary Purpose	Topics of Disclosure	Category of Reporting	Industry/ Sector	Intended Audience
UN Principles for Responsible Investment	United Nations	2006	To encourage the integration of ESG factors into investment decision-making and ownership practices	ESG integration, active ownership, investment and sustainability themes, industry specific issues	Standards	Investment, Financial, and Insurance sectors	Institutional investors
Carbon Disclosure Project	CDP	2007	To promote sustainable economies by providing data for companies, cities, and states to measure and manage their environmental impact	Climate change, water security, forest and commodity risks, and sustainable supply chains	Voluntary standards	All industries	Corporations, cities, states, and regions
Global Reporting Initiative	GRI	2008	To provide a standardized framework for sustainability reporting	Economic, environmental, social, governance performance and impacts	Standards	All industries	Organizations of all types, sizes, and sectors
Sustainability Accounting Standards Board	SASB	2011	To develop and disseminate sustainability accounting standards that help public corporations disclose financially material ESG information.	Industry-specific standards for financially material ESG topics such as water management, product lifecycle management, and labor practices.	Standards	All industries	Corporations and investors
Taskforce on Climate-related Financial Disclosures	TCFD	2015	To improve and increase reporting of climate-related financial information	Governance, strategy, risk management, metrics, and targets related to climate change	Voluntary standards	All industries	Companies, financial institutions, investors, and regulators
Sustainable Development Goals	United Nations	2015	To provide a blueprint for a more sustainable and equitable world	17 goals, including no poverty, zero hunger, good health and well-being, quality education, gender equality, clean water and sanitation, affordable and clean energy, decent work and economic growth, industry, innovation, and infrastructure, reduced	Voluntary standards	All industries	All stakeholders committed to sustainability

				inequalities, sustainable cities and communities, responsible consumption and production, climate action, life below water, life on land, peace, justice and strong institutions, partnerships for the goals			
Workforce Disclosure Initiative	ShareAction	2016	To improve corporate transparency and accountability in the areas of workforce diversity, supply chain labor standards, and employment rights	Workforce diversity, supply chain labor standards, and employment rights	Voluntary standards	All industries	Investors
OECD Guidelines	Organization for Economic Co-operation and Development	2016	To provide guidelines for responsible business conduct Environment, human rights, labor, anti-corruption, and consumer protection	Human rights, labor, environment, bribery and corruption, consumer interests, disclosure and transparency, competition, and taxation	Voluntary Standards	All industries	Multinational companies operating globally, and governments of countries where these companies are headquartered
Corporate Sustainability Reporting Directive	European Union	2018	To establish a uniform reporting framework for corporate sustainability	Environmental, social, and governance topics related to business strategy, policies, and outcomes	Standards	All industries	Companies operating within the European Union
Sustainable Finance Disclosure Regulation	European Union	2019	To ensure transparency and consistency in disclosing how sustainability is factored into investment decisions	Information on the integration of sustainability risks, the consideration of adverse sustainability impacts, and the promotion of environmental and social characteristics or objectives in financial market participants and products	Regulation	Financial and Investment	Financial market participants and products, such as fund managers, financial advisers, and institutional investors, operating within the European Union, as well as end-investors and beneficiaries that

							invest in those products.
International Sustainability Standards Board	International Federation of Accountants (IFAC)	Not yet released. Expected to be released in Q1 2022	To develop a set of international sustainability standards for corporate reporting, based on the Sustainable Development Goals (SDGs), that would enable organizations to consistently disclose material sustainability information to stakeholders.	Disclosures relating to climate environmental, social, and governance matters	Standards	All industries (The ISSB aims to create a globally accepted set of standards that can be used by organizations of all sizes and types, across all sectors and jurisdictions)	Companies, investors, lenders, regulators, and other stakeholders interested in sustainability reporting.

2.2 Market-based models

The recognition of ESG's significance in investment decision-making has created both a need and a market opportunity. As new players continue to emerge to meet this need, larger financial services firms are consolidating to become dominant players in the field. Although nearly 600 ESG ratings and rankings were identified, it should be noted that the number may not accurately reflect the current landscape due to mergers and acquisitions within the industry, according to ERM as of 2020 (Wong & Petroy, 2020).

Several notable examples of brands include the following acquisitions have taken place in the ESG industry in recent years (Hirai & Brady, 2021), including MSCI's acquisition of GMI Ratings in 2014 (MSCI, 2014) and Carbon Delta in 2019 (MSCI, 2019), S&P's acquisition of Trucost in 2016(S&P Global, 2016) and RobecoSAM's ESG rating business in 2019 (S&P Global, 2019), ISS's acquisition of Oekom Research in 2018 (ISSgovernance, 2018), Moody's acquisition of Vigeo Eiris (V.E) in 2019 (Moody's, 2019), Morningstar's acquisition of Sustainalytics in 2020 (Morningstar, 2020), and Deutsche Börse's acquisition of ISS in 2020 (Deutsche Börse, 2020).

The main international ESG rating systems with high recognition include FTSE Russell, Refinitiv, Moody's, Thomson Reuters, S&P Dow Jones, MSCI and Morningstar Sustainalytics. The following section offers a concise overview of each of the rating systems.

2.2.1 S&P Dow Jones

S&P Dow Jones has a well-established presence in the investment field, offering financial market indices like the S&P 500 and the Dow Jones Industrial Average. Following the introduction of the globally recognized Dow Jones Sustainability World Index in 1999, S&P Dow Jones continued to innovate by developing the ESG index and has been regarded as a trailblazer in this field for two decades. (Sariannidis et al., 2017). As of 2018, recent data indicates that S&P Dow Jones ESG scores encompass over 11,000 companies. S&P Dow Jones acquires ESG information and data for providing ratings from companies that participate in the Corporate Sustainability Assessment (CSA) since 2013. Figure 1 showcasing the aspect levels. S&P Global acquired RobecoSAM's ESG ratings business in 2019, and it's use information and data from the RobecoSAM/SAM ESG ratings business. (S&P Global, 2018).

Essentially the S&P Dow Jones ESG scores use input from the ESG scoring algorithm developed by SAM, but also improves upon it in several ways. It is essentially used to determine the constituents of the Dow Jones Sustainability Index, since it aggregates the ESG data of the relevant companies into a score through SAM's annual CSA (Sariannidis et al., 2017).

Standard scores and their corresponding weights are what ultimately weigh a company's overall ESG score. The individual score for each E, S and G element is also considered as the weighted average of all standard scores and their respective weights for each ESG element. Therefore, from the above analysis, it follows that the score referred to a given financial year is composed of the company's overall ESG score and the individual scores for E, S and G elements. The final ESG score in this method ranges from 0 to 100 (S&P Global, 2018).

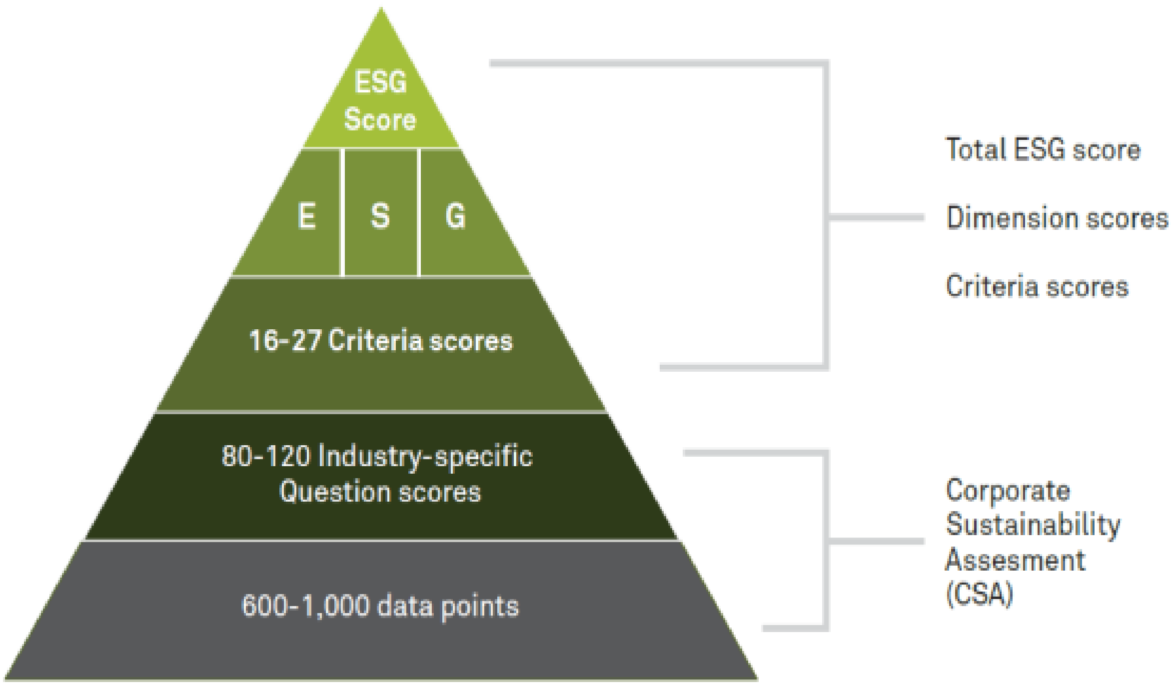

Figure 1: S&P DJI ESG Score Aspect Levels

Source: S&P Global, 2018

2.2.2 MSCI

MSCI also has a long-standing presence with more than 40 years of research experience in the field of ESG and can be described as commonly utilized (Afreen, 2021). Notably, MSCI is widely recognized as a trailblazer for incorporating a company's risk exposure into the economic benefit evaluation of the industry. MSCI currently rates more than 8500 companies worldwide, according to the latest data published, collecting thousands of data features for each company (MSCI, n.d.).

Figure 2 shows that the MSCI ESG rating system is comprised of 3 categories, 10 themes, and 35 fundamental indicators. These data points are used in combination to measure a company's ESG risk exposure and the ESG risk management practices they have adopted. The weighting ratios used to determine the impact of ratios on the industry and the time duration are two key aspects. The company's final score is weighted by rating indices and adjusted according to the industry. Companies are then assigned ESG scores that range from top AAA, AA to lag B, CCC, with necessary adjustments made as needed. (MSCI, 2020).

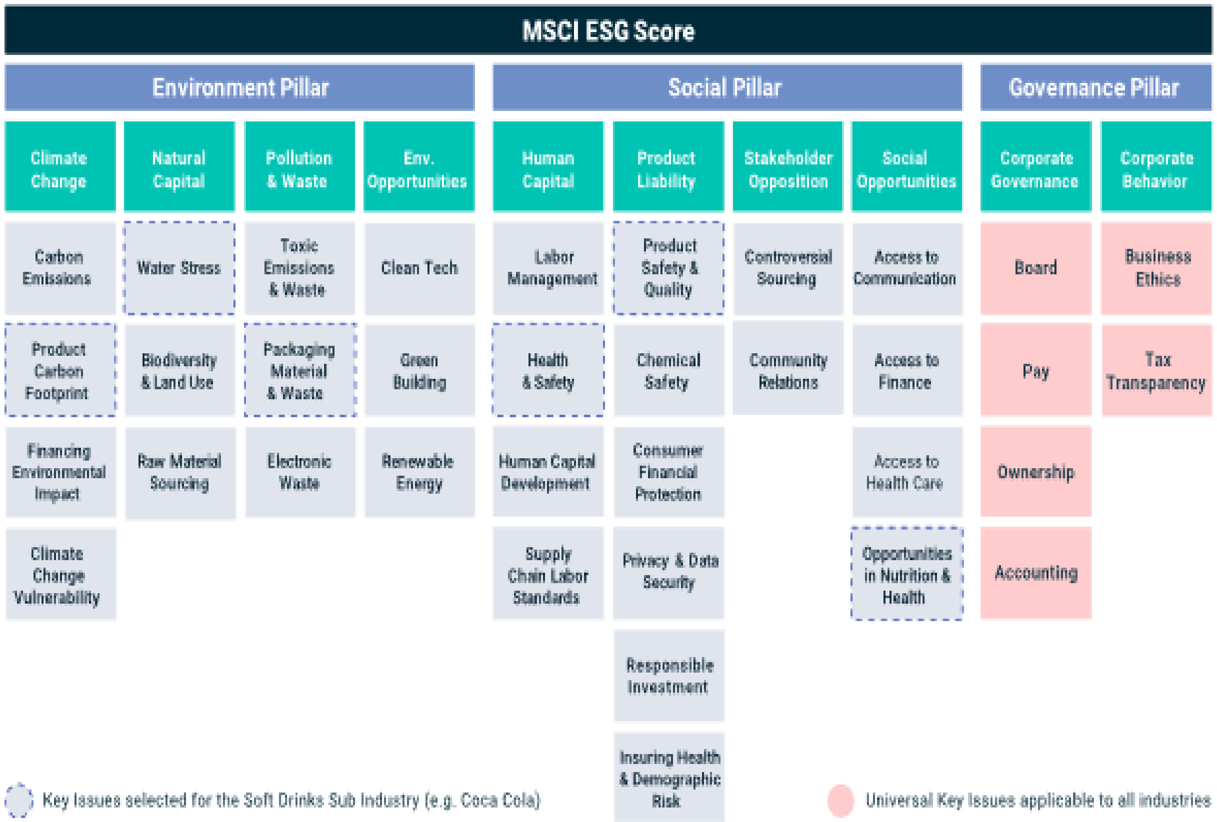

Figure 2: ESG Ratings Key Issue Framework

Source: MSCI, n.d.

2.2.3 Morningstar Sustainalytics

Morningstar's acquisition of Sustainalytics further strengthened its market position as a leading global provider of research and ratings. Its presence in the ESG field spans 25 years, during which it has provided ESG information to investors globally. Sustainalytics currently covers more than 14.000 companies worldwide, across 42 industries.

Sustainalytics measures companies' ESG performance in terms of risk and is essentially based on three main elements: corporate governance, key ESG issues and special issues (blank swans). Sustainalytics' ESG risk assessment considers two main dimensions combined into a single score. Those are risk exposure and risk management as shown also in Figure 3. The Sustainalytics rating system ranks ESG risk levels from Negligible to Severe, with scores ranging from 0 to 40+, with the lower the score, the lower the risk (Hale, 2019).

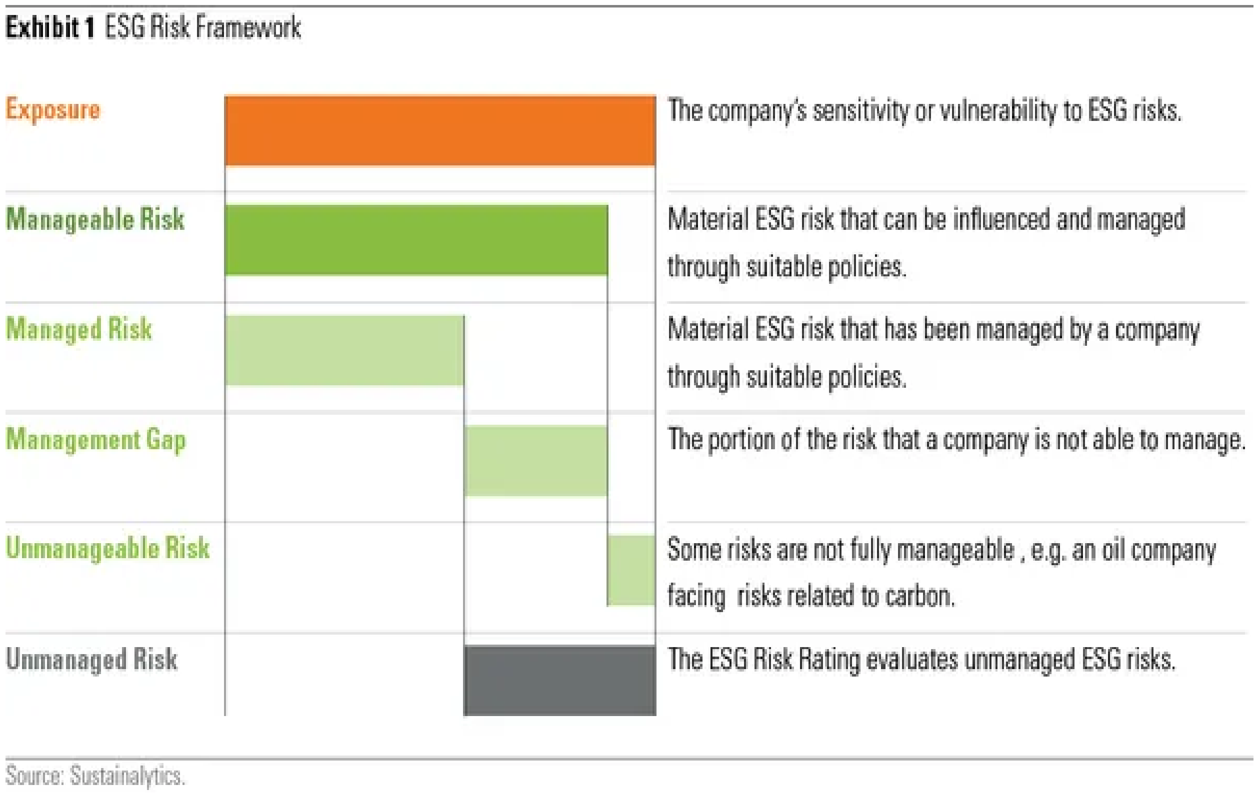

Figure 3: Morningstar Sustainalytics ESG Framework

Source: Hale, 2019

2.2.4 FTSE Russell

Another leading global indexing company is FTSE Russell. It has experience in ESG ratings for 20 years now and data models cover 7.200 securities in 47 developed and emerging markets (Gajo, 2020). It is worth noting that its part of the Information Services division of the London Stock Exchange Group (LSEG).

As presented in Figure 4, the FTSE Russell ESG rating system is divided into three main tiers, and it covers ESG issues in multiple dimensions referenced in each company's evaluation. As expected, the first tier refers to the three pillars of E, S and G. Under the second tier the 14 themes are used to measure the overall quality of companies' management of ESG issues, reflecting all 17 Sustainable Development Goals (SDGs) of the UN. The third tier consists of over 300 separate assessment indicators (Ratsimiveh & Haalebos, 2021).

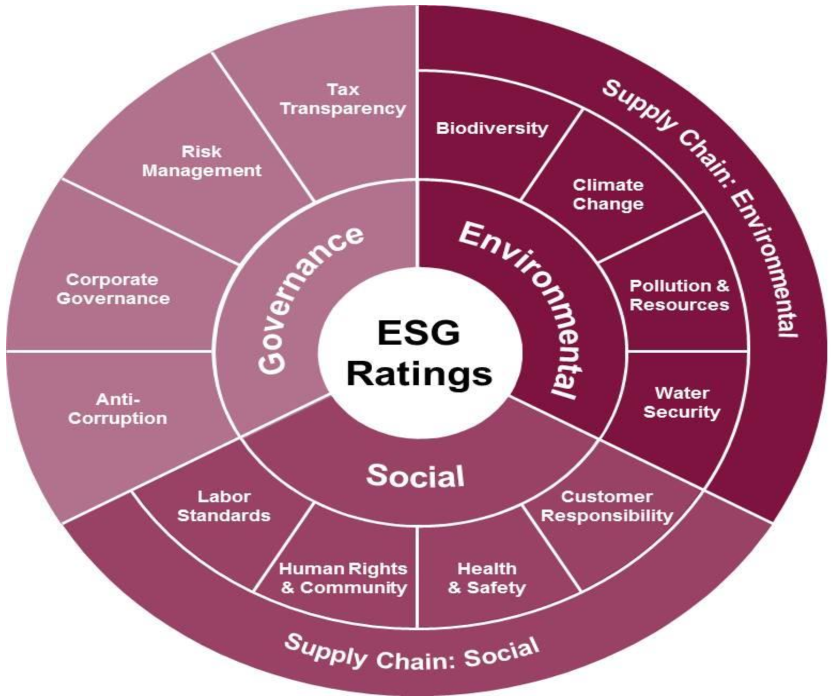

Figure 4: FTSE Russell Pillar and Theme Exposures

Source: FTSE Russell, n.d.

During the research cycle of 2018-2019, FTSE Russell's ESG methodology was primarily focused on the Water Security theme (FTSE Russell, 2018). Unlike other rating agencies, FTSE Russell contacts each rated company individually to ensure all relevant public information is provided. Its ESG rating system relies on publicly available information, including CSR reports, mandatory accounting disclosures, regulatory filings, stock exchanges, quarterly company reports, media etc.

2.2.5 Refinitiv (Thomson Reuters)

Refinitiv is a globally recognized financial market infrastructure data provider that has been operating for a significant period. Based on the latest information available on the website, the platform serves over 40.000 customers and 400.000 end users in more than 190 countries (Doronin et al., 2021).

Refinitiv's ESG division was previously known as the Finance and Risk Division of Thomson Reuters before it was acquired by Refinitiv. The combined ESG score is primarily based on a company’s relative ranking in its industry as it is shown in Figure 5. At the first level, the ESG score is structured into 10 themes, and it is further combined to generate a weighted average ESG score. Simultaneously, the second level pertains to the ESG controversies score, which is calculated by adjusting for market capitalization bias, depending on the number of conflicts that have emerged (Refinitiv, n.d.).

Refinitiv uses ESG controversies to assess a company's ESG performance. These controversies are typically events or incidents that reflect poorly on a company's ESG practices, such as environmental disasters, labor disputes, or corruption allegations. Refinitiv collects data on ESG controversies from a range of sources, including news articles, regulatory filings, and non-governmental organizations (NGOs). The number and severity of these controversies can impact a company's ESG rating (Refinitiv, n.d.).

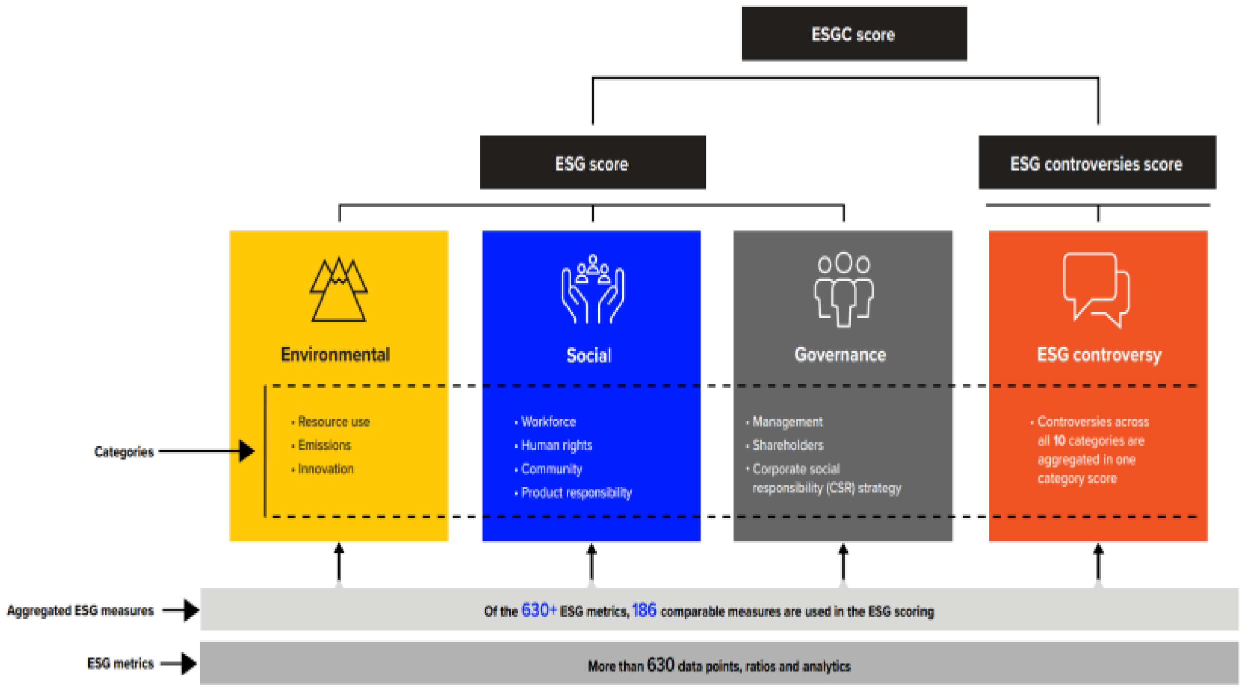

Figure 5: Refinitiv ESG rating indicators

Source: Refinitiv, n.d.

Refinitiv has been producing ESG scores for over 12.500 companies globally, which represents 80% of the global capitalization, since 2002. The Refinitiv rating system includes over 630 ESG rating metrics and 186 comparable indices for index allocation (Refinitiv, n.d.).

2.2.6 V.E. Moody's

Vigeo Eiris, which is part of Moody's ESG Solutions, is a global leader in research, data, benchmarks, and analytics with years of experience as a rating provider. Moody's provides ESG and climate information for over 300 million public and private companies globally, as shown in Figure 6. Moody's ESG assessments are based on double materiality and measure the extent to which companies manage ESG factors by leveraging historical data from 2004 to the present. This means that they evaluate the financial materiality of ESG factors for the company itself (internal factors) as well as the impact of the company's operations on the external environment and society (external factors). This approach is based on the idea that a company's financial performance is not only influenced by its internal ESG management but also by its impact on the external world. This is achieved by combining verified data, using analyst-led assessments, and creating predicted scores using a dataset containing over 100.000 firms to predict metrics for over 600 industries in 220 countries. Moody's ESG assessments consider external environmental and social risks to financial performance as well as the impact on the components of a business environment and societies in which a business operates (Moody’s, n.d.).

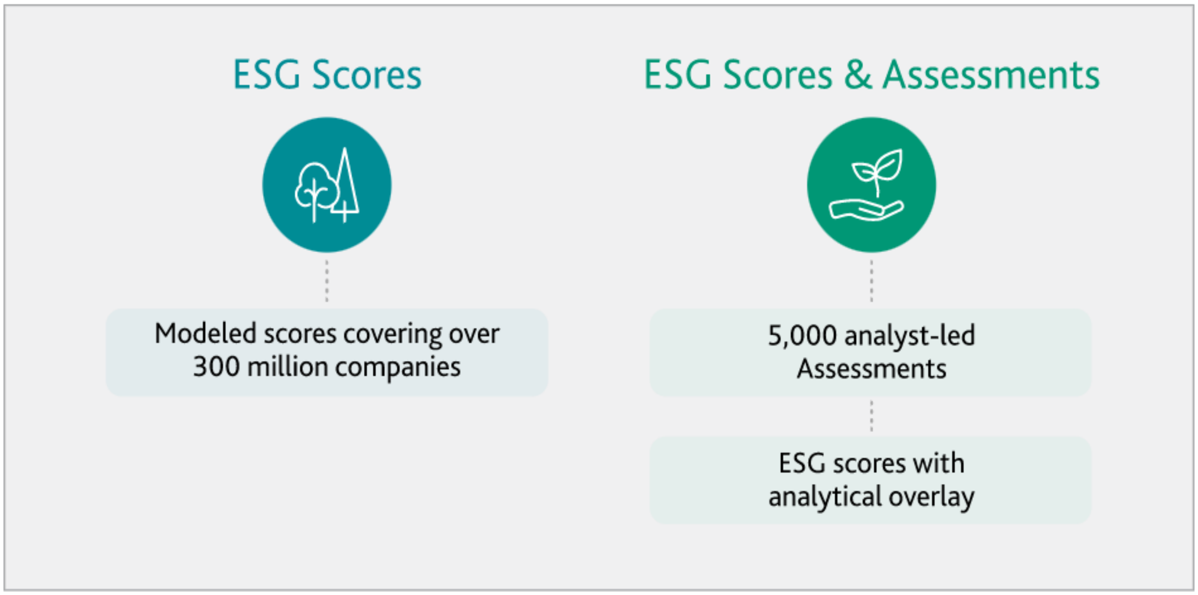

Figure 6: Moody’s ESG Supply Chain Coverage

Source: Moody’s, n.d.

Table 4 summarizes the key information on various ESG rating methodologies and parameters, based on the data and information analysed.

Table 4: Main ESG Data and Rating Providers

Research Process	S&P Dow Jones	MSCI	Refinitiv Thomson Reuters	FTSE Russell	Morningstar Sustainalytics	V.E. Moody's	ISS ESG
ESG Measure	Sustainability based on ESG risks and opportunities	Resilience to long term ESG risks, anticipated costs and/or opportunities	Relative ESG performance, commitment, and effectiveness	ESG risks based on publicly available information	Unmanaged ESG risks in a portfolio relative to a portfolio's peer group	Degrees of ESG factors management	
Model Logic	Quantitative	Quantitative & Rating Committee	Quantitative	Quantitative	Quantitative	Quantitative	Quantitative
Coverage	7000+	14000+	9000+	7000+	12000+	5000+	6000+
Score Upper	100	AAA	100	5	100	100	A+
Score Lower	0	CCC	0	0	0	0	D-
Rating Cycle	Annual*	Annual*	Weekly	Annual	Annual*	Annual*	Annual*
Rating Made Public	✓	✓	✓	×	✓	×	×
Indices Supplied	S&P, DJSI	MSCI, Bloomberg	Refinitiv	FTSE	Solactive, STOXX, S&P	Euronext	Solactive, STOXX
Data Verification by Companies	✓	✓	✓	✓	✓	✓	✓
Second Party Opinion	✓	✓	×	×	✓	✓	✓

* A full review is conducted annually, however some information, such as disputes involving the company, is updated daily and may change the rating at any time during the year.

Source: Hirai et al., 2021

This study reviewed the main methodologies, based on the world's largest rating agencies, it could also extend to other methodologies and metrics of i.e., CDP, RepRisk, Arabesque, InRate, State Street R-Factor etc. The ESG rating system is primarily based on qualitative information that each rating agency could access and thus produced results that varied significantly as they were evaluated using different scoring criteria and index systems. Even when this happened, they usually covered three levels of indicators with the main indicators always being in three dimensions, E, S and G while the rest of the secondary or even third level indicators were different due to the different focus of each agency.

2.3 Academic-based research

Numerous academic studies have examined the efficiency and precision of ESG rating methodologies, and there are several articles on ESG research conducted in various years. These sources highlight that the ratings produced by different organizations demonstrate variability, which results in significant divergence.

ESG rating discrepancies, such as the varying assessments of Tesla's ESG performance by FTSE Russell, MSCI, and Sustainalytics, highlight a lack of standardization and consensus within the ESG ratings industry. For instance, while FTSE Russell ranked Tesla last, MSCI placed it at the top among global automotive companies, and Sustainalytics positioned it somewhere in the middle. These disparities suggest that more uniform ESG rating standards and practices are needed to improve the reliability and comparability of these ratings (Allen, 2018).

The discrepancy in ESG ratings for a company like Tesla, where different methodologies place it at different positions, can be a problem because it can create confusion for investors, stakeholders, and other interested parties who rely on these ratings to make informed decisions (Vargas & Moneva, 2021).

If different ESG ratings methodologies are producing divergent results, then it becomes difficult to determine which assessment is the most accurate or relevant. This can lead to a lack of consensus on the company's ESG performance, which in turn can impact its reputation and financial performance (Rigoni et al., 2021). An empirical analysis of ESG ratings and suggest that discrepancies in ESG ratings can be attributed to differences in the weights assigned to different ESG factors, as well as differences in the data sources used by ESG ratings providers (Boubakri et al., 2021).

Additionally, companies may feel that they are being unfairly evaluated by one methodology versus another, which can create mistrust or scepticism of the ratings process. This can also lead to a lack of standardization in ESG ratings, which can make it difficult to compare companies across sectors or regions. (Preuss & Chapple, 2016)

Overall, the discrepancies in ESG ratings can be a problem because they can lead to confusion, mistrust, and a lack of standardization in the evaluation of companies' ESG performance. As a result, there is a growing need for greater transparency and standardization in the ESG ratings process to ensure that investors and other stakeholders have access to reliable and consistent information (Hoepner, 2017).

There are two primary areas of research interest that have emerged with respect to ESG ratings. The first relates to the limited predictive value of past ESG performance in forecasting future performance. The second pertains to the potential discrepancies and conflicting assessments of the same company by different ESG rating providers (Hoepner et al., 2019) (Krüger & Sautner, 2021).

Another research noted that there are three main types of differences between ESG ratings: differences in methodologies, differences in data, and differences in interpretation. It emphasizes the importance of transparency and communication between companies and ESG rating agencies to ensure that ESG ratings are accurate and reliable indicators of a company's sustainability performance. (Hirai & Brady, 2021).

According to a recent study, there is a lack of correlation between different ESG ratings methodologies. The study used data from Fortune magazine's "100 best companies" and compared it with data based on MSCI/KLD (formerly Kinder, Lydenberg and Domini Research & Analytics), finding that the relationship between rankings was only 14% of the total. This suggests that there is a significant difference in how ESG ratings providers evaluate companies, leading to potential inconsistencies in their assessments. These discrepancies highlight the need for greater standardization and transparency in ESG ratings methodologies. (Hawley, 2017).

Studies have compared ESG assessments in different ways, sometimes focusing on their validity and at other times on their correlation and agreement in outcomes. For example, one survey aimed to compare the environmental ratings of MSCI/KLD, ASSET4, and Global Engagement Services. The study attempted to assess the convergent validity of the ratings, which were based on similar dimensions, but found that they did not fully converge (Semenova & Hassel, 2015).

Accordingly, and after an initiative to examine the degree of convergence of the MSCI/KLD, FTSE4Good, DJSI, ASSET4 (Now Refinitiv), Calvert, and Innovest ratings for ESG factors, the conclusion reached by the research for the ix ratings mentioned above is that their degree of convergence is particularly low (Chatterji et al., 2016).

In another study, a sample of companies evaluated using three widely used ESG rating systems - ASSET4, MSCI/KLD, and ESG data from Bloomberg - between 2002 and 2012 were compared. The study analyzed individual E, S, and G scores, as well as aggregated ESG scores and financial data. The results showed that the correlations between ASSET4 and Bloomberg were quite high for both individual and aggregated ESG scores, while the correlations with MSCI/KLD were very low, indicating inconsistency between the rating systems (Dorfleitner et al., 2015).

It is clear that there are three distinct kinds of problems, theory, comparability and transparency, but they are also shown to be independent of each other. Just to explain the problems identified, the problem of comparability relates to the differences in the sources of information used to collect data, the problem of transparency refers to the reliability of the information published in sustainability reports, and the problem of theory reflects the different views on which ESG factors are considered economically relevant to the development of the ratings (Serafeim, G., 2015).

Finally, a quantitative analysis to review ESG research took place and found that ESG, and the economic performance of firms could be related. For this purpose, they used a tool called

CiteSpace to systematically analyze all previous literature on ESG including the theoretical foundations but also the measurements (Li et al. 2021).

Table 5 presents a summary of diverse academic research that has been analysed based on the available data and information.

Table 5: Academic Papers regarding Existing ESG Methodologies

Title	Authors	Main Objective	Main Results	Methodology
Do ratings of firms converge? Implications for managers, investors, and strategy researchers	Chatterji, A. K., Durand, R., Levine, D. I. and Touboul, S.	To examine the degree of convergence of ESG ratings produced by various organizations	The degree of convergence of ESG ratings is particularly low	Quantitative - Panel data analysis with fixed effects model and standard errors corrected for clustering.
Lies, damned lies and ESG rating methodologies	Allen, K.	To highlight the limitations and flaws in ESG rating methodologies	ESG ratings are subjective and can vary significantly between different agencies	Qualitative - Critical analysis of ESG rating methodologies and their limitations
The challenges of ESG ratings	Allen, F.	To analyse the challenges associated with ESG ratings	ESG ratings are often based on limited and inconsistent data	Qualitative - Discussion of the challenges and limitations of ESG ratings.
Disagreement in ESG ratings: A systematic literature review	Vargas, J. L., & Moneva, J. M.	To examine the reasons behind the disagreement among ESG ratings agencies	The disagreement among ESG ratings agencies is attributed to differences in their methodologies and data sources	Qualitative - Systematic literature review of academic studies on disagreement in ESG ratings.
Evaluating ESG ratings: An analysis of methodologies and sustainability reporting trends	Rigoni, D., Sartori, L., & Zoni, L.	To evaluate the methodologies and trends in sustainability reporting among ESG ratings agencies	There is a lack of standardization in ESG rating methodologies, and sustainability reporting trends vary between different industries	Qualitative - Quantitative Content analysis of sustainability reports and critical analysis of ESG rating methodologies
ESG ratings: An empirical analysis	Boubakri, N., Cosset, J.-C., & Saffar, W.	To examine the determinants of ESG ratings	The determinants of ESG ratings are related to firm characteristics, country factors, and industry factors	Quantitative - Multiple regression analysis, classification and regression tree analysis, and sensitivity analysis
Assessing the reliability and relevance of ESG rating systems	Preuss, L., & Chapple, W.	To assess the reliability and relevance of ESG rating systems	The reliability and relevance of ESG rating systems are limited by a lack of transparency, standardization, and consistency	Qualitative - Systematic literature review, content analysis, and qualitative data analysis using the method of constant comparison

Challenges and opportunities in ESG rating and sustainability reporting	Hoepner, A. G. F.	To examine the challenges and opportunities associated with ESG rating and sustainability reporting	ESG rating agencies should focus on improving transparency, consistency, and standardization in their methodologies	Qualitative - Literature review and critical analysis of ESG ratings and sustainability reporting.
ESG Shareholder Engagement and Downside Risk	Hoepner, A. G. F., Rezec, M., & Hebb, T.	To investigate the relationship between ESG shareholder engagement and downside risk	ESG shareholder engagement can reduce downside risk in firms	Quantitative - Panel data analysis and regression model to investigate the relationship between ESG shareholder engagement and downside risk.
Disagreement in Socially Responsible Investing	Krüger, P., & Sautner, Z.	To analyze the reasons behind the disagreement among socially responsible investing (SRI) funds	The disagreement among SRI funds is attributed to differences in their investment strategies and ESG rating methodologies	Quantitative -Panel regression analysis and instrumental variable approach
ESG Ratings and Rankings: All over the Map. What Does it Mean?	Hawley, J.	To discuss the implications of the variability in ESG ratings and rankings	The variability in ESG ratings and rankings raises questions about their usefulness and reliability	Qualitative - Critical analysis of ESG ratings and rankings.
Managing ESG Data and Rating Risk	Hirai A., Brady A.	To analyze the challenges and risks associated with managing ESG data and ratings	Companies should adopt a proactive approach to managing ESG data and ratings to mitigate risks	Qualitative - Interviews with industry experts
On the Validity of Environmental Performance Metrics	Semenova, N. and Hassel, L. G.	To investigate the validity of environmental performance metrics used by firms	The authors found that the current environmental performance metrics used by firms have limitations, and suggest alternative methods to address them	Qualitative - Systematic literature review
What's driving the sustainability movement?	Serafeim, G.	To identify the drivers of the sustainability movement	The author identifies the four main drivers of the sustainability movement: societal factors, investor pressure, regulatory pressure, and company leadership	Qualitative - Case study analysis and interviews with executives
ESG: Research progress and future prospects	Li T., Wang K., Sueyoshi T., Wang D.	To review the research progress and provide future prospects of ESG	The authors provide an overview of the progress made in ESG research and suggest future directions for research	Qualitative - Systematic literature review
Measuring the level and risk of	Dorfleitner G., G.	To compare the level and risk of	The authors find significant differences in	Quantitative - Factor analysis, clustering

corporate responsibility – An empirical comparison of different ESG rating approaches	Halbritter, and M. Nguyen	corporate responsibility measured by different ESG rating approaches	the ESG ratings provided by different agencies, indicating the need for standardization	analysis, and regression analysis

In addition to studies that evaluate the effectiveness and accuracy of existing ESG rating methodologies, there are also academic papers that propose new approaches and frameworks for evaluating ESG factors. These papers aim to address some of the limitations and challenges associated with current ESG rating methodologies and provide innovative solutions for measuring and assessing ESG performance.

One study conducted a comprehensive review of literature on ESG modelling and suggested a new integrated model for ESG performance evaluation (Sasikumar & Ramachandran, 2016). Similarly, another study proposed a machine learning model that combined financial and non-financial data to evaluate ESG performance of companies (Kumar & Ganesh, 2020).

In another research, a Delphi method-based expert survey was used to propose an ESG scorecard model for constructing ESG scores based on predefined indicators and weights (Biswas & Veliyath, 2018). Furthermore, a multi-criteria decision model was proposed to evaluate ESG performance based on predefined criteria, which combined financial and non-financial data (Ghosh et al., 2017).

Another study integrated financial and non-financial data to evaluate ESG performance based on predefined indicators and weights for equity investments (Wang et al., 2019). Similarly, ESG risk factors were analysed for their financial impacts using an event study methodology (Flammer, 2013).

Furthermore, a framework was proposed for integrating ESG performance metrics into supply chain management using a fuzzy analytic hierarchy process (FAHP) approach (Wong & Cheng, 2016). Additionally, a methodology for ESG scoring in the context of private equity investments was proposed in another study (Brown & Viehs, 2016). Lastly, one additional study proposed a methodology for modelling ESG scores and improving ESG scoring methodologies using a Bayesian network approach (Xiao & Yang, 2019).

Based on the analysed data and information, Table 6 presents a summary of essential details from various academic-based models.

Table 6: Academic Papers that Propose New Approaches

Title	Authors	Main Objective	Main Results	Methodology
Corporate Environmental, Social and Governance Modelling: A Review	Sasikumar, S.K., & Ramachandran, S.	Provide a comprehensive review of existing literature on ESG modelling and propose a new	Identified gaps in existing ESG models and proposed a new integrated ESG model incorporating	Qualitative - Literature review of existing ESG models

		integrated model for ESG performance evaluation	financial and non-financial data	
A Machine Learning Model for ESG Performance Evaluation of Companies	Kumar, V.S., & Ganesh, K.	Propose a machine learning model for ESG performance evaluation of companies	Developed a machine learning model that uses financial and non-financial data to predict ESG performance scores with high accuracy	Quantitative - Machine learning model based on K-means clustering, PCA, and decision tree analysis to evaluate ESG performance of companies
ESG Scorecard: A Model for Constructing ESG Scores	Biswas, C., & Veliyath, R.	Propose a model for constructing ESG scores based on a set of predefined indicators and weights	Developed an ESG scorecard model based on expert survey and Delphi method, which incorporates a set of predefined indicators and weights	Qualitative - Constructed ESG scorecard using principal component analysis and entropy-based weighting method to determine the weights of ESG criteria
A Multi-Criteria Decision Model for ESG Rating of Companies	Ghosh, P., Jana, R.K., & Mukherjee, K.	Propose a multi-criteria decision model for ESG rating of companies	Developed a model that combines financial and non-financial data to evaluate ESG performance of companies based on predefined criteria	Quantitative - Multi-criteria decision model based on AHP and Fuzzy TOPSIS methods to rate companies on ESG performance
ESG Rating Model for Equity Investment	Wang, L., Zhang, S., & Zhu, J.	Propose an ESG rating model for equity investment	Developed a model that integrates financial and non-financial data to evaluate ESG performance of companies based on predefined indicators and weights	Quantitative - Constructed ESG rating model using principal component analysis, entropy method, and grey relational analysis
ESG Risk Factors and Their Financial Impacts: An Event Study Analysis	Flammer, C.	Analyse the financial impacts of ESG risk factors	Conducted an event study analysis to show that ESG risk factors have a significant impact on financial performance	Quantitative - Empirical analysis of the relationship between ESG risk factors and financial risk using regression analysis
Integrating Environmental, Social, and Governance (ESG) Performance Metrics into	Wong, C.W.Y., & Cheng, T.C.E.	Propose a framework for integrating ESG performance metrics into supply chain management	Developed a framework that uses a fuzzy analytic hierarchy process (FAHP) approach to integrate ESG	Qualitative - Framework to integrate ESG performance metrics into supply chain management,

Supply Chain Management: A Framework and Case Study			performance metrics into supply chain management	illustrated with a case study
ESG Scoring for Private Equity	Brown, N., & Viehs, M.	Propose a methodology for ESG scoring in the context of private equity investments	Developed a methodology for ESG scoring which accounts for the specific characteristics of private equity investments	Qualitative - Literature review of existing ESG models, with a focus on private equity
Modelling ESG Scores and Improving ESG Scoring Methodologies	Xiao, J., & Yang, Y.	Propose a methodology for modelling ESG scores and improving ESG scoring methodologies	Developed a methodology that uses a Bayesian network approach to model ESG scores and identify areas for improvement in existing ESG scoring methodologies	Quantitative - Developed a two-step ESG scoring model, first using text mining and machine learning to extract ESG signals from news articles and company reports, then using principal component analysis to aggregate the signals into ESG scores. Evaluated the effectiveness of the model using correlation analysis, factor analysis, and other statistical tests.

Since there was a review of the literature review and the methodologies used to the development of the ratings, it is better to get an impression of the global regulation systems regarding ESG in the next section.

Chapter 4: Methodology

In recent years, logistic regression has become a popular statistical tool in research, as it allows for the modelling of the relationship between a dependent variable and multiple independent variables (Long & Freese, 2014). This technique is used to investigate the non-linear effect of a dependent categorical variable in relation to several independent variables. There are three types of logistic regression models, depending on the categories of the dependent variable: binomial regression (with only two categories), ordinal, and nominal (qualitative categories). (Petridis, 2015). The goal of logistic regression is to predict the outcome of a binary response variable Y based on probability theory, using a combination of predictor variables that may be nominal, ordinal, or quantitative (Petridis, 2015).

4.1 The Logistic Curve

In the language of statistics, LR is a methodology used to predict the probability of occurrence of a phenomenon by fitting the study data to the equation of the logistic curve.

According to the statistics literature, the binary dependent variable and the predicted value, the probability, must be bounded to fall within the same range. LR uses the logistic curve as shown in Figure 7 to represent the relationship between the independent variable and the dependent variable. Thus, it should define a relationship bounded by 0 and 1. For very high values of the independent variable, the probability approaches 1 but never reaches 1. In contrast, the predicted value on the curve decreases as the independent variable decreases, approaching 0 but never reaching 0 (Hilbe, 2011).

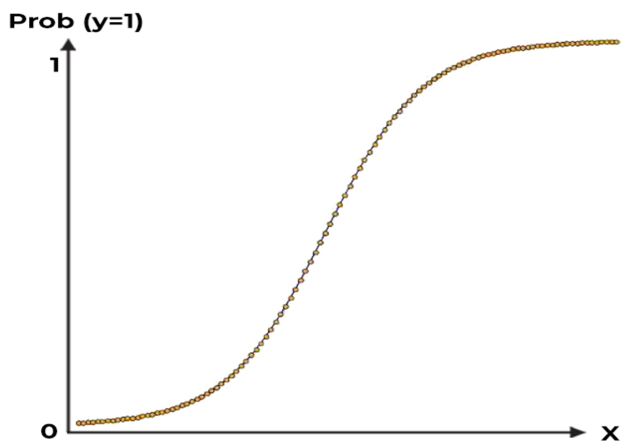

Figure 7: Sigmoid Curve

This curve has a sigmoid shape and is characterized by an exponential growth phase in which the growth rate gradually slows down and ends in the asymptotic saturation phase of growth, where the straight line finally runs parallel to the X axis.

4.2 Probability Transformation into Odds and Logit Values

It is important to ensure that the estimated values do not lie outside the range of 0 and 1. This is achieved with logistic transformation in two steps: firstly, the probability is transformed into odds, which is defined as the ratio between the probability that the event will occur and the probability that it will not. To limit the predicted values to a value between 0 and 1, the odds value can be converted back to a probability,

$$\text{Probability(event)} = \frac{\text{odds (event)}}{1 + \text{odds (event)}}$$

To prevent odds from falling below 0, which is the lower limit, so no upper limit is used in this case, the logit value must be calculated, which is the logarithm of the odds. Odds greater than 1.0 have a positive logit value. On the other hand, odds less than 1 have a negative logit value, and odds of 1.0 have a logit value of 0 (those are corresponding to a probability of 0.5) (Boateng & Abaye, 2019).

4.3 Odds Ratio (OR) Interpretation

Odds Ratio (OR) called when an independent variable X_i increases by one unit as (X_i+1). All other factors remain constant. Then the probability for the dependent variable increases by a factor exp(b_i) (Boateng & Abaye, 2019).

By creating the confusion matrix Table 7 of the two variables, in which the ratio OR estimates a relationship that develops between a cause and a result:

Table 7: Cause and result

		Result	
Cause		+	-
	+	a	b
	-	c	d

The OR ratio results from the relationship:

$$\text{OR} = \frac{\text{a d}}{c\ b}$$

OR ranges from zero (0) to positive infinity (+∞). It indicates the relative change amount by which if the value of the corresponding independent variable increases by one unit, the odds of the dependent variable decreases (OR < 1) or increases (OR > 1).

4.4 The Logistic Regression Model

Binary logistic regression is a binomial equation related to the response variable Y, which is the random result of the occurrence of one of two potential outcomes. These potential outcomes can be success or failure. In this paper, the dependent variable Y takes the value 1 if an event occurs, otherwise it takes the value 0 if no event occurs.

The logistic function has the form:

$$\mathrm{P} = \mathrm{f(z)} = \left[\frac{e^z}{1 + e^z}\right] = \left[\frac{1}{1 + e^{-z}}\right]$$

where z is the input variable and f(z) are the output variable. The f(z) represents the probability of a particular outcome due to the effect of that group, while the variable y represents the effect of a group of independent variables. The variable z also expresses the measure of the total contribution of all involved independent predictor variables x_1, x_2, …., x_k in the model and is defined as:

$$z = b_0 + b_1 x_1 + \cdots + b_k x_k$$

where b_0 is the height of the slope of the regression line and corresponds to the z-value when the values of all independent variables are equal to 0, while bi are the regression coefficients, each of which expresses the magnitude of the contribution of the corresponding variable. A negative value means that the variable decreases the probability of that outcome, while a positive value of the coefficient means that the explanatory variable increases the probability of the successful outcome. A high value of the coefficient means that the independent variable has a very strong influence on the probability that the event occurs or does not occur, while a low value indicates a small influence of the independent variable on the probability of the corresponding outcome. In logistic regression, each predictor is given a coefficient that measures its independent contribution to the variation in the dependent variable (Boateng & Abaye, 2019).

The model form for Predicted Probabilities is expressed as a logarithm (ln):

$$\ln(\mathrm{p}) = \ln\left[\frac{p}{1-p}\right] = b_0 + b_1 x_1 + \cdots + b_k x_k$$

The goal of LR is to estimate finally the k+1 unknown parameters b. This can be done by using Maximum Likelihood Estimate (MLE). The MLS entails finding the set of parameters for which the probability of the observed data is the greatest. The regression coefficients which indicate the degree of association between the outcome and each one independent variable are calculated using the form:

$$L = \prod_{i=1}^{n} f(x_i \theta)$$

Or better expressed as a logarithm (ln):

$$L = \sum_{i=1}^{n} ln\, f(x_i \theta)$$

where θ is a parameter of the variable which can vary freely. The predicted value for each observation will be equal to:

$$\hat{l} = \frac{1}{n} \ln L$$

Likewise, the probability function of the outcome of an event, represents the way an observed sample is described by the parameter values, e.g., mean, standard deviation, etc. By maximizing the outcome, the likelihood function determines those parameters that are most efficient in producing the observed data. The reliability of logistic regression results is usually affected by the sample size of the survey. There is a rule that states that the number of desired results must match the number of independent variables in the ratio of 10:1. From the point of view of statistical weighting, MLE is more suitable for large sample applications because it is flexible, easily adapted to the creation of many different types of models, analyses different elements, and contains more accurate measurements (Boateng & Abaye, 2019).

Valuable information on the properties of binomial models is described to another research (Cox & Snell, 1989) (Collett, 2003).

4.5 Selecting the Dependent Variables and the Potential Predictors

In most cases of selecting the dependent variables in a model, the outcome event is specifically categorized into two categories. These categories are the case where something has happened or else it hasn't. In some other cases, it is possible to reduce them to dichotomous variables if there are many multicategory or even continuous variables. Multiclass classification is a type of machine learning task that is used when there are multiple classes or categories that an observation can belong to. (Boateng & Abaye, 2019).

However, there is another important aspect to consider when developing a LR study. This mainly concerns the choice of variables to be analysed as potential predictors of the outcome. It is expected that the results of LR always depend on the independent variables selected as potential predictors. However, the decision of whether to include factors in the initial data set can affect the results since if a variable is not ultimately selected for analysis, it cannot be included in the final model (Levy & Stolte, 2000). Obviously, there are several disadvantages in the choice of predictor variables, and this may also concern the correlation between them. Particularly in small or medium-sized samples, collinearity can lead to overall significance levels of LR if the individual predictors do not by themselves predict the outcome, or the degree of relationship between a predictor and the outcome is not correctly determined (Kim et al., 2006). As a result, this can affect the analysis and can make the logistic model presented appear to explain more or less variance than it actually can. For this reason, it would be good to justify their choice as best as possible (Park, 2013). However, a careful review of the literature can always help to ensure that the full range of potential prognostic factors is considered (Reed & Wu, 2013).

The definition of small or medium-sized samples in machine learning can vary depending on the specific context and application. However, in general, small or medium-sized samples refer to datasets that contain a limited number of observations or instances relative to the number of variables or features.

There is no universal threshold for what constitutes a small or medium-sized sample, but in some cases, datasets with fewer than 1000 observations may be considered small, while datasets with up to 10,000 observations may be considered medium-sized. However, the size of the dataset alone may not always be the most important factor when considering the appropriate statistical methods and techniques to use for analysis. Other factors, such as the complexity of the relationships between variables, the distribution of the data, and the amount of noise or variability, can also influence the choice of methods for analysis.

In addition, there are interaction terms between the variables. It is important to follow the rules of thumb that can assess whether the selected predictors are appropriate. Some ways to check if the specificity (true negatives) and sensitivity (true positives) of the model are both above 80%. Then it is likely to have validity for the selected predictors (Oommen et al., 2011). These should definitely be taken into account, because omitting some variables could potentially bias the results. As expected, the solution is to not include as many variables as possible because there is a risk of adding irrelevant variables. Including variables unrelated to the outcome in question tends to inflate the apparent predictive power of the final model (Bender, 2009).

There are several methods for selecting independent variables in a statistical model. Some common methods include:

Forward selection: This method starts with an empty model and iteratively adds one independent variable at a time based on its statistical significance (Kuhn & Johnson, 2008).

Backward elimination: This method starts with a model containing all the independent variables and iteratively removes one variable at a time based on its statistical significance (Kuhn & Johnson, 2008).

Stepwise selection: This method combines the forward selection and backward elimination methods and iteratively adds and removes variables based on their statistical significance (Zhang, 2016).

Ridge regression: This method adds a penalty term to the regression coefficients to shrink them towards zero and avoid overfitting (Zou & Hastie, 2005).

Lasso regression: This method adds a penalty term to the sum of the absolute values of the regression coefficients to perform variable selection and shrinkage (Zou & Hastie, 2005).

Principal component analysis (PCA): This method transforms the original set of correlated variables into a smaller set of uncorrelated variables, called principal components, that capture most of the variation in the data (Jolliffe, 2002).

Decision trees: Decision trees are a popular method for variable selection in machine learning. The tree-based models partition the data into smaller subsets based on the most informative variables (Breiman et al., 1984).

Random Forest: Random Forest is an ensemble method that uses decision trees to select the most important variables. It creates multiple decision trees and selects the variables that are most frequently used across the trees (Breiman, 2001).

The choice of method depends on the specific problem and the nature of the data. It's important to carefully select the independent variables to avoid overfitting or underfitting the model and to ensure the results are interpretable and meaningful (Guyon & Elisseeff, 2003)

Furthermore, it is also important to use Domain knowledge which refers to knowledge and expertise in a specific field or domain, such as statistics or machine learning. In the context of independent variable selection, having domain knowledge is essential for choosing which variables to include in a statistical model (Bellazzi & Zupan, 2008). Domain knowledge can come from a variety of sources, such as previous research studies (Friedman et al., 2001), subject matter experts (Kuhn & Johnson, 2013), or personal experience (Altman, 1991).

In addition to selecting appropriate independent variables, domain knowledge is also important for interpreting the results of the model and drawing conclusions that are relevant to the domain (Bellazzi & Zupan, 2008). This is particularly important in fields such as data mining, machine learning, and clinical medicine where accurate predictions can have significant real-world consequences.

Overall, having domain knowledge in the relevant field is crucial for effective independent variable selection and for making meaningful insights and decisions based on the results of a statistical model (Guyon & Elisseeff, 2003).

The sample size is an important factor that can affect the number of variables that can be selected in a LR model. A general rule of thumb is to have a minimum of 10 instances for every variable studied to ensure adequate statistical power. However, there are ongoing debates about the appropriate sample size requirements for logistic regression analysis. (Agresti, 2007). Moreover, from a simulation study found that the performance of LR was generally reliable across a wide range of sample sizes and number of events per variable studied. This suggests that sample size is not the only important consideration in independent variable selection, and other factors such as the number of events per variable and the distribution of the data should also be taken into account (Menard, 2002).

Apart from the sample size, the absence of data can also pose limitations to the selection of independent variables. The exclusion of participants due to missing data can reduce the sample size, which can restrict the number of variables that can be incorporated into the model. This exclusion may also lead to the removal of specific variables, thus creating a bias in the variables selected for the LR model. To ensure the accuracy of the outcomes, it is crucial to address the problem of missing data in logistic regression analysis to avoid any potential biases. Moreover, self-selection bias is another obstacle that can affect the reliability of the sample and, ultimately, the validity of the results. To overcome such a constraint, it is crucial to select the sample carefully and increase its representativeness. (Jiang et al., 2011; Boateng et al., 2019).

There are several approaches for dealing with missing data in logistic regression analysis. One approach is to simply exclude any participants with missing data, but this can lead to bias in the results if the excluded participants differ systematically from those included. Another approach is to impute the missing values, which involves estimating the missing data based on the available data. This can be done using various methods, such as mean imputation, regression imputation, or multiple imputation. Each of these methods has its own strengths and weaknesses, and the appropriate method depends on the nature of the missing data and the research question at hand. Ultimately, the goal is to choose an approach that minimizes bias and maximizes the accuracy and validity of the results (Little & Rubin, 2014).

Finally, there are also constraints related to the properties of the data that are collected. More specifically, predictor variables that have extremely influential effect called outliers, will affect the results of a LR. Therefore, it is important to detect and address outliers in the data before fitting a LR model, either by removing the outliers or by using robust regression techniques that are less sensitive to extreme values (Mervis, 2017).

Although LR is particularly useful for finding a simple combination of the best predictor variables, such a procedure tends to exploit random sample characteristics (King et al., 2000). Each set of predictors that emerges from one sample may not apply to another sample. It is therefore considered desirable to correct for the exploitation of randomness by cross-replication with a new sample when this procedure is used (Boateng & Abaye, 2019).

4.6 LR Model Evaluation

The goodness of fit of the model LR can be assessed in several ways. First, the overall model is assessed (relationship between all independent variables and the dependent variable). Second, the significance of each independent variable must be evaluated. Third, the predictive accuracy or discriminative ability of the model must be assessed, and finally, the model must be validated. The predictive accuracy or discriminative ability of the model can be evaluated using measures such as the area under the receiver operating characteristic (ROC) curve or the Brier score. Finally, the model should be validated using techniques such as cross-validation or bootstrapping to assess its performance on new data (Boateng & Abaye, 2019).

4.6.1 CAP and ROC Curve

Usually in bibliography they say that properties of Receiver Operating Characteristic (ROC) curves are much more intuitive than the results for the Cumulative Accuracy Profile (CAP) (Engelmann et al, 2003).

The line that is highest of all represents the case of a perfect performance of the model being evaluated. Then the polygon shows the performance of the model being evaluated while the bottom line represents the simple case of zero information which is also the null hypothesis or random assignment of scores. Figure 8, illustrates the concept of CAP.

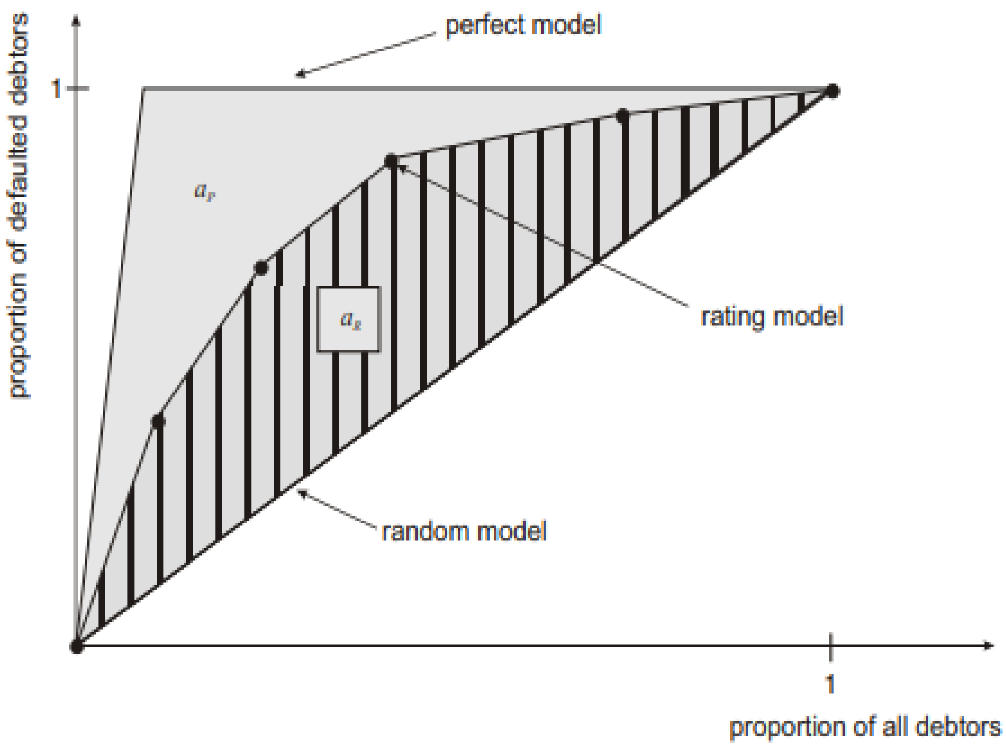

Figure 8: CAP Curve

The CAP curve is like ROC curve. Their main difference is that the CAP curve relates the hit rate to the rate of all alarms while the ROC curve compares it with the rate of only false alarms.

4.6.2 Confusion Matrix

A confusion matrix is a table used to evaluate the performance of a classification model by comparing the predicted class labels with the true class labels of a set of test data. It is a matrix that summarizes the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each class label.

The confusion matrix has two dimensions, the actual class, and the predicted class, as Table 8. The rows correspond to the actual class labels, and the columns correspond to the predicted class labels. The cells of the matrix contain the number of instances that were classified as belonging to a particular combination of actual and predicted class labels.

Table 8: Confusion Matrix

	Predicted Negative	**Predicted Positive**
Actual Negative	True Negative (TN)	False Positive (FP)
Actual Positive	False Negative (FN)	True Positive (TP)

True Positive (TP): Instances that belong to the positive class that are correctly classified as positive.

False Negative (FN): Instances that belong to the positive class that are incorrectly classified as negative.

False Positive (FP): Instances that belong to the negative class that are incorrectly classified as positive.

True Negative (TN): Instances that belong to the negative class that are correctly classified as negative.

From the confusion matrix, several metrics can be calculated to evaluate the performance of the classification model, such as accuracy, precision, recall, F1 score, and specificity. These metrics provide insights into the trade-off between false positives and false negatives and are useful for selecting the best model and tuning the decision threshold of the model.

Accuracy: This is the proportion of all test results that were correctly classified. It is calculated as

$$(TP + TN) / (TP + FP + TN + FN)$$

Sensitivity (also called recall or true positive rate): This is the proportion of true positives among all diseased cases. It is calculated as:

$$TP \, / \, (TP \, + \, FN)$$

Precision: This is the proportion of true positives among all positive cases. It is calculated as:

$$TP \, / \, (TP \, + \, FP)$$

F1 score: This is the harmonic mean of precision and recall. It provides a balance between the two metrics. It is calculated as:

$$2 \, * \, (precision \, * \, recall) \, / \, (precision \, + \, recall)$$

Misclassification rate: This is the proportion of all test results that were incorrectly classified. It is calculated as:

$$(FP \, + \, FN) \, / \, (TP \, + \, FP \, + \, TN \, + \, FN)$$

Specificity (also called true negative rate): This is the proportion of true negatives among all non-diseased cases. It is calculated as:

$$TN \, / \, (TN \, + \, FP)$$

Errors of type I and II: Type I error occurs when a test result is positive, but the individual does not have the disease (false positive). Type II error occurs when a test result is negative, but the individual does have the disease (false negative).

These metrics are useful for evaluating the performance of classification models, particularly in situations where there are imbalanced classes, or the cost of misclassification is asymmetric (Hosmer et al., 2013).

A confusion matrix table can be used to calculate the true positive rate (TPR), false positive rate (FPR), and other performance metrics that are used to evaluate the accuracy of a binary classifier. These metrics are also used to create the receiver operating characteristic (ROC) curve and calculate the area under the curve (AUC).

Table 9 is the confusion matrix table (a confusion matrix as called in machine learning classification projects) that reveals the complete static picture of model performance. Confusion matrix tables could be used as a mean of assessment of competing models only for common cut-off points.

Table 9: Confusion Matrix

Confusion Matrix

	Non-default forecast	**Default forecast**
Non-default	$\frac{Correct\ non-alarms}{Non-alarms}\%$	$\frac{False\ alarms}{Non-defaults}\%$
Default	$\frac{False\ non-alarms}{Defaults}\%$	$\frac{Correct\ alarms}{Defaults}\%$
Total	100	100

Source: Christodoulakis & Satchell, 2008

ROC curve is defined as the plot of the non-diagonal element combinations of a confusion matrix table for all possible cut-off points. It could be presented as a graph, using the plot of correct alarm rate (CAR or TPR) on the vertical axis, versus the false alarm rate (FAR or FPR) on the horizontal axis (Christodoulakis & Satchell, 2008).

$$TPR = \frac{Correct\ alarms}{Defaults}\%$$

$$\text{Errors of type I: FPR} = \frac{False\ alarms}{Non-defaults}\%$$

The Area Under the Curve (AUC) is an overall summary of diagnostic accuracy of the model. AUC equals 0.5 is the null hypothesis when the ROC curve corresponds to random chance. AUC equals 1.0 for perfect accuracy. When the AUC for the model is > 0.5 and very close to 1, it means that the model has a descent discriminatory power. On rare occasions, the estimated AUC is < 0.5, and in this case, it is indicating that the model test does worse than chance.

The area under the ROC curve (AUC) summarizes the information exhibited by the curve in a single statistic defined as

$$AUC = \int_0^1 TPR(FPR)\,\mathrm{d}FPR$$

Furthermore, since AUC is a known metric, it is possible to calculate the Accuracy Ratio (AR). Result comes out by calculating area under the prediction model and random model (aR) divided by calculating area under the perfect model and the random model (aP). An AR closer to 1, means better model. The AR is simply a linear transformation of the AUC, so the knowledge of any one of them could directly give knowledge of the other one (Engelmann et al, 2003).

$$Accuracy\ Rate\ (AR)\ =\ aR/aP\ \ =\ 2\ *\ AUC\ -1$$

or

$$AR\ =\ 2\ *\ (AUC — 0.5)$$

4.7 Statistical Significance of Individual Regression Coefficients

Once a model is performing well with a number of independent variables, it is important to understand the contribution of each independent variable to the predicted outcome. The LR coefficient for the i^{th} independent variable shows the change in the predicted log-likelihood of an outcome for a one-unit change in the i^{th} independent variable while all other things remain the same. In this case, the log likelihood of an outcome is expected to change by b units if the i^{th} independent variable with regression coefficient b_i has a change of one unit, but all other predictors remain the same. In LR there are several tests for evaluating the significance of each independent variable, including the Wald statistic and the likelihood ratio test (Menard, 2002).

4.7.1 Likelihood ratio test

The likelihood ratio test is a statistical test used to compare the fit of two models, one of which is nested within the other. It can be used to evaluate the overall fit of the model and also to assess the contribution of individual predictors to the model. By comparing the likelihood ratio test statistic for the full model (with all predictors) and a reduced model (without one or more predictors), one can determine whether the predictor(s) being removed make a significant contribution to the model's fit. A larger likelihood ratio test statistic indicates a better fit of the full model compared to the reduced model, indicating the contribution of the predictor(s) being tested. The likelihood ratio test for a given parameter compares the probability of obtaining the data when the parameter is zero, L0, with the probability L1 of obtaining the data evaluated at the maximum likelihood estimate of the parameter. It is calculated as follows:

$$\mathrm{G} = -2\ln\frac{L_0}{1-L_1} = \ -2\ln\left(L_0 - \ L_1\right)$$

It is characteristic that this statistic is compared with a X^2 distribution by having only 1 degrees of freedom (DoF). If the predictors are entered hierarchically, it is possible to evaluate the contribution of each predictor, and then in order to determine the contribution of each predictor can be compared each new model to the previous one (Agresti, 2018).

4.7.2 Wald statistic

Statistical significance tests such as Wald test can be applied to each variable coefficient. For each coefficient, the null hypothesis that the coefficient is zero is tested against the alternative that the coefficient is not zero using a Wald test, W_j. A Wald test can also be used to compare a full model that includes all predictor variables with a reduced model in which some coefficients are set to zero.

Asymptotically, the Wald statistic is distributed as a X^2 - distribution and is calculated as follows:

$$W_j = \frac{b_j^2}{SE_{b_j}^2}$$

The Wald statistic is the ratio between the square of the regression coefficient and the square of the standard error of the coefficient. The resulting statistic follows a chi-squared distribution with one degree of freedom under the null hypothesis that the coefficient is equal to zero. The Wald statistic can be used to evaluate the contribution of individual predictors or the significance of individual coefficients in a given model (Bewick et al., 2005). A higher Wald statistic indicates a more significant predictor, and a lower value suggests that the predictor may not be significant in the model (Boateng & Abaye, 2019).

4.7.3 Cox and Snell R-square

Cox and Snell is the name of a statistical method and a family of tests used in binary logistic regression analysis. This method was developed by D.R. Cox and E.J. Snell in 1989 (Cox & Snell, 1989).

The Cox and Snell method is used to determine the goodness-of-fit of a binary logistic regression model. This method compares the observed data with the predicted probabilities of the model and calculates the deviance statistic, which is a measure of the discrepancy between the observed data and the model predictions.

The Cox and Snell test is a type of likelihood ratio test that compares the deviance (that is, the lack of fit) of the fitted model with the deviance of a null model. The null model is a model with no predictor variables, and its deviance is the maximum deviance that can be observed by any model that does not include predictor variables. The test statistic is calculated as the difference between the deviance of the fitted model and the deviance of the null model, divided by the deviance of the null model.

The Cox and Snell method is widely used in binary logistic regression analysis to assess the goodness-of-fit of the model and to determine whether the model adequately explains the relationship between the predictor variables and the binary outcome variable.

4.7.4 Nagelkerke R-square

Nagelkerke R-square test is a statistical method used in logistic regression analysis to assess the predictive power of a model and it is actually an adjusted version of Cox and Snell statistic. This method was proposed by N.J.D. Nagelkerke in 1991 (Nagelkerke, 1991).

The Nagelkerke R-squared is a measure of the proportion of variability in the binary outcome variable that is explained by the predictor variables in the logistic regression model. As mentioned above, this metric is a modified version of the Cox and Snell R-squared and is an indicator of the proportional reduction in deviance of the fitted model compared to the null model.

The Nagelkerke R-squared is calculated by taking the ratio of the Cox and Snell R-squared to the maximum possible value of the Cox and Snell R-squared. The maximum possible value of the Cox and Snell R-squared is obtained by transforming the deviance of the null model using a logistic transformation. The resulting value is equivalent to the maximum possible value of R-squared in a linear regression model.

The Nagelkerke R-squared ranges from 0 to 1, with higher values indicating a stronger relationship between the predictor variables and the binary outcome variable. This measure is widely used in logistic regression analysis to compare the predictive power of different models and to determine the contribution of each predictor variable to the model's overall predictive power (Nagelkerke, 1991).

4.7.5 Hosmer and Lemeshow Test

The Hosmer-Lemeshow test is a statistical test used to assess the goodness of fit of a logistic regression model. This test was proposed by David Hosmer and Stanley Lemeshow in 1980 (Hosmer et al., 2013).

The Hosmer-Lemeshow test involves dividing the dataset into a number of equally sized groups based on the predicted probabilities from the logistic regression model. The observed frequencies and expected frequencies of the binary outcome variable are then calculated for each group. The expected frequencies are calculated based on the predicted probabilities from the model.

The test statistic is calculated by comparing the observed and expected frequencies for each group using a chi-square test. A small test statistic indicates a good fit between the observed and expected frequencies, while a large test statistic indicates a poor fit.

The Hosmer-Lemeshow test is often used to compare the goodness of fit of different logistic regression models and to identify the model that best fits the data. It is also used to assess the overall adequacy of a logistic regression model and to detect any departures from the assumptions of the model (Hosmer et al., 2013).

4.7.6 Statistical Significance

Statistical significance is a term used in Statistics to determine if a relationship or difference between two or more variables is real or simply due to chance. This is done by calculating a p-value, which is the probability of obtaining a result as extreme or more extreme than the one observed in the sample, assuming that there is no true difference or relationship in the population. If the p-value is below a predetermined significance level (related to type I errors), usually 0.05, the result is deemed statistically significant. This indicates strong evidence against the null hypothesis, which is the assumption that there is no true difference or relationship in the population.

It is important to note that statistical significance does not necessarily imply practical significance or importance. A statistically significant result may not be meaningful in the context of the research question or may have little practical significance. Hence, it is crucial to consider both statistical and practical significance when interpreting research findings (Gelman & Stern, 2006). In other words, the domain knowledge should always drive a final decision regarding significance.

4.8 Probit Regression

Probit regression is also called as a probit model and is a type of regression. The actual word is portmanteau, which is coming from two other words ‘probability’ and ‘unit’ (Bliss, 1934). The dependent variable can be classified in only two values, so probit as logit regression is used to model binary (dichotomous) outcome variables.

The model is used to estimate the probability to fall into a specific one of the categories. This refers to an observation with characteristics. There is more than that since the classification of the observations based on their predicted probabilities is a type of binary classification model.

The probit model is a type of generalized linear model (GLM) that is often used to model binary outcomes. It is similar to the logistic regression model, which is also a GLM used for modelling binary outcomes but uses a different link function. Specifically, the probit model uses a probit link function, which maps the linear predictor to the cumulative distribution function of the standard normal distribution, while the logistic regression model uses a logit link function, which maps the linear predictor to the log-odds of the outcome. Like logistic regression, the probit model is typically estimated using maximum likelihood estimation, and the resulting estimation is sometimes referred to as probit regression. (Aldrich et al., 1984).

In a probit model, the probability of an event occurring is modelled using the inverse standard normal distribution, with the linear combination of the predictor variables used as inputs to this distribution. Specifically, the probit model assumes that the latent variable underlying the observed binary outcome follows a normal distribution, and that the observed outcome is determined by whether the latent variable exceeds a certain threshold value. The probit model provides a way to estimate the coefficients of the linear combination of predictors that are associated with the probability of the outcome occurring, which can be used to make predictions and test hypotheses about the relationship between the predictors and the outcome. Thus, the probit regression uses an inverse normal link function as follows:

$$\ln(\mathrm{p}) = \Phi^{-1}(P)$$

The logistic regression on the other hand uses a logit link function, so the differences are obvious:

$$\ln(\mathrm{p}) = \ln\left[\frac{p}{1-p}\right]$$

They use different link functions and in generalized linear models, instead of using Y as the outcome, it is used a function that refers to the mean of Y. They are just the most common link functions, but they are not the only ones of course that can be used for modelling categorical data (Agresti, 2018).

As it is presented in Figure 9, the logit function is similar but has thinner tails than the normal distribution.

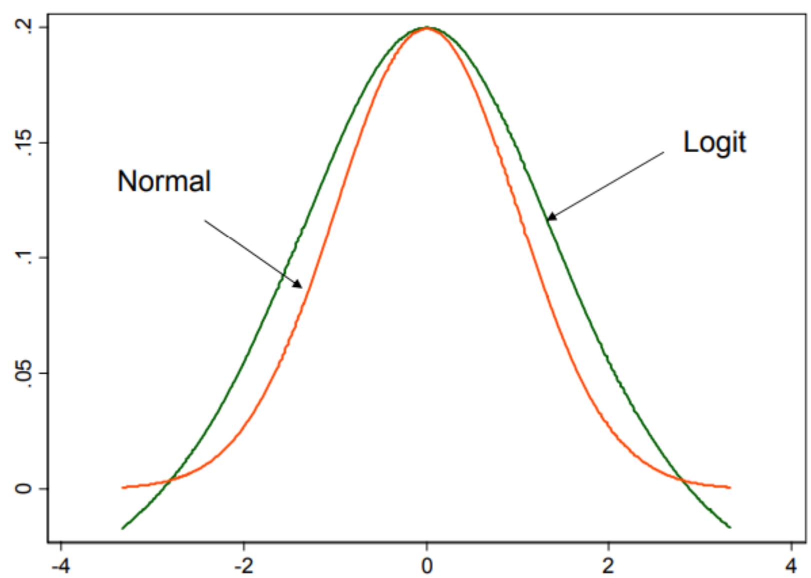

Figure 9: Logit vs Probit distribution

The difference in the overall results of the probit model is usually slight to non-existent. In a practical level and since the results are similar, the choice usually depends on the interpretation, and it is not a serious matter to use the one of them.

There are two ways to interpret coefficients for probit models. The difference in Z score is linked with the difference of each one-unit in the predictor variable and by using the model in order to calculate predicted probabilities at different values of X.

As per the sigmoidal relationship between a predictor and probability, in the logistic regression and the probit regression it is nearly identical. Just like in logistic regression, the difference in the probability in the predictor is not equal for each 1-unit change. In addition, the impact on the probability of a 1-unit difference in X will be bigger in the middle than near 0 or 1.

Additionally, the probit regression model, like any other statistical model, can be affected by outliers in the data. Outliers are observations that are significantly different from the rest of the data and can have a large influence on the estimated coefficients of the model.

However, the probit regression model can provide some degree of protection against the influence of outliers through the use of the probit link function. The probit link function maps the linear predictor to the cumulative distribution function of the standard normal distribution, which is a symmetric distribution that assigns low probabilities to extreme values. Therefore, even if there are outliers in the data that may cause the linear predictor to be far from the mean, the probit link function will assign a low probability to these extreme values, resulting in a smaller influence on the estimated coefficients of the model (Greene, 2018).

It is important to note that the strictness of treatment to outliers in probit regression, ultimately depends on the specific characteristics of the data and the modelling assumptions made.

A probit model can also be used to understand the conditions that lead to the outcome variable being close to zero or close to one, even if the outcome variable is a continuous variable between zero and one.

In this case, the probit model would be used to estimate the probability of the outcome variable being greater than or equal to a threshold value, which would represent the decision boundary between values close to zero and values close to one. The threshold value could be chosen based on the specific context of the problem, or it could be estimated from the data using a method such as the median or the mean of the outcome variable.

Once the threshold value is defined, the probit model would estimate the probability of the outcome variable being greater than or equal to the threshold value as a function of one or more independent variables. The coefficients of the model would provide information about the direction and strength of the relationship between the independent variables and the probability of the outcome variable being close to one.

For example, suppose we have a continuous outcome variable Y that takes on values between zero and one, and we want to understand the conditions that lead to Y being close to one. We can define a threshold value c, such that if Y >= c, then Y is considered to be close to one.

The probit model for this scenario can be written as:

$$P(Y \ >= \ c \mid X) \ = \ \Phi(\beta 0 \ + \ \beta 1 X)$$

where Φ is the cumulative distribution function of the standard normal distribution, β0 and β1 are the intercept and slope coefficients, respectively, and X is the independent variable.

The coefficients β0 and β1 represent the effect of the independent variable on the probability of Y being greater than or equal to c. By examining the coefficients and their significance, we can identify the conditions that are associated with a higher probability of Y being close to one. Similarly, we can use the model to understand the conditions that lead to Y being close to zero by defining a threshold value that is close to zero and estimating the corresponding probit model (Greene, 2018).

4.9 Panel Data Regression

Panel data analysis is a statistical method based on time-varying data. It is mainly used in the field of econometrics to analyse two-dimensional data presented in a table (Maddala, 2001). The data collected consists of records with a relevant chronological sequence. Their analysis process is done by the regression method in the two dimensions of the table. There is also the multidimensional analysis as it is called, and it is an econometric method in which the data collected is in more than two dimensions with one of them always being the dimension of time (Davies & Lahiri, 1995). For example, if the data is collected on a population of individuals, the dimensions could include variables such as age, gender, income, education level, etc. The dimension of time would refer to changes in these variables over time, such as how income levels or education levels change over different time periods. The dimensions can be continuous or categorical variables, and they can be analysed in various ways depending on the research question and the type of analysis being used (Goh & Lee, 2018).

Panel data regression allows to estimate the relationships between variables while controlling for both time-invariant and time-varying factors. This can help to identify the effects of particular variables over time, while also accounting for any differences between individuals or entities (Wooldridge, 2010). It has many applications, including in the fields of finance, economics, and social sciences. It is useful for analysing data that includes multiple time periods and different individual or group units, allowing researchers to identify patterns and trends that may not be visible with other types of analysis.

The table data can be seen as a combination of two categories that encompass features of both time series and cross-sectional data. This type of data can be modelled as a schedule in which the same entities are observed periodically. Time series data typically involves repeated observations of a single variable over time, while cross-sectional data involves multiple records and corresponding variables observed at a single point in time. By combining these two types of data, researchers can gain insights into how variables change over time and how they are related to other variables at a particular point in time. This combination comes from data as shows Figure 10:

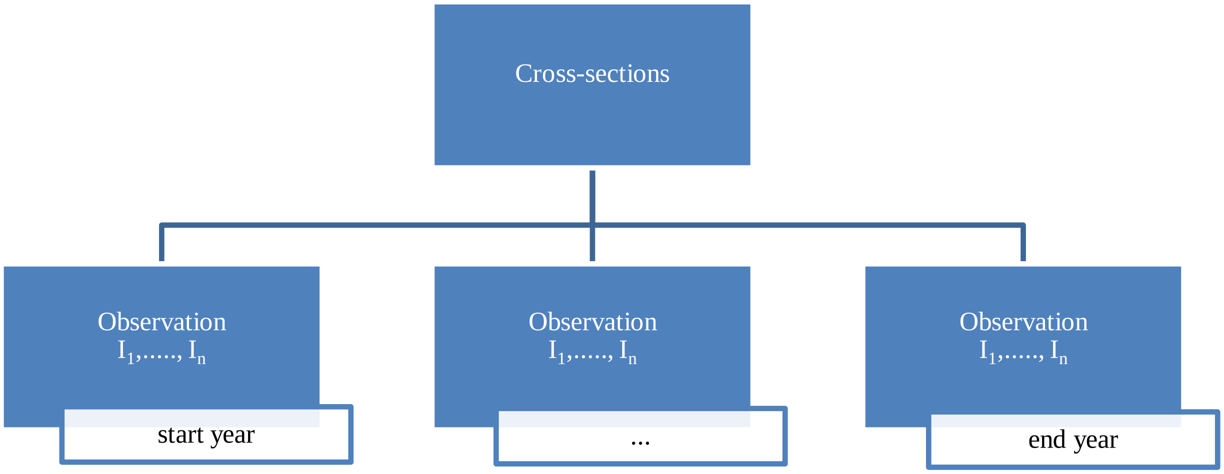

Figure 10: Illustration of Panel Data

A panel data regression model form is as follows:

$$y_{it} = a + bx_{it} + \varepsilon_{it}$$

To explain the previous model, it is important to say that y is the dependent variable, x is the independent variable, and a and b are the coefficients that estimate the effect of x on y. The indices i and t represent individual objects and time, respectively. The error term, represented by ε, captures the unexplained variation in y that is not accounted for by x and the coefficients. The error term can be decomposed into fixed effects or random effects models depending on whether the variation in ε is assumed to be non-stochastic or stochastic over i or t. However, this is a very general description, and the specifics of the regression model would depend on the research question and the type of data being analyzed (Hsiao et al., 1999).

The three main approaches to panel data analysis are Independently Pooled Panels (Pooled OLS), Fixed Effects Models (FE), and Random Effects Models (RE).

- OLS stands for Ordinary Least Squares, which is a method used to estimate the parameters of a linear regression model. It is a commonly used method in econometrics, and it aims to minimize the sum of squared residuals between the observed values and the predicted values of the dependent variable. OLS assumes that the independent variables are exogenous, meaning that they are not affected by the error term of the regression model.
- FE models, on the other hand, assume that there are unique individual-level effects that are correlated with the dependent variable but are time-invariant. This method controls for individual-level heterogeneity that could be driving the relationships between the variables.
- RE models assume that there are individual-level effects that are uncorrelated with the independent variables in the model. This method allows for more variation in the individual-level effects than FE models, and it assumes that these effects follow a normal distribution.

Choosing between these three methods depends on the research question, the nature of the data, and the assumptions that need to be met. In general, FE and RE models are preferred over Pooled OLS when there is individual-level heterogeneity that could affect the dependent variable. The choice between FE and RE models depends on the assumptions of the data, and the Hausman-Test can be used to determine which model is more appropriate (Wooldridge, 2010).

In panel data analysis, the assumptions of simple linear regression models still hold, but they may be more challenging to meet due to the longitudinal nature of the data. Here is some additional information on each of the assumptions:

- Linearity: The relationship between the dependent variable and each independent variable is linear. Non-linear relationships can be addressed through transformations or nonlinear models.
- Exogeneity: The independent variables are not correlated with the error term. In other words, the independent variables are not affected by the dependent variable or other factors in the model.
- Homoskedasticity: The variance of the error term is constant across all values of the independent variables. Heteroskedasticity, where the variance of the error term varies with the independent variables, can be addressed through robust standard errors or other methods.
- Non-autocorrelation: The error terms are not correlated with each other over time. Autocorrelation can be addressed through time series models, such as autoregressive models.
- Independent variables Non-Stochastic: The independent variables are not affected by random factors or measurement errors.
- No Multicollinearity: The independent variables are not highly correlated with each other. High multicollinearity can lead to unstable estimates and difficulty in interpreting the coefficients.

In panel data analysis, violating the assumptions can lead to biased or inefficient estimates, as well as incorrect conclusions (Wooldridge, 2010).

Chapter 5: Empirical Analysis

In the previous sections the various approaches in ESG definitions, methodologies for measuring the appropriate variables and the regulations that govern the ESG concept globally were discussed and analysed.

In the present section, the attention is turned to a specific case study using data relevant to the economic activity, pertinent to a number of different sectors and industries. Since the availability of data was limited, it was imperative to conduct a data collection phase from scratch using expert judgment based on criteria and regulations.

5.1 Data Collection

Initially, data collection focused on a number of sectors most relevant to the economy of Greece. While the adoption of ESG principles in Greece is still in its early stages, there has been some progress in recent years. Since, data collected included only a few ESG incidents and the trained models had a low discriminatory power, so it was necessary to conduct further research and include data from firms of other European Countries. In particular, Food, Beverage, Tobacco, Construction, Electric utilities, Gas utilities, Accommodation, Restaurants, and Travel Agencies.

However, since there is a relationship between some of the above industries, it was useful to create larger groups. Therefore, Table 10 presents the final grouping of the sectors integration.

Table 10: Sectors Integration

Sectors	Grouping
Food	Food, Beverage, Tobacco
Construction	Construction
Electric & Gas Utilities	Electric utilities, Gas utilities
Hospitality	Accommodation, Restaurants, Travel Agencies

The variable "Sector" is a dummy variable included in the model for the purpose of ranking based on necessity, with a focus on essential industries.

1 = Food, 2 = Construction, 3 = Electric & Gas Utilities, 4 = Hospitality

The analysis followed all published reports for 50 listed European corporates for 7 years (from 2015 to 2021) using Bloomberg terminal. The dataset was created entirely from scratch for the purpose of this study. More specifically, conducted a content analysis of published reports that have been read and analyzed using ESG criteria in order to be characterized as ESG negative incidents. It was a procedure that needed expert judgment based on ESG criteria and regulations to record each one incident. It is important to say that all incidents are unique records that do not take under consideration their certainty and severity but not even any recurring events during a certain year. As a result, the analysis concluded with a sample consisting of 350 observations (50 companies x 7 years), containing 76 ESG incidents in total, of which there is no Environmental, Social and Governance classification.

Furthermore, for each observation of the binary dependent variable with values 0 or 1 (which corresponds to a particular corporate and year combination) a number of 30 independent variables coming from each firm's financial accounts as well as macro and country risk measures have been collected. Variables measuring firm characteristics i.e., size, sector, liquidity, profitability, debt, etc. Other variables measure sector characteristics i.e., intensity of competition, etc., and others measure country characteristics. Table 11 provides the description of each independent variable:

*Table 11: Variables Definitions**

Feature Names	Feature Definition	Source
	a. Company's Data	
LogRevenues	Total Revenues	BBG
LogEmployees	Number of people employed by the company, based on the number of full-time equivalents. If unavailable, then the number of full-time employees is used, excluding part time employees.	BBG
	b. Company's Financial Data	
Current_ratio	Ratio to indicate the company's ability to pay back its short-term liabilities with its short-term assets. Unit: Actual. Calculated as: Current Assets / Current Liabilities	BBG
EBIT_margin_pct	Ratio which measures the company's profitability. Unit: Actual. Calculated as: (Trailing 12M Operating Inc (Loss) / Trailing 12M Net Sales) *100.	BBG
Return_on_Assets_pct	Indicator of how profitable a company is relative to its total assets, in percentage. Return on assets gives an idea as to how efficient management is at using its assets to generate earnings. Calculated as: (Trailing 12M Net Income / Average Total Assets) * 100	BBG
Debt_to_Assets_pct	Leverage ratio in percentage that defines the total amount of debt relative to assets. This enables comparisons of leverage to be made across different companies. Calculated as: Total Debt *100 / Total Assets	BBG
Debt_to_EBIT	Ratio of total debt to earnings before interest and taxes (EBIT). A low number indicates that the company can service its debt from current earnings, units in actual. Calculated as: Total Debt / Trailing 12 Month EBIT	BBG
Interest_to_EBIT	Interest Expense / EBIT	BBG
Earnings_per_share	Earnings Per Share (EPS) is the portion of a company's profit allocated to each shareholder. It is calculated based on Net Income Available for Common Shareholders divided by the Basic Weighted Average Shares Outstanding. This field returns Bottom-line Earnings Per Share when FPDF Settings for 'Non-GAAP Adjustments.'	BBG

Revenue_per_ share	Total Revenues / shares outstanding	BBG
Liabilities_to_ Assets	Total Assets / Total Liabilities	BBG
	c. Country's Data	
Unemployment	Country's unemployment rate	Eurostat + ONS
GDP_Growth	GDP growth (annual %)	Eurostat
Consumer_ Confidence_ Indicator	The Consumer Confidence Indicator (CCI) provides an indication of future developments of households' consumption and saving, based upon answers regarding their expected financial situation, their sentiment about the general economic situation, unemployment and capability of savings Values above 100 signals a boost in the consumers' confidence towards the future economic situation, as a consequence of which they are less prone to save, and more inclined to spend money on major purchases in the next 12 months. Values below 100 indicate a pessimistic attitude towards future developments in the economy, possibly resulting in a tendency to save more and consume less.	Eurostat
Res_Insolv_100	The score for resolving insolvency is the simple average of the scores for each of the component indicators: the recovery rate of insolvency proceedings involving domestic entities, as well as the strength of the legal framework applicable to judicial liquidation and reorganization proceedings.	World Bank
Starting_Bus__100	The score for starting a business is the simple average of the scores for each of the component indicators: the procedures, time, and cost for an entrepreneur to start and formally operate a business, as well as the paid-in minimum capital requirement.	World Bank
RR_100	The recovery rate is recorded as cents on the dollar recovered by secured creditors through judicial reorganization, liquidation, or debt enforcement (foreclosure or receivership) proceedings. The calculation takes into account the outcome: whether the business emerges from the proceedings as a going concern, or the assets are sold piecemeal.	World Bank
Voice_and_ Accountability	Reflects perceptions of the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media.	World Bank
Political_Stability	Political Stability and Absence of Violence/Terrorism measures perceptions of the likelihood of political instability and/or politically motivated violence including terrorism.	World Bank

Government_ Effectiveness	Reflects perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies.	World Bank
Regulatory_ Quality	Reflects perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development.	World Bank
Rule_of_Law	Reflects perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence.	World Bank
Control_of_ Corruption	Reflects perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests.	World Bank
Count_Climate_ 100	The CCPI assesses each country's performance in four categories: GHG Emissions (40% of the overall ranking), Renewable Energy (20%), Energy Use (20%) and Climate Policy (20%). In addition, the question is answered to what extent the respective country acts adequately in the areas of Emissions, Renewable Energies and Energy Use in order to achieve the Paris climate targets. In interpreting the results, it is important to note that the CCPI is calculated using production-based emissions only.	German Watch
	d. Sector's Data	
Revenues_share	Company's revenues share (Sector level defined by country variable)	Calculations
Tourist_arrivals_EU	Tourist arrivals in EU the specific year	INSETE
Tourist_arrivals_EU _chg	Change rate for Tourist arrivals between two years	Calculations

*All the independent variables have been collected based on the previous financial year.

5.2 Data Preparation

As mentioned earlier, the data available for ESG topic is limited and this created the need to collect the data for this research from scratch. So, it was deemed necessary to complete the following steps to collect and prepare the data for analysis.

Step 1: Data sources identification and access

This is the first but also the most essential data preparation step, it involves identifying the necessary data and their repositories (sources). Of course, it does not only include identifying all possible data sources, but also all those that apply to the desired analysis. This means that there must first be a plan that includes specific requirements to proceed to the next steps.

Step 2: Ingest

Once the data is identified, it needs to be transferred to the analysis tools, for this purpose it should come in a format. This could likely be some combination of structured and semi-structured data. In different types of repositories. But this is not enough and at a later stage all the data should be imported into a common repository. In this study the dataset was initially separated by sectors, but later sector became an independent variable, so the complete dataset had almost the same format. This step is an initial data validation action.

Step 3: Data cleaning and formatting

Data cleaning ensures that the dataset provides valid measurements in order to proceed further the analysis. There are many different potential problems that need to be addressed regarding the data received. Common cases can be missing values, out-of-range values, null values, ~~and~~ blanks that blur the values, and outliers that could skew the results of the analysis. In the present study since the sample was collected from scratch using real company and business data, it was not possible to recognise and exclude outliers ex ante. In addition, missing values regarding the independent variables, were determined from the beginning and those firms were excluded from the sample due to limited access to their financial data. However, there were only two independent variables referring to the tourism sector, with the result that in the consolidation of the dataset missing values emerged for the other sectors. In this case, it was necessary to exclude those two independent variables since the analysis of this research would not take account the sector segmentation.

After completing data cleaning some further formatting might me needful in order to assure that the data will have the appropriate format for the proposed mode of analysis and the desired results. For example, in this step we resolve issues like multiple data formats or inconsistent abbreviations. In this study this was observed mainly in the variable related to the reference year which should have been transformed into plain years and not in any other date format.

Step 4: Data combination

When the data set has been cleaned and formatted, then it can finally be consolidated as a holistic dataset. Once the data is finalised, there is a second chance for validation.

Step 6: Data Analysis

Once the analysis has begun, changes to the data set should be made with careful consideration and only based on the results (for example, in possible correlations and patterns) and with careful consideration. This is important because changes in the data can distort the results of the analysis. On some occasions, changes might even make it impossible to determine whether different results are caused by changes in the data or due to the algorithms used or even to any of the values of the hyperparameters.

After completing the above preparations and modifications, the final structure of the dataset is presented in Table 12:

Table 12: Data structure

FIRM	YEAR	DEFAULT INDICATOR Yi= 0 OR 1	FEATURE VALUES FROM THE END OF THE YEAR			
XAA	2015	0	Xi1	Xi2	…	Xik
XAB	2015	0	…	…	…	…
XAC	2015	1	…	…	…	…
XAD	2015	1	…	…	…	…
XAE	2015	0	…	…	…	…
…			…	…	…	…
XAA	2021	0	…	…	…	…
XAB	2021	1	…	…	…	…
XAC	2021	0	…	…	…	…
XAD	2021	0	…	…	…	…
XAE	2021	0	…	…	…	…

5.3 Descriptive Statistics

This section includes the use of descriptive statistics for both the binary categorical dependent variable, ESG, as well as the independent variables. The analysis was carried out using SPSS and Stata software for reasons of cross-validation of results enhancing the capability of implementing additional statistical tests since is a common place in scientific research to employ a combination of analytical methods scattered in various software products.

Binary data as the ESG dependent variable presented in Table 13, only takes one of two values such as ‘non-default’ or ‘default’. The values 0 and 1 are assigned to the two states. For a single variable there are only two ways of summarising the information: proportions that can be classified as risks or rates, and odds.

In general, if there are x events and y non-events, the odds of an event are x/y as was explained above and the proportion is x/(x+y). It is a simple matter to relate odds (o) to proportions (p). The odds of an event are o = p/(1-p). Thus, as this research refers to ESG, the odds are the ratio of the proportion of 1’s to the proportion of 0’s.

Table 13: ESG Frequencies and Percentages

ESG Events

		Frequency	Percent
Valid	0	274	78,3
	1	76	21,7
	Total	350	100,0

It is obvious from the results that 76 out of 350 observations were 1, i.e., a proportion, 0.217, or a percentage, 21.7%, were 1. This is defined as the number of ESG incidents. An alternative way of looking at the 350 observations, is to say that out of the 350 observations, 76 observations are 1 and 274 are 0 i.e., a ratio of 76:274.

Moreover, ESG events also show different frequencies if observed in relation to years or sectors. Tables 14 and 15 with Figures 12 and 13 provide somehow portray of how the events are presented classified in sectors and years.

Even if the data are presented by sector and year for the analysis, it is assumed that each observation is independent since the occurrence of one observation provides no information about the occurrence of the other observation.

Table 14: ESG event by Sector

Sector * ESG Crosstabulation

		ESG		
		0	1	Total
Sector[a]	1	44	26	70
	2	73	25	98
	3	69	15	84
	4	88	10	98
Total		274	76	350

a: 1 = Food, 2 = Construction, 3 = Electric & Gas Utilities, 4 = Hospitality

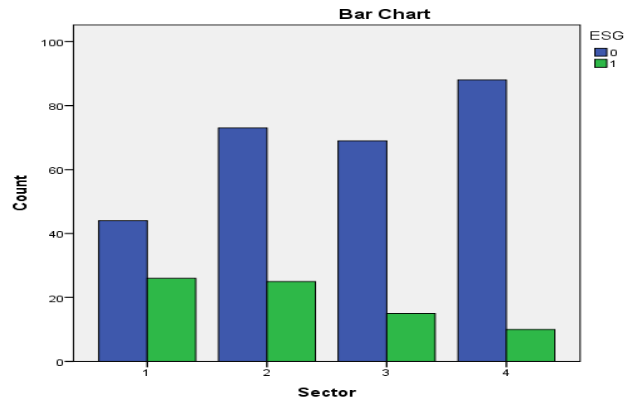

Figure 11: ESG events per Sector

Table 15: ESG event by Year

Year * ESG Crosstabulation

		ESG		
		0	1	Total
Year	2015	41	9	50
	2016	45	5	50
	2017	45	5	50
	2018	39	11	50
	2019	43	7	50
	2020	34	16	50
	2021	27	23	50
Total		274	76	350

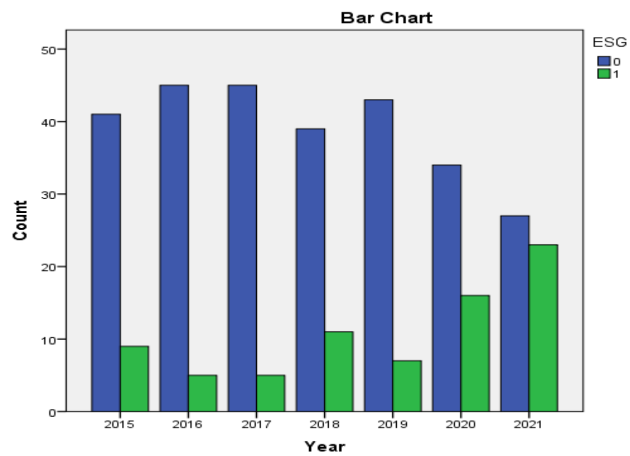

Figure 12: ESG events per Year

By observing Figure 13, at this stage it would be important to explain the business implications of the distribution of incidents, especially in relation to the years. It is obvious that there exists an upward trend of ESG incidents as time evolves, so it could be said that it follows an evolutionary process. The introduction of international standards in 2015 has led to increased awareness and reporting of ESG incidents, and as companies evolve in their behaviours and practices, more incidents can be considered as falling within the ESG category.

So, the increase in ESG incidents observed in the last years of the study can be attributed to changes in the market's perception of ESG issues and companies' efforts to adapt to new regulations and standards.

5.4 Explanatory Analysis (Diagnostic Stage)

Various descriptive statistics are applicable to the independent variables as Table 16 displays.

Table 16: Descriptive Statistics

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
		Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Sector	350	3	1	4	2,60	1,097	-,073	,130	-1,315	,260
LogRevenues	350	25,04	,00	25,04	20,7189	3,46975	-1,840	,130	6,521	,260
LogEmployees	350	13,06	,00	13,06	8,3211	2,97725	-,828	,130	,110	,260
Current_ratio	350	14,31	,00	14,31	1,5453	1,54561	4,167	,130	23,655	,260
EBIT_margin_pct	350	10689,54	-3268,93	7420,61	4,8362	478,16345	8,879	,130	177,544	,260
Return_on_Assets_pct	350	69,95	-19,95	50,00	2,8348	6,63797	1,732	,130	12,526	,260
Debt_to_Assets_pct	350	183,73	,00	183,73	33,7304	23,78814	2,444	,130	11,599	,260
Debt_to_FCF	350	4057,50	-3490,99	566,51	-23,2425	275,92054	-10,890	,130	130,335	,260
Debt_to_EBIT	350	83363,44	-707,44	82656,00	241,0629	4418,09548	18,706	,130	349,933	,260
Interest_to_EBIT	350	10154,11	-29,11	10125,00	29,1353	541,19588	18,708	,130	349,992	,260
Earnings_per_share	350	69,32	-9,29	60,03	2,4830	7,66448	4,890	,130	27,484	,260
Revenue_per_share	350	1185,46	,00	1185,46	46,0731	153,94113	6,416	,130	41,949	,260
Unemployment	350	23,40	3,10	26,50	10,9666	7,24649	,774	,130	-,804	,260
GDP_Growth	350	18,10	-10,80	7,30	,5400	3,71243	-1,798	,130	2,555	,260
Consumer_Confidence_Indicator	350	64,48	-62,88	1,60	-16,3702	17,85956	-1,427	,130	,798	,260
Res_Insolv_100	350	47,00	45,28	92,28	70,8607	13,49718	-,208	,130	-1,404	,260
Starting_Bus__100	350	18,07	77,93	96,00	89,8410	4,12912	-,725	,130	-,191	,260
RR_100	350	59,90	30,00	89,90	62,9869	22,71116	-,322	,130	-1,661	,260
Voice_and_Accountability	350	1,18	,43	1,61	1,0718	,32304	-,382	,130	-1,082	,260
Political_Stability	350	1,67	-,23	1,44	,3934	,37341	,736	,130	,524	,260
Government_Effectiveness	350	2,33	-,22	2,11	1,0472	,61350	-,397	,130	-1,200	,260
Regulatory_Quality	350	1,76	,15	1,91	1,0992	,57390	-,089	,130	-1,501	,260
Rule_of_Law	350	1,96	,07	2,03	1,0779	,64838	-,286	,130	-1,526	,260
Control_of_Corruption	350	2,37	-,18	2,19	,9950	,83286	-,149	,130	-1,636	,260
Count_Climate__100	350	34,62	41,66	76,28	59,6054	6,96732	-,040	,130	-,783	,260
Revenues_share	350	24,16	,00	24,16	2,8737	4,50356	2,105	,130	4,210	,260
Liabilities_to_Assets	350	203,19	3,28	206,47	67,9111	26,65764	1,163	,130	6,215	,260
Tourist_arrivals_EU	98	532673981	467120969	999794950	840269048	166531342	-1,439	,244	,973	,483
Tourist_arrivals_EU_chg	98	,59	-,53	,06	-,0386	,20185	-2,065	,244	2,327	,483

The table shows descriptive statistics for 31 independent variables, including the sample size (N), range, minimum, maximum, mean, standard deviation, skewness, and kurtosis.

Each row represents a different variable, including the sector, financial ratios (such as current ratio and debt-to-assets ratio), economic indicators (such as unemployment and GDP growth), governance indicators (such as political stability and rule of law), and environmental indicators (such as count of climate events).

The range shows the difference between the maximum and minimum values for each variable, while the mean shows the average value. The standard deviation indicates how much variation there is in the data, with a higher standard deviation indicating more variability.

Skewness measures the degree of asymmetry in the distribution of the data. A positive skewness indicates that the distribution has a longer tail on the positive side, while a negative skewness indicates a longer tail on the negative side.

Kurtosis measures the degree of peakedness or flatness in the distribution of the data. A positive kurtosis indicates a more peaked distribution, while a negative kurtosis indicates a flatter distribution.

The table provides useful information for understanding the distribution and variability of each variable, which can be helpful for further analysis and modelling.

Understanding the interactions and patterns between independent variables is essential for determining the significance of each variable and predicting the behaviour of the dependent variable. Identifying these interactions and patterns can also help to determine which variables to include in the models and how to interpret the results.

Interaction between independent variables occurs when the effect of one independent variable on the dependent variable depends on the level of another independent variable. This means that the relationship between the dependent variable and one independent variable change depending on the level of another independent variable.

Patterns between independent variables refer to the relationship or association between different independent variables. These patterns can be observed through statistical measures such as correlation coefficients, scatter plots, or regression analysis. Positive correlations indicate that the variables move in the same direction, while negative correlations indicate that they move in opposite directions.

Table 17 displays the correlation matrix which shows the Pearson correlation coefficient values between each pair of variables. The Pearson correlation coefficient measures the linear relationship between two variables, with values ranging from -1 to 1. A value of 1 indicates a perfect positive linear relationship, -1 indicates a perfect negative linear relationship, and 0 indicates no linear relationship.

Table 17: Correlations

Correlations

		Sector	LogRevenues	LogEmployees	Current_ratio	EBIT_margin_pct	Return_on_Assets_pct	Debt_to_Assets_pct	Debt_to_FCF	Debt_to_EBIT	Interest_to_EBIT	Earnings_per_share	Revenue_per_share	Unemployment	GDP_Growth	Consumer_Confidence_Indicator	Res_Insolv_100	Starting_Bus__100	RR_100	Voice_and_Accountability	Political_Stability	Government_Effectiveness	Regulatory_Quality	Rule_of_Law	Control_of_Corruption	Count_Climate__100	Revenues_share	Liabilities_to_Assets	Tourist_arrivals_EU	Tourist_arrivals_EU_chg
Sector	Pearson Correlation	1	-,311**	-,207**	-.045	.001	-,154**	-.036	-.057	.020	.020	-,304**	-,228**	-,251**	,107*	,228**	.079	-,210**	.047	-,131*	.018	-.017	.051	.031	.013	.067	.062	-.097	.c	.c
	Sig. (2-tailed)		.000	.000	.398	.984	.004	.506	.290	.707	.713	.000	.000	.000	.046	.000	.140	.000	.378	.014	.733	.746	.342	.562	.811	.212	.249	.071	0.000	0.000
LogRevenues	Pearson Correlation	-,311**	1	,898**	-,131*	.030	,222**	-,144**	-.026	.047	.047	,133*	,166**	-.093	-.055	,254**	,431**	.041	,509**	,533**	,179**	,559**	,432**	,481**	,476**	,187**	,560**	,212**	.096	.103
	Sig. (2-tailed)	.000		.000	.014	.570	.000	.007	.627	.376	.378	.013	.002	.082	.309	.000	.000	.450	.000	.000	.001	.000	.000	.000	.000	.000	.000	.000	.347	.314
LogEmployees	Pearson Correlation	-,207**	,898**	1	-,296**	.022	,191**	-,192**	-.063	.027	.027	,106*	,151**	-.105	-.049	,296**	,365**	.076	,478**	,448**	,176**	,507**	,364**	,441**	,412**	,190**	,630**	,166**	.032	.032
	Sig. (2-tailed)	.000	.000		.000	.684	.000	.000	.239	.611	.612	.048	.005	.050	.358	.000	.000	.155	.000	.000	.001	.000	.000	.000	.000	.000	.000	.002	.757	.752
Current_ratio	Pearson Correlation	-.045	-,131*	-,296**	1	.020	,181**	-,112*	.034	-.013	-.014	,132*	.001	-.091	-.010	.020	-.001	-,118*	-.065	-.009	.079	-.032	.012	-.017	-.002	.005	-,210**	-,284**	.016	-.049
	Sig. (2-tailed)	.398	.014	.000		.705	.001	.036	.526	.803	.797	.014	.989	.088	.859	.706	.983	.027	.224	.874	.139	.550	.822	.746	.970	.924	.000	.000	.878	.629
EBIT_margin_pct	Pearson Correlation	.001	.030	.022	.020	1	,217**	-.008	.003	.000	.000	.018	.002	-.047	-,113*	.073	.014	.045	.015	.055	.026	.032	.039	.035	.026	.013	.006	-.030	,456**	,425**
	Sig. (2-tailed)	.984	.570	.684	.705		.000	.874	.953	.999	.995	.731	.965	.380	.034	.172	.799	.397	.784	.304	.628	.553	.464	.513	.632	.807	.913	.581	.000	.000
Return_on_Assets_pct	Pearson Correlation	-,154**	,222**	,191**	,181**	,217**	1	-,320**	.090	-.033	-.033	,322**	.057	-,201**	,138**	,196**	,152**	,176**	,223**	,228**	,110*	,234**	,275**	,250**	,255**	,289**	.033	-,321**	,419**	,389**
	Sig. (2-tailed)	.004	.000	.000	.001	.000		.000	.093	.541	.539	.000	.284	.000	.010	.000	.004	.001	.000	.000	.039	.000	.000	.000	.000	.000	.533	.000	.000	.000
Debt_to_Assets_pct	Pearson Correlation	-.036	-,144**	-,192**	-,112*	-.008	-,320**	1	-.028	-.028	-.028	-,143**	-.061	,167**	-,149**	-,137*	-.008	-.027	.007	.021	-.035	.007	-.036	-.054	-.025	-,138**	-,135*	,771**	-.144	-.173
	Sig. (2-tailed)	.506	.007	.000	.036	.874	.000		.607	.603	.607	.008	.254	.002	.005	.010	.874	.614	.895	.702	.511	.897	.496	.311	.640	.010	.012	.000	.156	.088
Debt_to_FCF	Pearson Correlation	-.057	-.026	-.063	.034	.003	.090	-.028	1	.033	.033	.064	.020	-.079	-.050	-.008	.021	.027	.029	.051	.025	.052	.067	.065	.073	.096	.022	-.038	-.108	-.067
	Sig. (2-tailed)	.290	.627	.239	.526	.953	.093	.607		.536	.537	.235	.714	.142	.351	.880	.701	.615	.583	.339	.646	.334	.215	.224	.174	.072	.676	.476	.291	.514
Debt_to_EBIT	Pearson Correlation	.020	.047	.027	-.013	.000	-.033	-.028	.033	1	1,000**	-.029	.011	-.045	.025	.049	.084	-,106*	.049	.062	.078	.055	.057	.065	.054	.018	.018	.040	-.013	-.055
	Sig. (2-tailed)	.707	.376	.611	.803	.999	.541	.603	.536		0.000	.583	.837	.405	.643	.362	.117	.047	.361	.249	.147	.301	.288	.226	.313	.735	.737	.456	.901	.590
Interest_to_EBIT	Pearson Correlation	.020	.047	.027	-.014	.000	-.033	-.028	.033	1,000**	1	-.029	.011	-.044	.024	.048	.083	-,106*	.048	.061	.077	.055	.056	.064	.053	.018	.018	.040	-.052	-.072
	Sig. (2-tailed)	.713	.378	.612	.797	.995	.539	.607	.537	0.000		.584	.837	.409	.650	.366	.121	.049	.370	.253	.149	.307	.293	.231	.319	.739	.743	.450	.612	.480
Earnings_per_share	Pearson Correlation	-,304**	,133*	,106*	,132*	.018	,322**	-,143**	.064	-.029	-.029	1	,870**	-.090	.037	.045	-.070	.011	-.061	,218**	,280**	,223**	,156**	,184**	,180**	.047	-.018	-,136*	,319**	,273**
	Sig. (2-tailed)	.000	.013	.048	.014	.731	.000	.008	.235	.583	.584		.000	.091	.496	.398	.190	.831	.253	.000	.000	.000	.003	.001	.001	.379	.733	.011	.001	.007
Revenue_per_share	Pearson Correlation	-,228**	,166**	,151**	.001	.002	.057	-.061	.020	.011	.011	,870**	1	-,116*	.028	,108*	-.043	-.071	-.050	,247**	,342**	,272**	,184**	,224**	,208**	.050	.065	.000	.038	.030
	Sig. (2-tailed)	.000	.002	.005	.989	.965	.284	.254	.714	.837	.837	.000		.030	.599	.044	.418	.187	.352	.000	.000	.000	.001	.000	.000	.347	.222	.995	.714	.771
Unemployment	Pearson Correlation	-,251**	-.093	-.105	-.091	-.047	-,201**	,167**	-.079	-.045	-.044	-.090	-,116*	1	-.059	-,789**	-,573**	-.050	-,567**	-,641**	-,603**	-,629**	-,781**	-,725**	-,748**	-,544**	-,200**	,132*	-.019	.129
	Sig. (2-tailed)	.000	.082	.050	.088	.380	.000	.002	.142	.405	.409	.091	.030		.274	.000	.000	.354	.000	.000	.000	.000	.000	.000	.000	.000	.000	.013	.856	.205
GDP_Growth	Pearson Correlation	,107*	-.055	-.049	-.010	-,113*	,138**	-,149**	-.050	.025	.024	.037	.028	-.059	1	,252**	.029	-,225**	.017	-.011	.076	.083	.103	.092	.047	,170**	.001	-,144**	,846**	,894**
	Sig. (2-tailed)	.046	.309	.358	.859	.034	.010	.005	.351	.643	.650	.496	.599	.274		.000	.582	.000	.750	.840	.156	.122	.053	.085	.383	.001	.990	.007	.000	.000
Consumer_Confidence_Indicator	Pearson Correlation	,228**	,254**	,296**	.020	.073	,196**	-,137*	-.008	.049	.048	.045	,108*	-,789**	,252**	1	,662**	-,182**	,665**	,655**	,610**	,700**	,753**	,734**	,699**	,435**	,295**	-.084	,429**	,355**
	Sig. (2-tailed)	.000	.000	.000	.706	.172	.000	.010	.880	.362	.366	.398	.044	.000	.000		.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.115	.000	.000
Res_Insolv_100	Pearson Correlation	.079	,431**	,365**	-.001	.014	,152**	-.008	.021	.084	.083	-.070	-.043	-,573**	.029	,662**	1	-,146**	,934**	,686**	,261**	,683**	,721**	,717**	,701**	,372**	,386**	,181**	.016	-.027
	Sig. (2-tailed)	.140	.000	.000	.983	.799	.004	.874	.701	.117	.121	.190	.418	.000	.582	.000		.006	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.879	.790
Starting_Bus__100	Pearson Correlation	-,210**	.041	.076	-,118*	.045	,176**	-.027	.027	-,106*	-,106*	.011	-.071	-.050	-,225**	-,182**	-,146**	1	.041	.074	-,305**	-.017	.041	.028	.053	,266**	.005	.039	-.023	-.062
	Sig. (2-tailed)	.000	.450	.155	.027	.397	.001	.614	.615	.047	.049	.831	.187	.354	.000	.001	.006		.442	.169	.000	.758	.439	.604	.323	.000	.931	.463	.819	.547
RR_100	Pearson Correlation	.047	,509**	,478**	-.065	.015	,223**	.007	.029	.049	.048	-.061	-.050	-,567**	.017	,665**	,934**	.041	1	,751**	,250**	,769**	,788**	,798**	,784**	,479**	,410**	,186**	.011	-.010
	Sig. (2-tailed)	.378	.000	.000	.224	.784	.000	.895	.583	.361	.370	.253	.352	.000	.750	.000	.000	.442		.000	.000	.000	.000	.000	.000	.000	.000	.000	.916	.925
Voice_and_Accountability	Pearson Correlation	-,131*	,533**	,448**	-.009	.055	,228**	.021	.051	.062	.061	,218**	,247**	-,641**	-.011	,655**	,686**	.074	,751**	1	,614**	,938**	,919**	,930**	,946**	,487**	,305**	,223**	.016	.019
	Sig. (2-tailed)	.014	.000	.000	.874	.304	.000	.702	.339	.249	.253	.000	.000	.000	.840	.000	.000	.169	.000		.000	.000	.000	.000	.000	.000	.000	.000	.873	.849
Political_Stability	Pearson Correlation	.018	,179**	,176**	.079	.026	,110*	-.035	.025	.078	.077	,280**	,342**	-,603**	.076	,610**	,261**	-,305**	,250**	,614**	1	,621**	,590**	,597**	,610**	,282**	-.008	-.030	-.144	-.176
	Sig. (2-tailed)	.733	.001	.001	.139	.628	.039	.511	.646	.147	.149	.000	.000	.000	.156	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000	.879	.581	.156	.084
Government_Effectiveness	Pearson Correlation	-.017	,559**	,507**	-.032	.032	,234**	.007	.052	.055	.055	,223**	,272**	-,629**	.083	,700**	,683**	-.017	,769**	,938**	,621**	1	,929**	,963**	,958**	,566**	,364**	,196**	.045	.097
	Sig. (2-tailed)	.746	.000	.000	.550	.553	.000	.897	.334	.301	.307	.000	.000	.000	.122	.000	.000	.758	.000	.000	.000		.000	.000	.000	.000	.000	.000	.657	.341
Regulatory_Quality	Pearson Correlation	.051	,432**	,364**	.012	.039	,275**	-.036	.067	.057	.056	,156**	,184**	-,781**	.103	,753**	,721**	.041	,788**	,919**	,590**	,929**	1	,963**	,972**	,604**	,287**	,121*	.078	.086

	Sig. (2-tailed)	.342	.000	.000	.822	.464	.000	.496	.215	.288	.293	.003	.001	.000	.053	.000	.000	.439	.000	.000	.000	.000		.000	.000	.000	.000	.024	.445	.398
Rule_of_Law	Pearson Correlation	.031	,481**	,441**	-.017	.035	,250**	-.054	.065	.065	.064	,184**	,224**	-,725**	.092	,734**	,717**	.028	,798**	,930**	,597**	,963**	,963**	1	,978**	,620**	,359**	,128*	.050	.065
	Sig. (2-tailed)	.562	.000	.000	.746	.513	.000	.311	.224	.226	.231	.001	.000	.000	.085	.000	.000	.604	.000	.000	.000	.000	.000		.000	.000	.000	.017	.628	.524
Control_of_Corruption	Pearson Correlation	.013	,476**	,412**	-.002	.026	,255**	-.025	.073	.054	.053	,180**	,208**	-,748**	.047	,699**	,701**	.053	,784**	,946**	,610**	,958**	,972**	,978**	1	,616**	,320**	,145**	.017	.028
	Sig. (2-tailed)	.811	.000	.000	.970	.632	.000	.640	.174	.313	.319	.001	.000	.000	.383	.000	.000	.323	.000	.000	.000	.000	.000	.000		.000	.000	.007	.869	.784
Count_Climate__100	Pearson Correlation	.067	,187**	,190**	.005	.013	,289**	-,138**	.096	.018	.018	.047	.050	-,544**	,170**	,435**	,372**	,266**	,479**	,487**	,282**	,566**	,604**	,620**	,616**	1	,137*	-.073	.094	,203*
	Sig. (2-tailed)	.212	.000	.000	.924	.807	.000	.010	.072	.735	.739	.379	.347	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000		.010	.176	.357	.044
Revenues_share	Pearson Correlation	.062	,560**	,630**	-,210**	.006	.033	-,135*	.022	.018	.018	-.018	.065	-,200**	.001	,295**	,386**	.005	,410**	,305**	-.008	,364**	,287**	,359**	,320**	,137*	1	,187**	.001	.004
	Sig. (2-tailed)	.249	.000	.000	.000	.913	.533	.012	.676	.737	.743	.733	.222	.000	.990	.000	.000	.931	.000	.000	.879	.000	.000	.000	.000	.010		.000	.991	.971
Liabilities_to_Assets	Pearson Correlation	-.097	,212**	,166**	-,284**	-.030	-,321**	,771**	-.038	.040	.040	-,136*	.000	,132*	-,144**	-.084	,181**	.039	,186**	,223**	-.030	,196**	,121*	,128*	,145**	-.073	,187**	1	-.062	-.071
	Sig. (2-tailed)	.071	.000	.002	.000	.581	.000	.000	.476	.456	.450	.011	.995	.013	.007	.115	.001	.463	.000	.000	.581	.000	.024	.017	.007	.176	.000		.543	.487
Tourist_arrivals_EU	Pearson Correlation	.[c]	.096	.032	.016	,456**	,419**	-.144	-.108	-.013	-.052	,319**	.038	-.019	,846**	,429**	.016	-.023	.011	.016	-.144	.045	.078	.050	.017	.094	.001	-.062	1	,914**
	Sig. (2-tailed)	0.000	.347	.757	.878	.000	.000	.156	.291	.901	.612	.001	.714	.856	.000	.000	.879	.819	.916	.873	.156	.657	.445	.628	.869	.357	.991	.543		.000
Tourist_arrivals_EU_chg	Pearson Correlation	.[c]	.103	.032	-.049	,425**	,389**	-.173	-.067	-.055	-.072	,273**	.030	.129	,894**	,355**	-.027	-.062	-.010	.019	-.176	.097	.086	.065	.028	,203*	.004	-.071	,914**	1
	Sig. (2-tailed)	0.000	.314	.752	.629	.000	.000	.088	.514	.590	.480	.007	.771	.205	.000	.000	.790	.547	.925	.849	.084	.341	.398	.524	.784	.044	.971	.487	.000	

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

c. Cannot be computed because at least one of the variables is constant.

The table also includes the significance of the correlation coefficients, indicating the probability of obtaining the observed correlation coefficient if the true correlation coefficient were zero. A significance level is commonly used to determine whether a correlation coefficient is statistically significant. If there is significance, it can be concluded that the correlation is statistically significant, meaning that it is unlikely to have occurred by chance. Based on the coefficients, it is possible to observe the following patterns:

LogRevenues and LogEmployees have a strong positive correlation with each other. This suggests that as revenues increase, the number of employees tends to increase as well.

LogRevenues has a moderate correlation with most of the financial ratios such as Current_ratio, EBIT_margin_pct, and Debt_to_FCF, which indicates that as the revenues of the company increase, so does its financial health.

Return_on_Assets_pct has a moderate positive correlation with EBIT_margin_pct, indicating that companies with higher EBIT margins tend to have higher returns on their assets.

There is a weak negative correlation between Sector and LogRevenues and between Sector and LogEmployees. This suggests that different sectors may have varying levels of revenues and employees.

Unemployment has a weak negative correlation with Res_Insolv_100, indicating that as unemployment decreases, the number of business insolvencies tends to decrease as well.

GDP_Growth has a weak positive correlation with most of the financial ratios and revenue measures, which indicates that as the GDP grows, so does the financial health of the companies.

The indicators of good governance such as Voice_and_Accountability, Political_Stability, Government_Effectiveness, Regulatory_Quality, Rule_of_Law, and Control_of_Corruption have weak to moderate positive correlations with most of the financial ratios and revenue measures. This suggests that better governance is associated with better financial health of companies.

In general, a high correlation between two variables could indicate that they are measuring similar aspects of the phenomenon being studied, and this could lead to multicollinearity issues in a regression model. It is possible that in the specific context of the ESG risk rating model being developed, a correlation higher than 0.6 between certain variables could be problematic due to concerns about multicollinearity or overfitting the model.

Descriptive statistics and explanatory analysis gave a first impression of the data final structure in order to decide the more effective methodologies for the empirical analysis.

5.5 Empirical Results

The first method chosen is binary logistic regression which is used for classification problems when the output or dependent variable is a dichotomous categorical variable. That's because provides useful insights of how relevant an independent variable is (i.e., the (coefficient size), but also the direction of the relationship. It also works well for cases where the dataset is linearly separable. Moreover, it is much easier to implement than other methods, especially in the context of machine learning. The only case should be mentioned is that it may not be accurate if the sample size is too small.

The second method chosen is probit regression. The advantage is that it overcomes the challenges of a logistic regression model. The dataset is derived from actual financial data, and the values of each variable are subject to volatility. The probit regression model can provide some degree of protection against the influence of outliers using the probit link function. However, instead of modelling the natural logarithm of the odds ratio like in logistic regression, probit regression models the cumulative distribution function of a standard normal random variable. The predicted probabilities from probit regression are always between 0 and 1, and the probate incorporates non-linear effects of X, as well. Furthermore, a probit model can also be used to understand the conditions that lead to the outcome variable being close to zero or close to one, even if the outcome variable is a continuous variable between zero and one. As expected, it uses the Maximum Likelihood Estimator (MLE) due to the probit is not linear in the parameters bj. However, a potential disadvantage is that the coefficients are difficult to interpret.

The collected data consists of records arranged in a relevant chronological sequence, which led to the selection of panel data regression as the third method. Panel data regression combines cross-sectional data and time series data, where measurements for the same unit (in this case, each firm) are taken at different times, specifically from 2015 to 2021.

The following section presents the results obtained from each of the methods used in this study.

5.5.1 Logistic Regression

5.5.1.1 Model Selection

Initially, an attempt was made to create a logit model for each sector. However, due to the inadequate dataset size and a low number of ESG incidents, it was not possible to establish the relationship between the variables and ESG incidents. Additionally, the independent variables remained the same across all sectors except for the Hospitality sector. In this sector, it was possible to include two additional variables related to Tourism_aririvals_EU and Tourism_aririvals_EU_chg. The variables and their statistical significance are displayed on Tables 18.

Table 18: Variables in the Equation by Sector

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Sector 1	Political_Stability	1,668	,848	3,866	1	,049	5,299
	Revenues_share	,699	,175	16,025	1	,000	2,011
	Constant	-3,425	,943	13,189	1	,000	,033
Sector 2	GDP_Growth	-,392	,122	10,399	1	,001	,676
	Consumer_Confidence_Indicator	,173	,062	7,794	1	,005	1,189
	Constant	1,151	,682	2,846	1	,092	3,160
Sector 3	GDP_Growth	-,336	,100	11,194	1	,001	,714
	Revenues_share	,364	,096	14,393	1	,000	1,439
	Constant	-3,506	,711	24,301	1	,000	,030
Sector 4	Revenue_per_share	,024	,008	10,493	1	,001	1,025
	Liabilities_to_Assets	,041	,013	9,261	1	,002	1,042
	Constant	-5,965	1,338	19,881	1	,000	,003

This is a table showing the results of regression analyses for the four different sectors. For Sector 1, the independent variables are Political_Stability and Revenues_share, while for Sector 2, they are GDP_Growth and Consumer_Confidence_Indicator. For Sector 3, the independent variables are GDP_Growth and Revenues_share, and for Sector 4, they are Revenue_per_share and Liabilities_to_Assets. The variables Tourism_aririvals_EU and Tourism_aririvals_EU_chg were not considered on Sector 4.

The results suggest that the significance of each variable varies across sectors. For example, in Sector 1, both Political_Stability and Revenues_share are statistically significant, while in Sector 2, only GDP_Growth and Consumer_Confidence_Indicator are significant. The table also provides information on the strength and direction of the relationship between the independent variables and the dependent variable, as indicated by the beta coefficients.

Table 19 displays the classification table which can be used to evaluate the practical performance of the models.

Table 19: Classification Table by Sectore

Classification Table[a]

			Predicted		
			ESG		Percentage
	Observed		Non-Default	Default	Correct
Sector 1	ESG	Non-Default	40	4	90,9
		Default	8	18	69,2
	Overall Percentage				82,9
Sector 2	ESG	Non-Default	69	4	94,5
		Default	17	8	32,0
	Overall Percentage				78,6
Sector 3	ESG	Non-Default	67	2	97,1
		Default	7	8	53,3
	Overall Percentage				89,3
Sector 4	ESG	Non-Default	84	4	95,5
		Default	4	6	60,0
	Overall Percentage				91,8

a. The cut value is ,500

Table 20 shows that the Cox and Snell R-Square and Nagelkerke R-Square values are not close to value 1, but it is better to check other tests to argued if the models have the capacity to predict the dependent variable by examining the relationship between one or more independent variables that already exist.

Table 20: Model Performance by Sector

Model Summary

	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
Sector 1	52,102	,437	,597
Sector 2	79,202	,279	,411
Sector 3	48,016	,307	,504
Sector 4	39,590	,225	,466

Conversely, a significant result the Hosmer and Lemeshow Test suggests that the model is not suitable, whereas a non-significant test suggests a good fit. In Table 21, the values > 0.557 for Sectors indicate non-significant tests. A high value of Chi-squared with a small p-value (< 0.05) suggests a poor fit, while small Chi-squared values with larger p-values closer to 1 indicate a good logistic regression model fit.

Table 21: Hosmer and Lemeshow Test by Sector

Hosmer and Lemeshow Test

	Chi-square	df	Sig.
Sector 1	6,810	8	,557
Sector 2	3,039	8	,932
Sector 3	8,934	8	,348
Sector 4	3,700	8	,883

According to the Table 22, the correlation between the variables is as follows:
Sector 1: Revenues_share, Political_Stability (positive correlation 0.476)

Sector 2: GDP_Growth, Cinsumer_Confidence_Indicator (negative correlation -0.855)

Sector 3: GDP_Growth, Revenues_share (negative correlation -0.538)

Sector 4: Revenue_per_share, Liability_to_Assets (positive correlation 0.148)

Table 22: Correlations of variables in the equations by Sector

Correlations

		Political_Stability	Revenues_share	GDP_Growth	Consumer_Confidence_Indicator	Revenue_per_share	Liabilities_to_Assets
Political_Stability	Pearson Correlation	1	,020	-,164	,071	-,192	,161
	Sig. (2-tailed)		,848	,106	,490	,059	,114
Revenues_share	Pearson Correlation	,020	1	-,106	,229	,799	,379
	Sig. (2-tailed)	,848		,298	,023	,000	,000
GDP_Growth	Pearson Correlation	-,164	-,106	1	,410	-,084	-,271
	Sig. (2-tailed)	,106	,298		,000	,409	,007
Consumer_Confidence_Indicator	Pearson Correlation	,071	,229	,410	1	,027	,237
	Sig. (2-tailed)	,490	,023	,000		,793	,019
Revenue_per_share	Pearson Correlation	-,192	,799	-,084	,027	1	,286
	Sig. (2-tailed)	,059	,000	,409	,793		,004
Liabilities_to_Assets	Pearson Correlation	,161	,379	-,271	,237	,286	1
	Sig. (2-tailed)	,114	,000	,007	,019	,004	

The outcomes obtained from the sector-specific models cannot be generalized to the overall model results for all sectors due to the limited dataset size and fewer incidents. In the models by sector, found no more than two variables that were statistically significant, and in some cases, these variables had a high correlation between them. In this case such a sector analysis would not produce valid results.

After considering the correlation between independent variables, the logit model was attempted by gradually adding an independent variable and examining the likelihood ratio and p-value. The log-likelihood ratio test is used to compare two nested models, one with fewer predictors and the other with more predictors. If the difference in log-likelihood values between the two models is statistically significant, it suggests that the model with more predictors provides a better fit to the data. In logistic regression, a rule of thumb is that a difference of 3.84 or more in the log-likelihood values between two nested models is statistically significant at the 0.05 level. However, the significance level can vary depending on the degrees of freedom and the sample size.

Then, the forward LR method of SPSS was used, which is a stepwise regression approach that starts from the null model and adds a variable that improves the model the most, one at a time, until the stopping criterion is met. The resulting model was similar to the one created manually but with an even better likelihood ratio. The final model included the variables Sector, GDP_Growth, LogEmployees, and the other two variables Res_Insolv_100 and RR_100, which were highly correlated variables. The results of five model iterations are presented in Table 23, along with the statistical tests and results in Tables 24 and 25.

Table 23: Logit Model Variables in the Equation (initial)

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 5	Sector	-,625	,166	14,243	1	,000	,535
	LogEmployees	,583	,112	27,125	1	,000	1,792
	GDP_Growth	-,202	,046	19,369	1	,000	,817
	Res_Insolv_100	-,074	,034	4,666	1	,031	,929
	RR_100	,075	,023	10,895	1	,001	1,078
	Constant	-5,110	1,666	9,410	1	,002	,006

Table 24: Logit Model Performance (initial)

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
5	232,580[a]	,318	,489

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Table 25: Logit Model Classification Table (initial)

Classification Table[a]

Observed			Predicted: ESG Non-Default	Predicted: ESG Default	Percentage Correct
Model 1	ESG	Non-Default	257	17	93,8
		Default	39	37	48,7
	Overall Percentage				84,0

a. The cut value is .500

As a final step, it was important to reassess the correlation between the variables that were identified as statistically significant and remove any highly correlated variables. The findings are presented in Table 26, which clearly indicates that two variables have a strong positive correlation.

Table 26: Logit Model Variables Correlations

Correlations

		Sector	GDP_Growth	Res_Insolv_100	RR_100	LogEmployees
Sector	Pearson Correlation	1	,107*	,079	,047	-,207**
	Sig. (2-tailed)		,046	,140	,378	,000
GDP_Growth	Pearson Correlation	,107*	1	,029	,017	-,049
	Sig. (2-tailed)	,046		,582	,750	,358
Res_Insolv_100	Pearson Correlation	,079	,029	1	,934**	,365**
	Sig. (2-tailed)	,140	,582		,000	,000
RR_100	Pearson Correlation	,047	,017	,934**	1	,478**
	Sig. (2-tailed)	,378	,750	,000		,000
LogEmployees	Pearson Correlation	-,207**	-,049	,365**	,478**	1
	Sig. (2-tailed)	,000	,358	,000	,000	

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

As anticipated, this necessitates running the model again by incorporating only one of these two variables at a time and evaluating the final models based not only on the likelihood ratio but also their discriminatory power to identify the best model among all the inferior ones. Both models showed similar performance, something that was obviously expected given the high correlation between the variables Res_Insolv_100 and RR_100. In this instance, the model with the best discriminatory power was retained. It is now critical to elucidate the statistical tests and their significance for the resulting model.

-2 Log likelihood (-2LL) is a measure of how well the estimated model fits the likelihood. The measure -2LL is the best possible based on the previous explained steps. A good model is one that results in a high likelihood of the observed results. This translates to a small number for -2LL. If a model fits perfectly, the likelihood is 1, and -2 times the log likelihood is 0.

Cox and Snell R Square is an alternative index of goodness of fit related to the R^2 value from linear regression. It is based on the log likelihood for the model compared to the log likelihood for a baseline model. However, with categorical outcomes, it has a theoretical maximum value of less than 1, even for a 'perfect' model.

To evaluate the goodness of fit of the logistic regression model, Nagelkerke R square is calculated from the result of generalized linear models. It represents the power of explanation of the model. Nagelkerke R Square is an adjusted version of Cox and Snell R Square. The values fall between 0 and 1. It measures the proportion of the total variation of the dependent variable that can be explained by the independent variables in the current model.

Table 27 shows that while the Cox and Snell R-Square and Nagelkerke R-Square values are not close to value 1, it can still be argued that a model has the capacity to predict the dependent variable by examining the relationship between one or more independent variables that already exist. Comparing the two models, it seems that Model 1 has better fit than Model 2.

Table 27: Logit Models Performance

Model Summary

Models	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	237,185[a]	,308	,475
2	247,165[a]	,288	,445

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than ,001.

Conversely, a significant result the Hosmer and Lemeshow Test suggests that the model is not suitable, whereas a non-significant test suggests a good fit. In Table 28, the value of 0.559 of Model 1 indicates a non-significant test. A high value of Chi-squared with a small p-value (< 0.05) suggests a poor fit, while small Chi-squared values with larger p-values closer to 1 indicate a good logistic regression model fit.

Table 28: Logit Models Hosmer and Lemeshow Test

Hosmer and Lemeshow

Models	Chi-square	df	Sig.
1	6,792	8	,559
2	8,812	8	,358

The logistic regression model's practical performance can be evaluated through the classification table (contingency table, confusion matrix). For each case, the predicted response category is determined by selecting the category with the highest model-predicted probability. It should be noted that the Classification cut-off value for this model is 0.5, which is strict as indicated in Table 29. The correct predictions are shown in cells on the left diagonal of the table and the overall percentage is 83,4%.

Table 29: Logit Modesl Classification Table

Classification Table[a]

			Predicted		
			ESG		
	Observed		Non-Default	Default	Percentage Correct
Model1	ESG	Non-Default	257	17	93,8
		Default	41	35	46,1
	Overall Percentage				83,4
Model2	ESG	Non-Default	255	19	93,8
		Default	43	33	46,1
	Overall Percentage				82,3

a. The cut value is ,500

Table 30 presents coefficients and corresponding p-values (Sig.) obtained from the analysis. These two factors are essential in determining which relationships within a model are statistically significant and the nature of those relationships. The p-values associated with the coefficients help to indicate whether these relationships are statistically significant or not, while the regression coefficients describe the mathematical relationship between each independent variable and the dependent variable.

In general, a higher Wald statistic indicates a stronger relationship between the predictor variable and the outcome variable, and if the p-value associated with the Wald statistic is below the chosen level of significance (e.g., 0.05), then the null hypothesis that the coefficient for the predictor is zero can be rejected. This leads to the conclusion that the predictor is statistically significant in predicting the outcome variable.

Furthermore, the standard error (SE) is a measure of the variability or spread of a sampling distribution. It is the standard deviation of the sample mean's distribution, and it measures the amount of variation or dispersion of sample means around the population mean. The standard error is an important concept in statistical inference, as it is used to calculate confidence intervals and test hypotheses about population parameters based on sample statistics.

Table 30: Logit Models, Variables in the Equation

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Model1	Sector	-,650	,165	15,616	1	,000	,522
	LogEmployees	,601	,114	27,988	1	,000	1,824
	GDP_Growth	-,199	,045	19,244	1	,000	,820
	RR_100	,032	,011	9,007	1	,003	1,033
	Constant	-7,690	1,213	40,214	1	,000	,000
Model2	Sector	-,604	,163	13,730	1	,000	,547
	LogEmployees	,733	,106	48,114	1	,000	2,081
	GDP_Growth	-,185	,043	18,959	1	,000	,831
	Constant	-6,840	1,075	40,518	1	,000	,001

The results demonstrate that Res_Insolv_100 is not statistically significant in Model 2 and is therefore not included in the equation on the third step of the Forward LR method. One the other hand 100 is statistically significant in Model1 and the statistical test showed that it has better fit.

Thus, the Model1 equation with the best fit includes the variables Sector, LogEmployees, GDP_Growth, RR_100 and presented as:

$$Z = a + bX1 + cX2 + dX3 + eX4$$

is becoming,

$$Z = -7.690 + (-0.650 * Sector) + 0.601 * LogEmployees + (-0.199 * GDP\_Growth) + 0.032 * RR\_100$$

And the probability is calculated as:

$$P = exp\,(a + bX1 + cX2 + dX3 + eX4)\;/\;[\,1 + exp\,(a + bX1 + cX2 + dX3 + eX4)]$$
$$= exp\,(Z)\;/\;[\,1 + exp\,(Z)]$$

A given negative coefficient means that for an increase in the predictor X_i there is a decrease in the predicted probability. On the other hand, a given positive coefficient means that for an increase in the predictor, an increase follows in the predicted probability. A high value of the coefficient means that the independent variable has a very strong influence on the probability that the event occurs or does not occur, while a low value indicates a small influence of the independent variable on the probability of the corresponding outcome.

Table 31 contains some of the misclassified cases. It's easy to calculate the misclassification rate as the number of total incorrect predictions divided by the total number of predictions. The misclassification rate for this model is 16.6%.

Table 31: Logit Model 1, Casewise List

Casewise List[b]

Case	Selected Status[a]	Observed ESG	Predicted	Predicted Group	Temporary Variable Resid	Temporary Variable Zresid
60	S	1**	,076	0	,924	3,490
62	S	1**	,089	0	,911	3,195
69	S	1**	,069	0	,931	3,680
106	S	1**	,108	0	,892	2,877
146	S	1**	,122	0	,878	2,683
176	S	1**	,034	0	,966	5,299
192	S	0**	,949	1	-,949	-4,313
323	S	0**	,924	1	-,924	-3,476

a. S = Selected, U = Unselected cases, and ** = Misclassified cases.
b. Cases with studentized residuals greater than 2,000 are listed.

5.5.1.2 ROC Analysis

Figure 14 presents a visual representation of the ROC curves for the different models that were tested in this analysis. It allows for a comparison of the AUC values, which are presented in the legend. The progression of model selection is also depicted, illustrating how the addition of variables improved the accuracy of the model. Overall, ROC analysis provides a useful tool for evaluating and comparing the performance of statistical models.

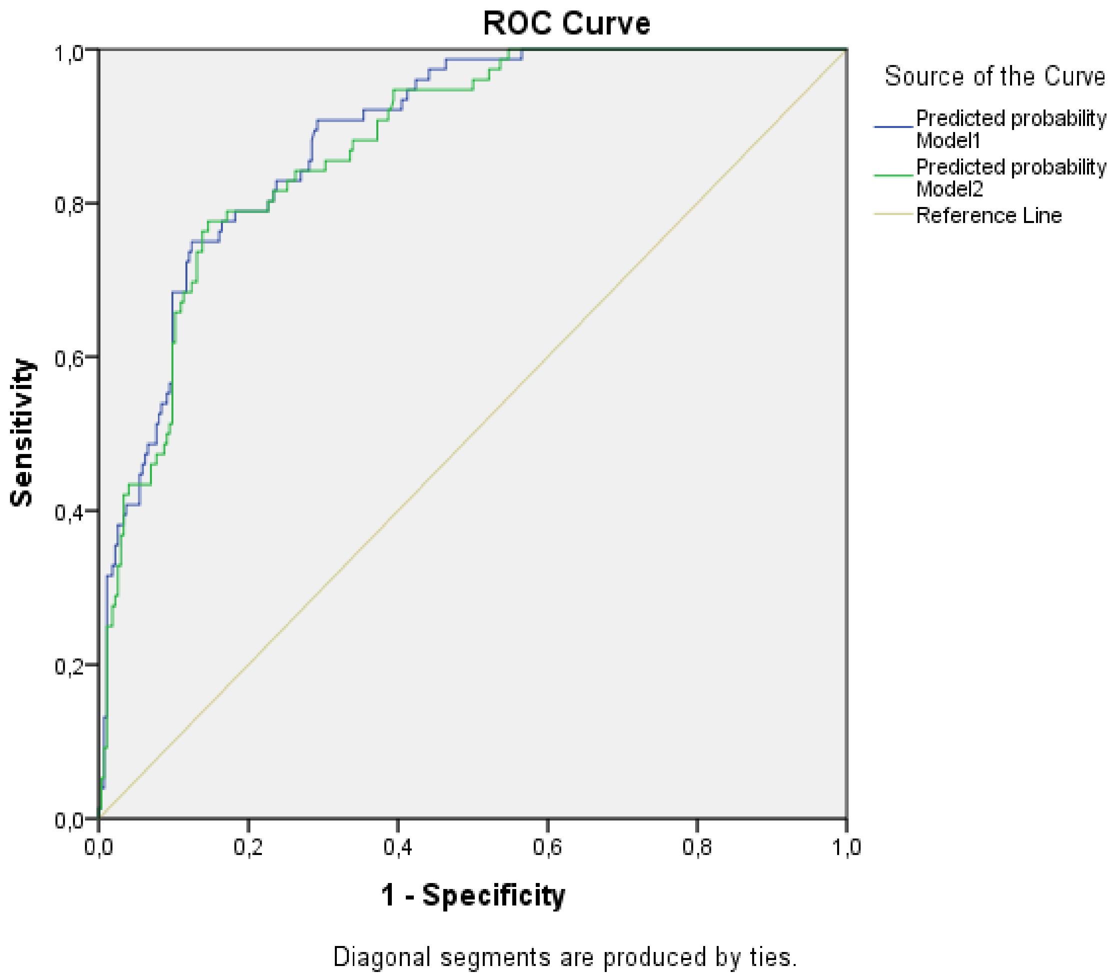

Figure 13: Logit Model, ROC Curve

The Area Under the Curve (AUC) is an overall summary of diagnostic accuracy of the model. AUC equals 0,5 is the null hypothesis when the ROC curve corresponds to random chance. AUC equals 1 for perfect accuracy. As it is shown to the Table 32 the AUC for the specific models is > 0,5 and very close to 1, so this means that the models have descent discriminatory power. On rare occasions, the estimated AUC is < 0,5, and in this case, it is indicating that the model test does worse than chance. Comparing the two models, it is obvious that Model1 has better discriminatory power than Model2.

Table 32: Logit Model, Area under ROC

Area Under the Curve

Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Predicted probability Model1	,886	,019	,000	,848	,923
Predicted probability Model2	,872	,021	,000	,830	,914

The test result variable(s): Predicted probability Model1, Predicted probability Model2 has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption
b. Null hypothesis: true area = 0.5

Furthermore, once the AUC is known, it is possible to calculate the Accuracy Ratio (AR) for Model1. Result comes out by calculating area under the prediction model and random model (aR) divided by calculating area under the perfect model and the random model (aP). An AR closer to 1, means better model. It is also proven that there is another formula to calculate accuracy ratio.

$$Accuracy\ Rate\ (AR)\ =\ 2\ *\ AUC\ -1$$

Using this formula for Model1:

$$AR\ =\ 2 * 0.886 - 1\ =\ 0.772$$

5.5.2 Probit Regression

5.5.2.1 Model Selection

The process for Probit regression was similar to that of the Logit model, but the conclusions were somewhat different. Initially, the model was selected manually by observing the likelihood ratio and testing the performance of the model using correlated variables. Then, the automated selection of statistically significant variables was checked and compared to those selected manually. The results showed that the same variables selected in the Logit model were also significant in the Probit model.

At this point it is important to say that -2LL was not the lowest of all models runs. The best result comes from a mix of both -2LL and Chi-Square as the likelihood-ratio chi-square statistic (G^2) is based on the ratio of the observed to the expected frequencies. The estimations are detailed in Tables 33.

Table 33: Probit Model, Fitting Information

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	364,901			
Final	235,463	129,438	4	,000

Link function: Probit.

Pearson's chi-square statistic test is applied to categorical data to express goodness of fit. It essentially determines if your data is significantly different than expected. Using the null hypothesis that there are no differences between the classes in the population, and the p-value is close to 1, it would be assuming the null hypothesis is true. The null hypothesis for Pearson's chi-square test is that there is no significant difference between the observed and expected frequencies in the categorical data. In other words, the observed data follows the expected pattern and any deviation from this pattern is due to random chance. The estimations are detailed in Tables 34.

Table 34: Probit Model, Goodness-of-Fit

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	264,205	343	,999
Deviance	234,077	343	1,000

Link function: Probit.

Table 35 presents the coefficients and p-values in the regression analysis for probit regression. The coefficients and p-values work together to indicate which relationships in the model are statistically significant and their nature. The p-values show whether these relationships are statistically significant, and the coefficients describe the mathematical relationship between each independent variable and the dependent variable. In the probit regression model, the coefficients show the change in the z-score or probit index for a one-unit change in the predictor.

$$E(Y|X) = P(Y = 1|X) = \Phi(a + bX1 + cX2 + dX3 + eX4)$$

Table 35: Probit Model, Variables Significance

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Threshold	[ESG = 0]	4,332	,644	45,314	1	,000	3,071	5,594
Location	Sector	-,361	,092	15,306	1	,000	-,542	-,180
	LogEmployees	,332	,063	28,161	1	,000	,209	,454
	GDP_Growth	-,113	,025	20,240	1	,000	-,162	-,064
	RR_100	,019	,006	10,101	1	,001	,007	,030

Link function: Probit.

The table of coefficients shows the results of a probit regression model that is used to estimate the probability of an outcome variable being greater than or equal to a certain threshold value, given values of the independent variables.

The first row of the table shows the coefficient estimate for the threshold variable, which represents the constant term in the model. In this case, the estimated coefficient is 4.332 with a standard error of 0.644. This means that when all other independent variables are equal to zero, the estimated probability of the outcome variable being greater than or equal to the threshold value is very high, since the coefficient is positive and statistically significant at a high level of confidence (p-value less than 0.001).

The other rows of the table show the coefficient estimates for each of the independent variables in the model, along with their standard errors, Wald test statistics, degrees of freedom, and p-values. The negative coefficient estimate for the Sector variable suggests that companies in certain sectors are associated with lower probabilities of the outcome variable being close to one. The positive coefficient estimate for the LogEmployees variable suggests that larger companies are associated with higher probabilities of the outcome variable being close to one. The negative coefficient estimates for the GDP_Growth variable suggests that lower GDP growth rates are associated with higher probabilities of the outcome variable being close to one. Finally, the positive coefficient estimates for the RR_100 variable suggests that higher levels of regulatory compliance are associated with higher probabilities of the outcome variable being close to one.

Overall, this probit model can be used to understand the factors that influence the probability of the outcome variable being close to one, and how different independent variables are related to this probability. The coefficients can be interpreted in terms of their signs and magnitudes, and their statistical significance can be used to assess the strength of the relationship between each independent variable and the outcome variable.

5.5.2.2 ROC Analysis

Even though previous statistical tests may have provided a clear answer regarding the goodness of fit of the model, the ROC curve, AUC, and AR provide another way to evaluate a model's performance. Figure 15 presents the different model performances according to the explained process of model selection.

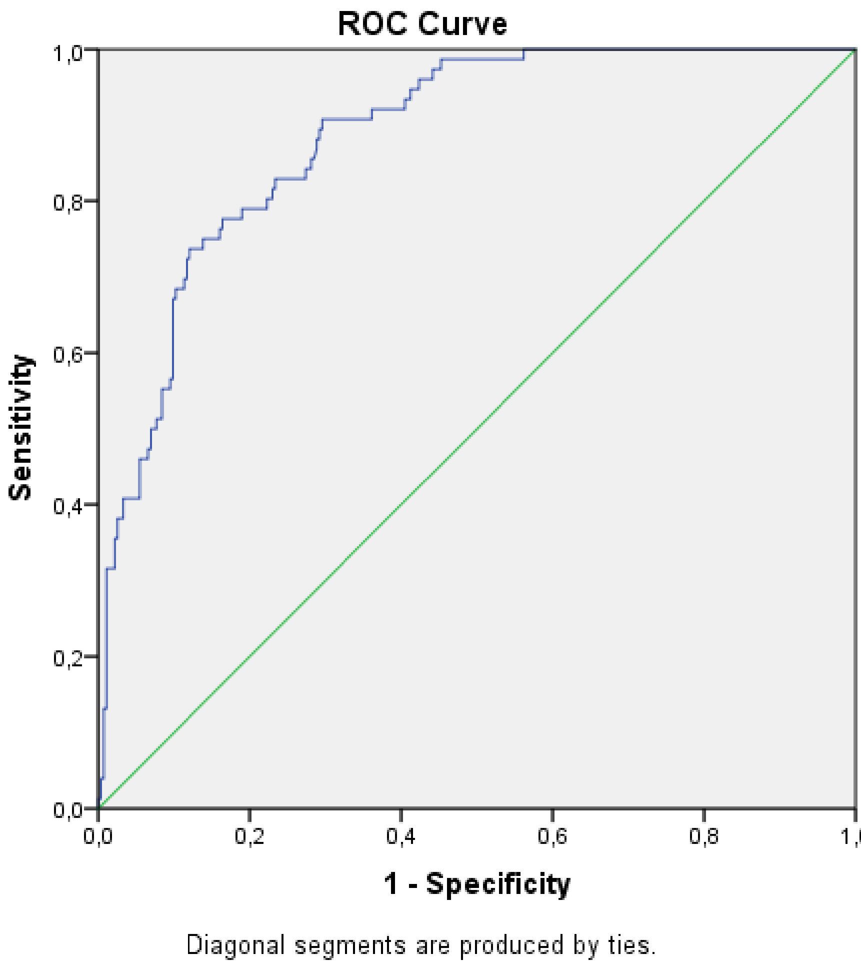

Figure 14: Probit Model, ROC Curve

This Model is finally the one that has the best fit comparing to others by selecting the same independent variables as logit model. The difference in the overall results of the probit and logit models is usually slight to non-existent. Furthermore, Table 36 presents the Area Under the Curve for the model's estimations where it is obvious that logit and probit results are similar.

Using the formula of AR as a linear transformation of the AUROC, it is possible to calculate the accuracy ratio for probit model 3 as the one with the best performance:

$$AR = 2 * 0.885 - 1 = 0.77$$

Table 36: Probit Model, Area under ROC

Area Under the Curve

Test Result Variable(s) Estimated Cell Probability for Response Category: 1	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Model	,885	,019	,000	,848	,923

The test result variable(s): Model has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption
b. Null hypothesis: true area = 0.5

5.4.3 Panel Data Regression

As mentioned earlier, the data consisted of records in relative chronological order and lead to the use of panel data regression fixed effects models. Panel data regression is a method that considers time and CompanyID variable as fixed factor for the estimation. Based on the data structure explained, it appears that no further restructuring is necessary for conducting panel regression analysis. Various attempts were made to estimate a panel data model using pseudo-variables relating to firms, as well as a univariate model, which yielded similar results as anticipated.

The analysis on Table 37 shows only one independent variable as statistically significant, called GDB Growth with a negative relationship with the ESG dependent variable.

Table 37: Panel Data Regression

Tests of Between-Subjects Effects

Dependent Variable: ESG

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	26,479[a]	52	,509	4,580	,000	,445
Intercept	,096	1	,096	,863	,354	,003
Sector	,000	0	.	.	.	,000
LogEmployees	,314	1	,314	2,823	,094	,009
GDP_Growth	3,021	1	3,021	27,179	,000	,084
RR_100	,049	1	,049	,441	,507	,001
CompanyID	11,027	48	,230	2,067	,000	,250
Error	33,018	297	,111			
Total	76,000	350				
Corrected Total	59,497	349				

a. R Squared = ,445 (Adjusted R Squared = ,348)

The table shows that the model is statistically significant (F=4.580, p<0.001), which means that the independent variables together have a significant effect on the dependent variable. The R-squared value of 0.445 indicates that the model explains 44.5% of the variance in ESG, and the adjusted R-squared of 0.348 suggests that the model may not fit the data well after accounting for the number of independent variables included.

Table provides information on the overall fit of the model. The Corrected Model row indicates that the model explains a significant amount of the variance in ESG, with an F-value of 4.580 and a corresponding p-value of 0.000. The Partial Eta Squared column indicates the effect size of each independent variable on the dependent variable, which ranges from 0 (no effect) to 1 (complete effect). For instance, CompanyID has the largest effect size of 0.250, indicating that it has a relatively strong effect on ESG.

The next section shows the results for each independent variable. The Type III Sum of Squares column indicates how much of the variance in ESG is explained by each independent variable, after controlling for the other variables in the model. The df column shows the degrees of freedom for each variable, which is the number of observations minus the number of parameters estimated. The Mean Square column is the Type III Sum of Squares divided by the df, which represents the variance of the estimate. The F column provides the F-value for each variable, which tests the null hypothesis that the variable has no effect on ESG. The Sig. column shows the corresponding p-value, and all variables except "Sector" have a p-value below 0.05, indicating that they are statistically significant predictors of ESG.

Table 38 presents the parameter estimates from the panel regression with ESG as the dependent variable and several independent variables including sector, LogEmployees, GDP_Growth, RR_100, and several dummy variables for different company IDs. The table includes the estimated values of the parameters, their standard errors, t-values, significance levels, and confidence intervals, as well as partial eta-squared values.

Table 38: Parameter Estimates

Parameter Estimates

Dependent Variable: ESG

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound	Partial Eta Squared
Intercept	-,639	,331	-1,932	,054	-1,290	,012	,012
Sector	,085	,339	,249	,803	-,583	,753	,000
LogEmployees	,073	,043	1,680	,094	-,012	,158	,009
GDP_Growth	-,026	,005	-5,213	,000	-,036	-,016	,084
RR_100	,003	,005	,664	,507	-,007	,013	,001
[CompanyID=1]	-,328	,830	-,395	,693	-1,961	1,306	,001
[CompanyID=2]	-,693	,773	-,897	,370	-2,215	,828	,003
[CompanyID=3]	-,265	,982	-,270	,787	-2,198	1,668	,000
[CompanyID=4]	,145	1,058	,137	,891	-1,937	2,227	,000
[CompanyID=5]	-,388	,790	-,492	,623	-1,942	1,166	,001

[CompanyID=6]	,181	1,068	,170	,865	-1,921	2,283	,000
[CompanyID=7]	-,271	,860	-,315	,753	-1,963	1,421	,000
[CompanyID=8]	-,090	1,001	-,090	,929	-2,060	1,881	,000
[CompanyID=9]	-,207	,740	-,280	,780	-1,663	1,249	,000
[CompanyID=10]	-,459	,745	-,617	,538	-1,926	1,007	,001
[CompanyID=11]	-,307	,776	-,396	,692	-1,835	1,220	,001
[CompanyID=12]	-,614	,741	-,829	,408	-2,073	,844	,002
[CompanyID=13]	-,349	,969	-,360	,719	-2,256	1,559	,000
[CompanyID=14]	-,609	,726	-,839	,402	-2,037	,819	,002
[CompanyID=15]	,037	,593	,063	,950	-1,129	1,204	,000
[CompanyID=16]	-,249	,571	-,436	,663	-1,373	,875	,001
[CompanyID=17]	-,372	,410	-,908	,365	-1,180	,435	,003
[CompanyID=18]	-,591	,428	-1,379	,169	-1,434	,252	,006
[CompanyID=19]	-,200	,505	-,396	,692	-1,194	,794	,001
[CompanyID=20]	-,006	,467	-,012	,990	-,925	,913	,000
[CompanyID=21]	-,173	,454	-,380	,704	-1,065	,720	,000
[CompanyID=22]	-,301	,635	-,475	,635	-1,550	,948	,001
[CompanyID=23]	-,453	,622	-,727	,468	-1,678	,772	,002
[CompanyID=24]	-,476	,427	-1,114	,266	-1,317	,365	,004
[CompanyID=25]	-,336	,412	-,817	,415	-1,147	,474	,002
[CompanyID=26]	,005	,661	,008	,994	-1,296	1,307	,000
[CompanyID=27]	-,183	,181	-1,014	,311	-,538	,172	,003
[CompanyID=28]	-,361	,249	-1,446	,149	-,851	,130	,007
[CompanyID=29]	,046	,210	,218	,827	-,367	,458	,000
[CompanyID=30]	-,204	,325	-,628	,531	-,843	,435	,001
[CompanyID=31]	,023	,192	,117	,907	-,355	,401	,000
[CompanyID=32]	,186	,418	,445	,657	-,637	1,008	,001
[CompanyID=33]	-,338	,225	-1,502	,134	-,781	,105	,008
[CompanyID=34]	-,147	,324	-,455	,649	-,784	,490	,001
[CompanyID=35]	-,334	,243	-1,374	,170	-,811	,144	,006
[CompanyID=36]	-,080	,296	-,269	,788	-,662	,502	,000
[CompanyID=37]	-,015	,228	-,067	,947	-,463	,433	,000
[CompanyID=38]	-,229	,324	-,706	,481	-,866	,408	,002
[CompanyID=39]	-,092	,338	-,272	,786	-,757	,574	,000
[CompanyID=40]	0[a]	.	.	.	.	.	.
[CompanyID=41]	,042	,296	,143	,886	-,541	,626	,000
[CompanyID=42]	,399	,420	,952	,342	-,427	1,226	,003
[CompanyID=43]	,045	,240	,188	,851	-,427	,517	,000
[CompanyID=44]	,439	,399	1,101	,272	-,346	1,223	,004
[CompanyID=45]	,052	,356	,145	,885	-,649	,752	,000
[CompanyID=46]	,009	,255	,037	,971	-,493	,512	,000

[CompanyID=47]	-,195	,373	-,524	,601	-,928	,538	,001
[CompanyID=48]	,066	,182	,361	,718	-,293	,425	,000
[CompanyID=49]	-,033	,179	-,185	,853	-,386	,320	,000
[CompanyID=50]	0[a]	.	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

The intercept has a negative value of -0.639, indicating that the average ESG score across all companies is lower than the reference value (which is not specified in the table).

The parameter estimate for sector is positive (0.085) but not statistically significant ($p=0.249$), suggesting that there is no significant difference in ESG scores across different sectors.

The LogEmployees variable has a positive coefficient of 0.073, indicating that larger companies tend to have higher ESG scores, although the effect is only marginally significant ($p=0.094$).

The GDP_Growth variable has a negative coefficient of -0.026, which is statistically significant ($p<0.001$) and suggests that ESG scores tend to be lower in periods of higher GDP growth.

The RR_100 variable has a small and statistically non-significant coefficient of 0.003, suggesting that there is no clear relationship between ESG scores, and the number of regulatory actions taken against the company.

The dummy variables for different CompanyIDs provide estimates of the difference in ESG scores between each company and the reference company (which is not specified in the table). Most of these coefficients are negative, suggesting that the average ESG score for these companies is lower than the reference value. However, only a few of these coefficients are statistically significant ($p<0.05$), indicating that the differences are only meaningful for some companies.

In summary, this regression analysis shows that GDP growth and the specific company are the main factors that affect the ESG score, while the number of employees and the rate of return do not have a significant effect. The GDB Growth variable is selected as statistically significant among others for the logit and probit models. This leads to the conclusion that the shape of the time series does not help to fit the particular model to interpret the data. In other words, the analysis fails to detect statistical significance of other variables, which are present in the other models (logit, probit) and are empirically shown to have explanatory power. This shows that imposing the time series on the panel model tends to weaken the cross-sectional characteristics of this dataset. Similar results have been obtained by both statistical software used for this research.

5.4.4 Comparative Analysis

Based on the information provided and the focus on investigating models to explain the relationship between ESG scores and other variables, the study utilized logistic regression, probit regression, and panel data regression to compare the performance of the models.

The logistic regression model included the variables "Sector", "LogEmployees", "GDP_Growth", and "RR_100" in the equation, with all of them being statistically significant ($p<0.05$). The classification table indicated that the overall percentage of correctly predicted observations was 83.4%. The area under the ROC curve was 0.886, indicating good discriminatory power.

Similarly, the probit regression model also included the variables "Sector", "LogEmployees", "GDP_Growth", and "RR_100". The regression coefficients suggested that the variables had a statistically significant effect on ESG scores ($p<0.05$). The area under the ROC curve was also high, at 0.885, indicating good discriminatory power.

The panel data regression analysis suggested that GDP growth and the CompanyID were the main factors that affected the ESG score, while the number of employees and the rate of return did not have a significant effect. However, only a few of the dummy variable coefficients for different CompanyIDs were statistically significant, indicating that the differences were only meaningful for some companies.

In addition, it is worth noting that while both the logistic and probit regression models show good discriminatory power and are able to explain the relationship between ESG scores and other variables, the logistic regression model appears to have a slightly better performance, with a slightly higher overall percentage of correctly predicted observations and a slightly larger area under the ROC curve. However, the difference in performance between the two models is relatively small and both can be considered appropriate for analysing the data. On the other side, the panel data regression analysis suggested that the time series component of the data might not be as important in explaining the relationship.

5.4.5 Sensitivity Analysis

Sensitivity analysis in logistic regression involves testing the stability of the model by examining the impact of small changes in the model's assumptions, inputs, or parameters on the model's results. This analysis helps to identify the robustness of the model and its sensitivity to changes.

The logistic regression model is represented as:

$$P = exp(a + bX1 + cX2 + \cdots + nXn) \,/\, [\, 1 + exp(a + bX1 + cX2 + \cdots + nXn) \,]$$

$$= exp(.)\,[1 + exp(.)]^{-1}$$

where P is the predicted probability of the outcome, a, b, and c are coefficients, and X1 to Xn are the predictor variables.

To calculate the sensitivity of P to small changes in X1 while holding everything else constant, the partial derivative dP/dX1 is taken. The result is bP(1-P), where b is the coefficient for X1, and P is the predicted probability of the outcome. This equation shows that the sensitivity of P to changes in X1 depends on the value of P: the closer P is to 0 or 1, the less sensitive it is to changes in X1, while the middle range is where the maximal effect of changes occurs. The sign of b determines whether the effect of changes in X1 on P is positive or negative.

$$dP/dX1 = exp(.)\,(-1)\,[\,1 + exp(.)\,]\text{^} - 2\,exp(.)\,b + [\,1 + exp(.)\,]\text{^} - 1\,exp(.)\,b$$

$$= b\,exp(.)\,[1 + exp(.)]\text{^} - 1\,\{\,1 - exp(.)[1 + exp(.)]\text{^} - 1]\}$$

$$= b\,P\,(1 - P)$$

The derivative of the sum total with respect to small changes in X1, X2, ... Xn can also be calculated, provided that there are no changes in the vicinity of the initial point and the variables are modified in a minor way. This can be expressed as:

$$DP = dP/dX1\,DX1 + dP/dX2\,DX2 + \ldots + dP/dXn\,DXn$$

Using the best performance model and trying to understand the business insights based on each independent variable that entered to the model as statistically significant variable and its affect to the overall probability of ESG event, a sensitivity analysis took place using STATA software. There are concepts like elasticity, value range and rates of change that can be helpful to understand how each one affects the probability of default. In other words, sensitivity analysis assesses how “sensitive” the model is to fluctuations in the parameters and data on which it is built.

Elasticity refers to the responsiveness of a variable to changes in another variable. In the context of a logistic regression model, the elasticity represents the percentage change in the predicted probability of the outcome for a given percentage change in the value of the independent variable. Elasticity is calculated as the product of the coefficient estimate and the ratio of the mean of the independent variable to the mean of the dependent variable.

Value range refers to the range of values that a variable can take. In the context of sensitivity analysis, value range is important because it can help to identify the upper and lower limits of a variable and determine how changes in the variable within that range affect the outcome.

Rates of change refer to the amount of change in a variable over a given period of time. In sensitivity analysis, rates of change can be used to identify how changes in a variable over time affect the outcome. By calculating the rates of change for each variable, one can assess which variables have a greater impact on the outcome and which variables have a smaller impact.

Overall, sensitivity analysis is an important tool for understanding the impact of changes in variables on the predicted outcome of a logistic regression model. By considering concepts such as elasticity, value range, and rates of change, one can gain a better understanding of how changes in variables affect the outcome and make more informed decisions based on the results of the analysis (Hosmer et al., 2013). The estimations are detailed in Tables 39.

Table 39: Logit Model, Marginal Effects

Marginal Effects

	dy/dx	Std. Err.	Z	P>z	[95% Conf.	Interval]
sector	-0.07053	0.016007	-4.41	0	-0.10191	-0.03916
logemployees	0.065203	0.010346	6.3	0	0.044925	0.085482
gdp_growth	-0.02155	0.004283	-5.03	0	-0.02995	-0.01316
rr_100	0.003477	0.001104	3.15	0.002	0.001315	0.00564

The table shows the marginal effects of four variables: sector, logemployees, gdp_growth, and rr_100. The first column (dy/dx) shows the change in the dependent variable (y) associated with a small change in the corresponding independent variable (x), holding all other variables constant. The second column (Std. Err.) shows the standard error of the estimated marginal effect. The third column (Z) shows the test statistic for the null hypothesis that the marginal effect is equal to zero. The fourth column (P>z) shows the p-value associated with the test statistic, indicating the level of significance. In this case, all p-values are less than 0.05, indicating that the marginal effects are statistically significant at the 5% level. The fifth column ([95% Conf. Interval]) shows the 95% confidence interval for the estimated marginal effect. If the interval does not include zero, it indicates that the marginal effect is statistically significant.

From dy/dx it is possible to understand the relationship between ESG probability and the explanatory variables. So, when GDP_Growth increases then the probability of ESG decreases since they have a negative relationship. On the other hand, when RR_100 increases then the probability of ESG increases since they have a positive relationship.

Specifically, the first row shows that a one-unit increase in the sector variable is associated with a decrease of 0.07053 in the dependent variable, holding all other variables constant. This marginal effect is statistically significant at the 5% level, as indicated by the p-value of 0.

Similarly, a one-unit increase in logemployees is associated with an increase of 0.065203 in the dependent variable, holding all other variables constant. This marginal effect is also statistically significant at the 5% level, with a p-value of 0.

On the other hand, a one-unit increase in gdp_growth is associated with a decrease of 0.02155 in the dependent variable, holding all other variables constant. This marginal effect is statistically significant at the 5% level, with a p-value of 0.

Finally, a one-unit increase in rr_100 is associated with an increase of 0.003477 in the dependent variable, holding all other variables constant. This marginal effect is statistically significant at the 5% level, with a p-value of 0.002.

Let's now use the table provided earlier to explain the concepts of elasticity, value range and rates of change:

Elasticity: Elasticity measures the percentage change in the outcome variable (in this case, the probability of ESG event) for a one percent change in an explanatory variable. It is calculated as the product of the marginal effect and the ratio of the explanatory variable to the mean of the explanatory variable. For example, let's take the marginal effect for the variable "logemployees", which is 0.065203. To calculate the elasticity of logemployees, we multiply this marginal effect by the ratio of the variable to its mean. The mean of logemployees is 8.3211. Then the elasticity of logemployees is 0.065203 * (8.3211/1) = 0.54256. This means that a one percent increase in logemployees is associated with a 0.65203 percent increase in the probability of ESG event, holding all other variables constant. So, if the mean value of logemployees is used as a reference, it can be said that a 1% increase in logemployees is associated with a 0.54256% increase in the probability of default, holding all other variables constant.

Value range: The value range is the range of values that an explanatory variable can take, and it is used to compare the magnitudes of the marginal effects of different variables. Looking at the table, we can see that the variable "logemployees" has a larger marginal effect (0.065203) than the other variables, which suggests that it has a stronger impact on the probability of ESG event. However, the value range of "logemployees" is also larger than the other variables, which means that it has more room to vary and therefore its marginal effect may be less robust.

Rates of change: Rates of change refer to how much the outcome variable changes for a one unit change in an explanatory variable. In the table, the marginal effect for "rr_100" is 0.003477, which means that a one unit increase in "rr_100" is associated with a 0.003477 increase in the probability of ESG event, holding all other variables constant. This information can clarify the economic significance of the variable and how it affects the outcome variable.

5.4.6 Out-of-Sample Forecasting

In the real economy, historical data is commonly used to make informed estimates and forecast future trends. The historical data spanning seven years up to 2021 could be used for out-of-sample forecasting. The explanatory variables for the financial year ending in 2021 could be collected, and a prediction could be made to estimate the probability of these firms experiencing ESG events during the year 2022. In early 2023, the real ESG events of 2022 will be collected and added to the historical data. At this point, the actual incidents can be compared with the model estimates made as forecasts for that specific year, allowing for an evaluation of the model's discriminatory power. This process can be repeated over several years to assess the model's predictive accuracy over time.

In this study and using the best performing model with the explanatory variables selected as statistically significant variables and their coefficients, an out-of-sample forecast was applied. For this purpose, the dataset was spitted in two samples by separating the years 2015 to 2020 as historical data to predict the probability of ESG default of 2021 with explanatory variables data from the financial year of 2020. This is the last year with collected data at the initial dataset. It is already known from dataset's descriptive statistics that 2021 presents more ESG incidents than any other year. It is easy to understand that year 2021 is decisive for the model performance and accuracy.

In this case, it was a 5% of the total of 50 observations of 2021 was included back to the dataset. At this point, the logit model runs again with this new number of 305 total observations and to estimate the coefficients and the estimations are shown on Tables 40, 41, 42.

Table 40: In-Sample, Classification Table

Classification Table[a]

			Predicted		
			ESG		
Observed			Non-Default	Default	Percentage Correct
Step 1	ESG	Non-Default	233	16	93,6
		Default	32	24	42,9
	Overall Percentage				84,3

a. The cut value is ,500

Table 41: In-Sample, Variables in the Equation

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Sector	-,615	,186	10,935	1	,001	,540
	LogEmployees	,742	,143	27,052	1	,000	2,100
	GDP_Growth	-,242	,123	3,868	1	,049	,785
	RR_100	,043	,014	9,290	1	,002	1,044
	Constant	-10,017	1,747	32,891	1	,000	,000

a. Variable(s) entered on step 1: Sector, LogEmployees, GDP_Growth, RR_100.

Table 42: In-Sample, Model Performance

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	185,620[a]	,292	,475

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than ,001.

The next step is to compare the impact of each coefficient in this model with the one derived from the full dataset. The results indicate that the in-sample coefficients have a comparable effect on the probability as the logit model using the full dataset. To assess its performance, a ROC curve analysis is depicted in Figure 16 and Table 43.

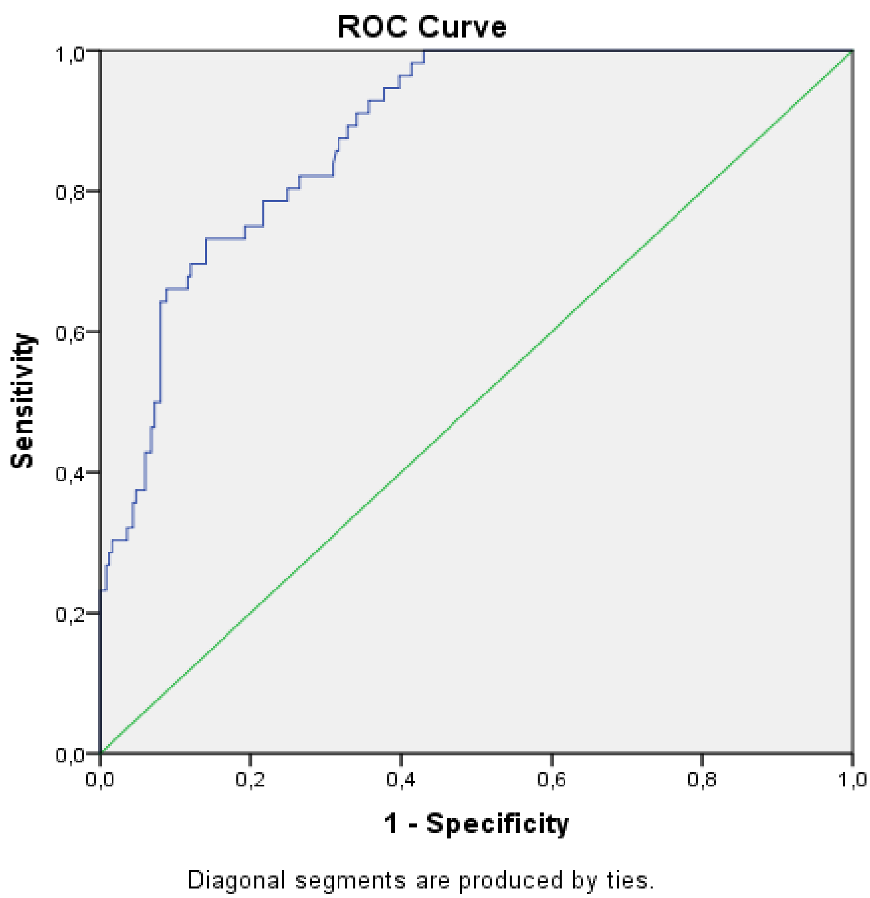

Figure 15: In-Sample, ROC Curve

Table 43: In-Sample, Area Under the Curve

Area Under the Curve

Test Result Variable(s): Predicted probability

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval Lower Bound	Asymptotic 95% Confidence Interval Upper Bound
,881	,021	,000	,840	,923

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption
b. Null hypothesis: true area = 0.5

Subsequently, the model equation and probability formula were utilized to perform an out-of-sample forecast for the remaining 45 observations of 2021, which included 20 ESG incidents, as presented in Table 44. By solely estimating the probabilities of these observations, a ROC Curve Analysis was conducted to evaluate the model's discriminatory power, as illustrated in Figure 17, and the AUC was determined and reported in Table 45.

Table 44: Out-of-Sample Forecasting, ESG Observations

Case Processing Summary

ESG[a]	Valid N (listwise)
Positive[b]	20
Negative	25

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.
a. The test result variable(s): P has at least one tie between the positive actual state group and the negative actual state group.
b. The positive actual state is 1.

Table 45: Out-of-Sample Forecasting, Area Under the Curve

Area Under the Curve

Test Result Variable(s): P

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval Lower Bound	Asymptotic 95% Confidence Interval Upper Bound
,816	,065	,000	,688	,944

The test result variable(s): P has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption
b. Null hypothesis: true area = 0.5

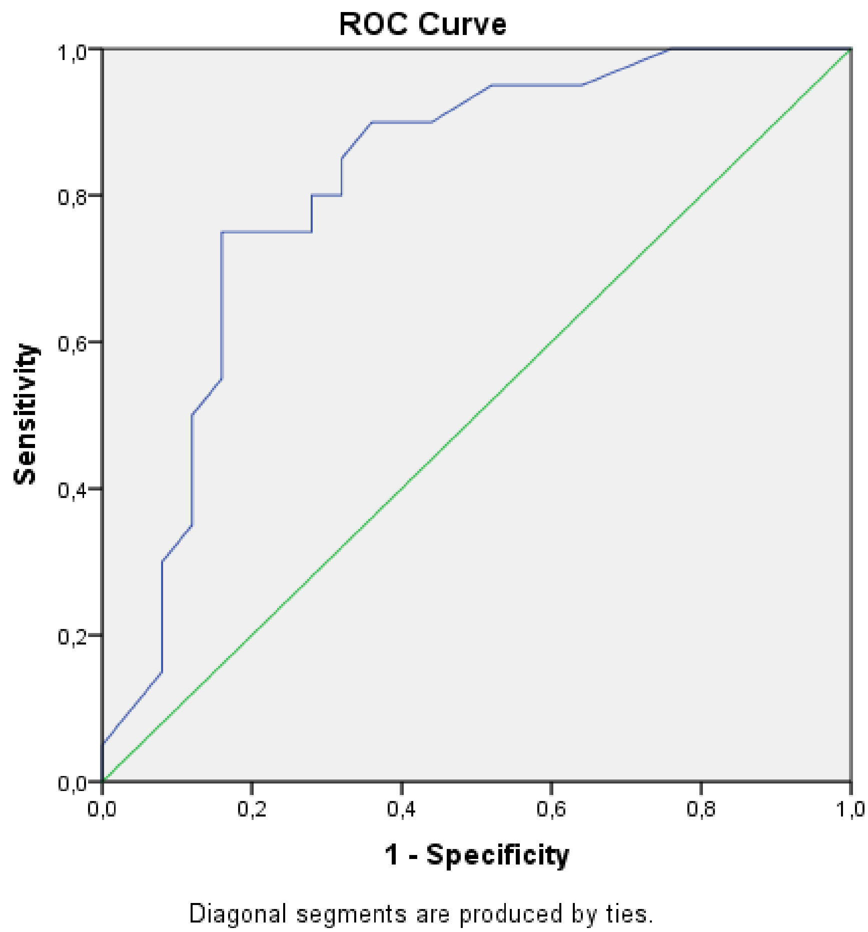

Figure 16: Out-of-Sample Forecasting, ROC Curve

It was anticipated that the model's performance would decrease when using less data compared to the initial model. Furthermore, the performance of the model was found to be even weaker when evaluated using out-of-sample forecasting analysis.

5.6 Summary of Results

To provide a business context for the analysis results, it would be helpful to reiterate the definitions and descriptions of the explanatory variables used in the analysis. This would help to better understand how these variables were used to determine the outcome and make informed decisions based on the results. Additionally, providing insights into how the explanatory variables may impact the ESG probability in the future could help the model development with any potential changes in these variables. The model equation includes the variables:

Sector: a dummy variable included in the model for the purpose of ranking based on necessity, with a focus on essential industries.

1 = Food, 2 = Construction, 3 = Electric & Gas Utilities, 4 = Hospitality

LogEmployees: Number of people employed by the company, based on the number of full-time equivalents. If unavailable, then the number of full-time employees is used, excluding part time employees.

GDP_Growth: The GDP growth (annual %) in Eurostat refers to the annual percentage change in the Gross Domestic Product (GDP) of the European Union member states. This indicator is used to measure the rate of economic growth in the region and provides insights into the overall health of the EU's economy.

RR_100: The recovery rate is recorded as cents on the dollar recovered by secured creditors through judicial reorganization, liquidation, or debt enforcement (foreclosure or receivership) proceedings. The calculation takes into account the outcome: whether the business emerges from the proceedings as a going concern, or the assets are sold piecemeal.

Regarding of the sensitivity analysis of the logit model it was possible to understand the relationship between the probability of the appearance of an ESG incident with the explanatory variables. The variables Sector and GDP_Growth have a negative relationship with the ESG probability, meaning that as they increase, the probability of ESG decreases. On the other hand, the variables LogEmployees and RR_100 have a positive relationship with the ESG probability, meaning that as they increase, the probability of ESG increases.

The binary logistic regression model provides insights into the probability of an ESG incident occurring based on certain predictors in the business and finance context. The variable Sector, which represents the industry sector of a company, has a negative relationship with the ESG probability, indicating that certain industries may have a higher likelihood of ESG incidents than others. As a cross check for this hypothesis, it is possible to use the ESG incident percentages per sector. According to Table 46, Food sector has the higher percentage of 37,1% in ESG incident appearance, the next one is the Construction sector with 25,5%, following the Electric and Gas Utilities with 17,9% and the last one with the lower percentage is the Hospitality sector with 10,2%. The variable GDP_Growth, which measures the rate of economic growth in the European Union member states, also has a negative relationship with the ESG probability, suggesting that economic stability may decrease the likelihood of ESG incidents. However, the variables LogEmployees and RR_100 have positive relationships with the ESG probability, implying that companies with more employees and higher recovery rates may be at a higher risk of ESG incidents. These insights can help businesses and investors make more informed decisions about risk management and sustainability practices.

Table 46: Percentages of ESG incidents per sector

Sector * ESG Crosstabulation

			ESG		
			0	1	Total
Sector	1*	Count	44	26	70
		% within Sector	62,9%	37,1%	100,0%
	2*	Count	73	25	98
		% within Sector	74,5%	25,5%	100,0%
	3*	Count	69	15	84
		% within Sector	82,1%	17,9%	100,0%
	4*	Count	88	10	98
		% within Sector	89,8%	10,2%	100,0%
Total		Count	274	76	350
		% within Sector	78,3%	21,7%	100,0%

*Sectors: 1 = Food, 2 = Construction, 3 = Electric & Gas Utilities, 4 = Hospitality

As an example, and according to the previous results, a company in the hospitality sector with a high recovery rate in case of debt enforcement proceedings and a low number of employees may have a higher probability of experiencing an ESG incident than a company in the food sector with a higher number of employees and a lower recovery rate. Additionally, the GDP growth of the European Union member states can also impact the probability of ESG incidents for companies operating in different sectors. Companies operating in sectors with a negative relationship with GDP growth may have a higher probability of ESG incidents compared to companies in sectors with a positive relationship with GDP growth. Therefore, companies should proactively manage ESG risks avoiding negative impacts on their business operations and reputation, taking into account factors such as the sector they operate in, their number of employees, their recovery rate, and the GDP growth in their region.

Chapter 6: Conclusion

From the review of ESG methodologies and regulations, it is evident that there exist various qualitative and quantitative models with distinct scoring criteria, indicators, and interpretations. In this study, these models are categorized into market-based and academic research-based models, based on their differences. ESG data and rating providers adopt their own methodologies based on their needs and the primary ESG definition, as there are no global reporting standards. It is crucial to reach a consensus on what should be considered material for each sector. Moreover, regulatory cooperation among institutions in different countries or sectors is crucial for achieving goals. However, the global ESG measurement systems are still incomplete, leaving room for further improvement.

This study utilized three multivariable methods: LR, Probit, and Panel data regression, with a focus on ESG research. These methods are commonly used to model and understand the relationship between multiple independent variables and a binary dependent variable.

The dataset used in this research was created from Bloomberg published news and articles. It included 50 listed European firms from 4 main sectors along with their incidents (if any) over a period of 7 years. Independent variables were derived from each firm's financial accounts, as well as macro and country risk measures. The incidents were not classified by sector, and incidents occurring in the same year for a company were treated as one.

The results have provided financial significance for each independent variable included in the model for both Logit and Probit methodologies. Due to the real corporate data and the volatility of the dataset's values, the Probit model did not provide significantly different results, as it has a special treatment for outliers. However, the panel methodology did not produce meaningful results concerning the parameter of time and cross-sections, possibly due to the small size of the dataset.

The final model primarily utilized financial data as independent variables for each company, suggesting that corporate financial performance could influence ESG scores. The results indicate that larger and financially stronger companies, particularly in essential industries, are more prone to experiencing ESG incidents. However, the macroeconomic conditions seem to have a favorable effect on companies' adoption and compliance with ESG standards.

One limitation of this study is that the number of firms and years data is relatively small comparing to a real dataset for all EU or even global firms and sectors. This makes it debatable whether the results obtained from the regression are sufficiently informative. Despite this limitation and based on the lack of global reporting standards and agreement on what should be deemed as material for ESG methodologies, the research certainly adds value, and it is obvious there is a relationship between corporate financial performance and ESG score.

As a suggestion for future research, it would be worthwhile to investigate the ESG incidents further categorized as E, S and G by focusing on the independent variables of each model and how they impact the probability of ESG default. Additionally, studying the incidents across different sectors and countries could provide valuable insights beyond the previous results. It would also be worthwhile to consider the severity and certainty levels of the incidents and classify them into different categories. Another area for potential research is the dynamic correlation between ESG incidents.

Moreover, corporate financial data such as PD, Credit Risk, Sustainability, Enterprise Value, etc., could be meaningful for ESG scoring and predicting the probability of incident occurrence. Investigating any potential relationships between these variables and ESG ratings would be insightful.

An interesting recommendation is to use other databases like, for example, Moody's or MSCI to replicate this research and compare the results. In this respect, the model could be fine-tuned and optimized accordingly by adjusting the explanatory variables according to specific characteristics of each sector or even by including additional variables.

Overall, the future of ESG as an aspect of business and investment looks promising, as global standards for environmental, social, and governance protection are becoming increasingly specific and rigorous. Investments are one of the pillars in global economy and the value added from ESG ratings open new directions. Based on records from 2022, the ESG ranking podium is exclusively occupied by Nordic countries, with Sweden and Iceland in second and third place, respectively, while Finland takes the top spot (Global Risk Profile, 2021). It will be really challenging to encourage more and more companies to make a long-term commitment to environmental, social and governance goals. It could be said that this is like a crossroad and the direction taken could determine how companies but also the wider society will evolve over the following years.

References

Afreen, M. (2021). Open Access Original Research Building Sustainability Resiliency During Post Covid World Focusing ESG And Climate Risk Issue. *International Journal of Social Sciences and Economic Review*, pp.21-27.

Agresti, A. (2018). An introduction to categorical data analysis. John Wiley & Sons.

Agresti A. (2015). Foundations of Linear and Generalized Linear Models. *New York: Wiley*. Pp. 183–186. ISBN 978-1-118-73003-4.

Agresti, A. (2007) Building and Applying Logistic Regression Models. In: An Introduction to Categorical Data Analysis, *2nd Edition, John Wiley & Sons, Hoboken, NJ*, 137-172.

Aldrich J., Nelson F., Adler S. (1984). Linear Probability, Logit, and Probit Models. *Sage*. Pp. 48–65. ISBN 0-8039-2133-0.

Allen, K. (2018), Lies, damned lies and ESG rating methodologies, *Financial Times*.

Allen, F. (2018). The challenges of ESG ratings. CFA Institute.

Altman, N. S. (1991). An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician, 45(4), 276-280.

Atkins B. (2020). Demystifying ESG: Its History & Current Status. Report available online: https://www.forbes.com/sites/betsyatkins/2020/06/08/demystifying-esgits-history--current-status/?sh=3860610e2cdd

Bellazzi, R., & Zupan, B. (2008). Predictive data mining in clinical medicine: Current issues and guidelines. International Journal of Medical Informatics, 77(2), 81-97.

Bender, R. (2009) Introduction to the Use of Regression Models in Epidemiology. In: Verma, M., Ed., Cancer Epidemiology. Methods in Molecular Biology, 471, 179-195.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. CRC press.

Bewick, V., Cheek, L., Ball, J. (2005) Statistics Review 14: Logistic Regression. *Critical Care*, 9, 112-118. https://doi.org/10.1186/cc3045

Biswas, C., & Veliyath, R. (2018). ESG scorecard: A model for constructing ESG scores. Journal of Applied Accounting Research, 19(2), 240-254.

Bliss C. I. (1934). The Method of Probits. Oxford English Dictionary, 3rd ed. s.v. probit (article dated June 2007). *Science*. 79 (2037): 38–39. Bibcode:1934Sci....79...38B. doi:10.1126/science.79.2037.38. PMID 17813446. These arbitrary probability units have been called 'probits'.

Boateng, E.Y., & Abaye, D.A. (2019) A Review of the Logistic Regression Model with Emphasis on Medical Research. *Journal of Data Analysis and Information Processing*, 7, 190-207. https://doi.org/10.4236/jdaip.2019.74012.

Boffo R., & Patalano R. (2020). ESG Investing: Practices, Progress and Challenges. OECD Paris. Report available online: www.oecd.org/finance/ESG-Investing-Practices-Progress-and-Challenges.pdf

Boubakri, N., Cosset, J.-C., Saffar, W. (2021). ESG ratings: An empirical analysis. *Journal of Corporate Finance*, 67, 101935. https://doi.org/10.1016/j.jcorpfin.2020.101935

Brown, N., & Viehs, M. (2019). ESG scoring for private equity. Journal of Sustainable Finance & Investment, 9(4), 355-374.

CDSB (n.d.). Framework for reporting environmental and social information. Report available online: https://www.cdsb.net/what-we-do/reporting-frameworks/environmental-information-natural-capital.

CDSB Framework (n.d.). CDSB Framework. Report available online: https://www.cdsb.net/what-we-do/reporting-frameworks

Chatterji, A. K., Durand, R., Levine, D. I. and Touboul, S. (2016), Do ratings of firms converge? Implications for managers, investors, and strategy researchers, *Strategic Management Journal*, Vol. 37 No. 8, pp. 1597-1614.

Chesebrough D. (2022). UN Global Compact. Integrate the Principles for Responsible Investment Report available online: https://www.unglobalcompact.org/take-action/action/responsible-investment

Christodoulakis, G.A., & Satchell, S.E. (2008). The validity of credit risk model validation methods, *Working Paper,* Manchester Business School.

Collett D. (2003). Modelling Binary Data, *2nd ed. Chapman & Hall*, London, 344 p.

Cox D. R., & Snell E. J. (1989). The Analysis of Binary Data, 2nd ed. *Chapman and Hall,* London, 236p.

Davies A., & Lahiri K. (1995). A New Framework for Testing Rationality and Measuring Aggregate Shocks Using Panel Data. *Journal of Econometrics*. 68 (1): 205–227. doi:10.1016/0304-4076(94)01649-K.

Davis P. (2013), Corporations, global governance, and post-conflict reconstruction. *Routledge*, London.

Deutsche Börse (2020). Deutsche Börse acquires leading governance, ESG data and analytics provider ISS. Report available online: https://www.deutsche-boerse.com/dbg-en/media/press-releases/Deutsche-B-rse-acquires-leading-governance-ESG-data-and-analytics-provider-ISS-2343868

Dorfleitner G., G. Halbritter, M. Nguyen (2015). Measuring the level and risk of corporate responsibility – An empirical comparison of different ESG rating approaches, *Journal of Asset Management*, Vol. 16 No. 7, pp. 450-466.

Engelmann, B., Hayden, E., Tasche, D. (2003). Testing rating accuracy. *Risk*, 16, 82–6.

Engelmann, B., Hayden, E., Tasche, D. (2003). Measuring the Discriminative Power of Rating Systems. *Series 2: Banking and Financial Supervision* No 01/2003

Ernst, Young (2021). The future of sustainability reporting standards. Report available online: https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/sustainability/ey-the-future-of-sustainability-reporting-standards-june-2021.pdf

Flammer, C. (2015). ESG risk factors and their financial impacts: An event study analysis. Journal of Sustainable Finance & Investment, 5(4), 240-262.

Friedman, J., Hastie, T., Tibshirani, R. (2001). The elements of statistical learning: Data mining, inference, and prediction. Springer Science & Business Media.

FTSE Russell (2018). Methodology Enhancements 2018 – 2019 Research Cycle. Report available online: https://www.jse.co.za/sites/default/files/jse_document_manager/RW/Internal/Indices/FTSEJSE%20Responsible%20Investment%20index%20series/2018%20ESG%20Model%20Enhancements/FTSE%20Russell%20ESG%20Ratings_Methodology%20changes%20for%20the%20research%20cycle%20April.pdf

FTSE Russell (n.d.). Sustainability and ESG data/ESG ratings. Report available online: https://www.ftserussell.com/data/sustainability-and-esg-data/esg-ratings

Gajo M. (2020). Börse Singapur und FTSE Russell erweitern Partnerschaft, Die Aktiengesellschaft, 65(19), pp. r285-r285. https://doi. org/10.9785/ag-2020-651918

Gelman, A., & Stern, H. (2006). The difference between “significant” and “not significant” is not itself statistically significant. The American Statistician, 60(4), 328-331.

Ghosh, P., Jana, R.K., Mukherjee, K. (2017). A multi-criteria decision model for ESG rating of companies. Decision Science Letters, 6(2), 157-174.

Global Risk Profile (2021). The 2021 ESG Index is Out! Nordic countries at the top of the 2021 ESG Index. Report available online: https://globalriskprofile.com/2021-esg/

Greene, W. H. (2018). Econometric analysis (8th ed.). Pearson. Chapter 18 discusses the probit model and its assumptions.

GRI (2022), The GRI Perspective. ESG standards, frameworks and everything in between. Report available online: https://www.globalreporting.org/media/jxkgrggd/gri-perspective-esg-standards-frameworks.pdf

GRI (2022). Four-in-five largest global companies report with GRI. Report available online: https://www.globalreporting.org/news/news-center/four-in-five-largest-global-companies-report-with-gri/

Goh, C. F., & Lee, T. H. (2018). Multidimensional data analysis in business research: A review on current practices and trends. *Journal of Business Research*, 89, 198-210.

Ground J. (2022). ESG Global Study 2022. *Harvard Law School Forum on Corporate Governance.* Report available online: https://corpgov.law.harvard.edu/2022/06/17/esg-global-study-2022/

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar), 1157-1182.

Hale J. (2019). Enhancement to Sustainability Rating Emphasizes Material ESG Risk. Report available online: https://www.morningstar.com/articles/954595/enhancement-to-sustainability-rating-emphasizes-material-esg-risk

Hawley, J. (2017), ESG Ratings and Rankings: All over the Map. What Does it Mean? (Truevalue Labs).

Hilbe J. (2011), LOGISTIC REGRESSION, Arizona State University. Report available online: https://encyclopediaofmath.org/images/6/69/Logistic_regression.pdf

Hirai A., & Brady A. (2021). Managing ESG Data and Rating Risk. *Harvard Law School Forum on Corporate Governance*. Report available online: https://corpgov.law.harvard.edu/2021/07/28/managing-esg-data-and-rating-risk/

Hoepner, A. G. F. (2017). Challenges and opportunities in ESG rating and sustainability reporting. *Journal of Sustainable Finance & Investment*, 7(2), 95–110. https://doi.org/10.1080/20430795.2017.1321145

Hoepner, A. G. F., Rezec, M., Hebb, T. (2019). ESG Shareholder Engagement and Downside Risk. *Journal of Business Ethics*, 157(3), 767-785.

Hosmer, D.W., Lemeshow, S., & Sturdivant, R.X. (2013). Applied logistic regression (3rd ed.). Wiley.

Hsiao C., Lahiri K., Lee L., et al., eds. (1999). Analysis of Panels and Limited Dependent Variable Models. *Cambridge: Cambridge University Press*. ISBN 0-521-63169-6.

ISSgovernance (2018). Deal Furthers ISS' Responsible Investment Business Expansion. Report available online: https://www.issgovernance.com/oekom-research-ag-join-institutional-shareholder-services/

J.P. Morgan (n.d.). EU SFDR Explained: A guide to the EU Sustainable Finance Disclosure Regulation for investors. Report available online: https://am.jpmorgan.com/se/en/asset-management/liq/investment-themes/sustainable-investing/understanding-SFDR/

Jiang, X., El-Kareh, R., Ohno-Machado, L. (2011) Improving Predictions in Imbalanced Data Using Pairwise Expanded Logistic Regression. AMIA Annual Symposium Proceedings, 2011, 625-634. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3243279/

Jolliffe, I. T. (2002). Principal component analysis. Wiley Online Library.

Kim, Y., Kwon, S., Song, S.H. (2006) Multiclass Sparse Logistic Regression for Classification of Multiple Cancer Types Using Gene Expression Data. Computational Statistics & Data Analysis, 51, 1643-1655. https://doi.org/10.1016/j.csda.2006.06.007

King, G., Tomz, M., Wittenberg, J. (2000) Making the Most of Statistical Analyses: Improving Interpretation and Presentation. *American Journal of Political Science*, 44, 341-355. https://doi.org/10.2307/2669316

KPMG (n.d.). Corporate Sustainability Reporting Directive. Report available online: https://home.kpmg/nl/en/home/topics/environmental-social-governance/corporate-sustainability-reporting-directive.html

Krüger, P., & Sautner, Z. (2021). Disagreement in Socially Responsible Investing. *Journal of Financial Economics*, 139(3), 869-893.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1-26.

Kuhn, M., & Johnson, K. (2013). Applied predictive modelling. Springer.

Kumar, V.S., & Ganesh, K. (2020). A machine learning model for ESG performance evaluation of companies. Journal of Cleaner Production, 268, 122209.

Levy, P.S., & Stolte, K. (2000) Statistical Methods in Public Health and Epidemiology: A Look at the Recent Past and Projections for the Next Decade. Statistical Methods in Medical Research, 9, 41-55. https://doi.org/10.1177/096228020000900106

Li T., Wang K., Sueyoshi T., Wang D. (2021). ESG: Research progress and future prospects. Sustainability 2021, 13(21), 11663. https://doi.org/10.3390/su132111663

Little, R. J., & Rubin, D. B. (2014). Statistical analysis with missing data. John Wiley & Sons.

Long J.C., & Freese J. (2014). Regression Models for Categorical Dependent Variables Using Stata, *3rd ed. College Station*: Stata Press, 589 p.

Maddala G. S. (2001). Introduction to Econometrics (Third ed.). *New York: Wiley*. ISBN 0-471-49728-2.

Menard, S. (2002) Applied Logistic Regression Analysis. *2nd Edition*, Volume 106, Sage, New York.

Mervis, J. (2017). What Are Outliers and Why Are They Important? Harvard Business Review. Retrieved from https://hbr.org/2017/02/what-are-outliers-and-why-are-they-important

Moody's (2019). Moody's Acquires Majority Stake in Vigeo Eiris, a Global Leader in ESG Assessments. Report available online: https://ir.moodys.com/press-releases/news-details/2019/Moodys-Acquires-Majority-Stake-in-Vigeo-Eiris-a-Global-Leader-in-ESG-Assessments/

Moody's (n.d.). ESG Capabilities. Report available online: https://esg.moodys.io/esg-measures

Morningstar (2020). Morningstar, Inc. Completes Acquisition of Sustainalytics. Report available online: https://newsroom.morningstar.com/newsroom/news-archive/press-release-details/2020/Morningstar-Inc-Completes-Acquisition-of-Sustainalytics/default.aspx

MSCI (2014). MSCI Completes Acquisition of GMI Ratings. Report available online: https://ir.msci.com/news-releases/news-release-details/msci-completes-acquisition-gmi-ratings

MSCI (2016). MSCI Completes Acquisition of Carbon Delta. Report available online: https://ir.msci.com/news-releases/news-release-details/msci-completes-acquisition-carbon-delta

MSCI (2020). MSCI ESG Metrics Calculation Methodology. Report available online: https://www.msci.com/documents/10199/1283513/MSCI_ESG_Metrics_Calc_Methodology_Dec2020.pdf/92a299cb-0dbc-63ba-debb-e821bd2e2b08

MSCI (n.d.). ESG Ratings Key Issue Framework. Report available online: https://www.msci.com/our-solutions/esg-investing/esg-ratings/esg-ratings-key-issue-framework

Naftemporiki (2022). On development: ESG criteria and the opportunity for Greek businesses to attract new investments and financing. Report available online: https://www.naftemporiki.gr/finance/1376977/peri-anaptyxis-ta-kritiria-esg-kai-i-efkairia-ton-ellinikon-epicheiriseon-gia-proselkysi-neon-ependyseon-kai-chrimatodotiseon/

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. Biometrika, 78(3), 691-692.

Niemoller J. (2021). ESG Reporting Frameworks: Comparing CDP, GRI & More. Report available online: https://www.perillon.com/blog/esg-reporting-frameworks

OECD (2021). ESG Investing and Climate Transition: Market Practices, Issues and Policy Considerations, OECD Paris. Report available online: https://www.oecd.org/finance/ESG-investing-and-climatetransition-Market-practices-issues-and-policy-considerations.pdf

Oommen, T., Baise, L.G., Vogel, R.M. (2011) Sampling Bias and Class Imbalance in Maximum-Likelihood Logistic Regression. Mathematical Geosciences, 43, 99-120. https://doi.org/10.1007/s11004-010-9311-8

Parrino, R. (2019). The origins and development of ESG: A review of the academic and practitioner literature. *Journal of Investment Management*, 17(3), 1-13. https://doi.org/10.3905/jim.2019.17.3.001

Park, H.A. (2013) An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain. *Journal of Korean Academy of Nursing*, 43, 154-164. https://doi.org/10.4040/jkan.2013.43.2.154

Peterdy K. (2023). ESG (Environmental, Social and Governance). Corporate Finance Institute. Report available online: https://corporatefinanceinstitute.com/resources/esg/esg-environmental-social-governance/

Petridis, D. (2015). LOGISTIC REGRESSION [Chapter]. In Petridis, D. 2015. ANALYSIS OF MULTIVARIATE TECHNIQUES [Undergraduate textbook]. *Kallipos, Open Academic Editions*. chapter 2. http://hdl.handle.net/11419/2128

Preuss, L., & Chapple, W. (2016). Assessing the reliability and relevance of ESG rating systems. *Journal of Business Ethics*, 136(2), 217–232. https://doi.org/10.1007/s10551-014-2511-5

Ratsimiveh K., & Haalebos R. (2021). ESG scores and beyond. Report available online: https://content.ftserussell.com/sites/default/files/esg_scores_and_beyond-part_2_final.pdf.

Reed, P., & Wu, Y. (2013) Logistic Regression for Risk Factor Modelling in Stuttering Research. *Journal of Fluency Disorders*, 38, 88-101. https://doi.org/10.1016/j.jfludis.2012.09.003

Refinitiv, (n.d.). Refinitiv ESG company scores. Report available online: https://www.refinitiv.com/en/sustainable-finance/esg-scores

Rigoni, D., Sartori, L., Zoni, L. (2021). Evaluating ESG ratings: An analysis of methodologies and sustainability reporting trends. *Journal of Business Ethics*. Advance online publication. https://doi.org/10.1007/s10551-020-04785-4

Ruggie J. G. (2013). Just business: Multinational corporations and human rights. New York: W.W. Norton & Company

Sasikumar, S.K., & Ramachandran, S. (2016). Corporate environmental, social and governance modelling: A review. International Journal of Business Governance and Ethics, 11(3), 221-244.

Schulte, L. (2021). EU's New ESG Reporting Rules Will Apply to Many US Issuers. Bloomberg Law. Report available online: https://news.bloomberglaw.com/securities-law/eus-new-esg-reporting-rules-will-apply-to-many-u-s-issuers

Serafeim, G. (2015). What's driving the sustainability movement? Harvard Business Review, 93(4), 1-9. Available online: https://hbr.org/2015/04/whats-driving-the-sustainability-movement

Serafeim G. (2022). ESG Reporting Guide 2022. Athens Stock Exchange. Report available online: https://www.athexgroup.gr/documents/10180/6599246/ESG+Reporting+Guide+2022-2202.pdf

Standard and Poor's Global (2016). S&P Dow Jones Indices acquires Trucost. Report available online: https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/sp-dow-jones-indices-acquires-trucost

Standard and Poor's Global (2018). S&P DJI ESG SCORES. Report available online: https://www.spglobal.com/spdji/en/documents/additional-material/faq-spdji-esg-scores.pdf

Standard and Poor's Global (2019). S&P Global Acquires the ESG Ratings Business of RobecoSAM. Report available online: https://www.spglobal.com/marketintelligence/en/media-center/press-release/sp-global-acquires-the-esg-ratings-business-of-robecosam

Safety4Sea (2022). How ESG relates with UN Sustainable Development Goals. Maritime Knowledge, Sustainability. Report available online: https://safety4sea.com/cm-how-esg-relates-with-un-sustainable-development-goals/

Sariannidis N., Giannarakis G., Prtalidou X., Evangelos B. (2017) The effect of Dow Jones Sustainability Index on Consumer Sentiment Index. Investment Management and Financial Innovations, 14(1), 89-95. doi:10.21511/imfi.14(1).2017.09 http://dx.doi.org/10.21511/imfi.14(1).2017.09.

SASB Standards (n.d.). SASB Standards & Other ESG Frameworks. Report available online: https://www.sasb.org/about/sasb-and-other-esg-frameworks/

Semenova, N., & Hassel, L. G. (2015), On the Validity of Environmental Performance Metrics, *Journal of Business Ethics*, Vol. 132 No. 2, pp. 249-258.

Tavanti M. (2015). Global Sustainability Reporting Initiatives: Integrated Pathways for Economic, Environmental, Social, and Governance Organizational Performance. In Agata Stachowicz-Stanusch (Ed.) Corporate Social Performance: Paradoxes, Pitfalls and Pathways to The Better World (pp. 301-323). Contemporary Perspectives in Corporate Social Performance and Policy book Series, Charlotte, NC: Information Age Publishing (IGI). Report available online: http://www.truevaluemetrics.org/DBpdfs/Sustainability/Global_Sustainability_Reporting_Initiati.pdf

TCFD (n.d.). Financial Stability Board (FSB). Report available online: https://www.fsb-tcfd.org/about/.

The GRI Perspective (2022), ESG standards, frameworks and everything in between. Report available online: https://www.globalreporting.org/media/jxkgrggd/gri-perspective-esg-standards-frameworks.pdf

United Nations Industrial Development Organization (n.d.). What is CSR? Report available online: https://www.unido.org/our-focus/advancing-economic-competitiveness/competitive-trade-capacities-and-corporate-responsibility/corporate-social-responsibility-market-integration/what-csr#:~:text=Corporate%20Social%20Responsibility%20is%20a,and%20interactions%20with%20their%20stakeholders

Vargas, J. L., & Moneva, J. M. (2021). Disagreement in ESG ratings: A systematic literature review. Sustainability, 13(3), 1143. https://doi.org/10.3390/su13031143

Wang, L., Zhang, S., Zhu, J. (2019). ESG rating model for equity investment. Journal of Cleaner Production, 233, 765-774.

Wettstein F. (2009). Multinational corporations and global justice: Human rights obligations of a quasi-governmental institution. *Stanford, Calif: Stanford Business Books.*

Wolf K. D., Flohr A., Rieth L., Schwindenhammer S. (2010). Then role of business in global governance: Corporations as norm-entrepreneurs. Basingstoke, England: Palgrave Macmillan.

Wong C., & Petroy E. (2020). Rate the Raters 2020. The SustainAbility Institute by ERM. https://www.sustainability.com/thinking/rate-the-raters-2020/

Wong, C.W.Y., & Cheng, T.C.E. (2016). Integrating environmental, social, and governance (ESG) performance metrics into supply chain management: A framework and case study. International Journal of Production Economics, 182, 420-433.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data (2nd ed.). The MIT Press.

Xiao, J., & Yang, Y. (2020). Modeling ESG scores and improving ESG scoring methodologies. Annals of Operations Research, 288(2), 623-639.

Zhang, Z. (2016). Variable selection with stepwise and best subset approaches. Annals of Translational Medicine, 4(7), 136.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 67(2), 301-320.