Norges miljø- og biovitenskapelige universitet

Master's thesis (2024) 30 ECTS
School of Economics and Business

Predictive validity of NLP-based ESG controversy scores

Melka Yonas Hambissa
Master of Science in Business Administration: Business Analytics

Acknowledgements

This thesis was written in the spring of 2024 as the final stage of my master's degree in business and administration from the Norwegian University of Life Sciences (NMBU). With specialization in Business Analytics, I have gained analytical competence and acquired set of skills to address future challenges. Writing a thesis that combines different aspects of sustainability and technology has been amazing and educating.

There are several people I would like to thank. Firstly, I would like to express my deepest gratitude to my supervisor Joachim Scholderer. From the initial shaping of my thesis topic to the final stages of my project, his continuous support and excellent guidance, has been exceptional. I am truly grateful for the time and efforts he has devoted to my thesis.

Secondly, I want to express my sincere gratitude to the former NMBU student, Selma Kazini and Astrid Sofie Schjetne, for allowing me to use the data they have prepared during their master's thesis in 2020 for my second study. This was very important to the results of my research.

Lastly, I would like to deeply thank my two sons Masara and Bonsa for their support and understanding during this academic year, especially during this master's thesis intensive period.

Abstract

ESG controversy scores based on natural language processing (NLP) of media coverage, ticker messages and corporate disclosures are often considered a simple, cheap and broadly applicable alternative to the classical ESG rating systems sold by financial information providers. However, there is insufficient research on their predictive validity. The aim of the research presented in this thesis was to investigate (1) how well classical rating-based ESG scores can be predicted from NLP-based ESG controversy scores, (2) if the predictive relationship differs by industry, and (3) how consistent the effects are when compared across different approaches for constructing ESG controversy scores and rating-based ESG scores. Two studies were conducted to address these questions. Study 1 used data for OSX-listed companies, extracted from LSEG/Refinitiv Datastream. Study 2 focused on unlisted companies, re-analysing a data set developed by Kazinic and Valheim (2020). The results of Study 1 suggest that the LSEG/Refinitiv ESG controversies score has no substantial relationship with the actual LSEG/Refinitiv ESG score; all explainable variation was due to global differences between industries. In financial analysis and portfolio management, such a low level of predictive validity would clearly be unacceptable. Using the LSEG/Refinitiv ESG controversies score as a replacement for classical, rating-based ESG scores cannot be recommended. The results of Study 2 indicate that the controversies scoring method developed by Kazinic and Valheim (2020) is substantially better than the LSEG/Refinitiv ESG controversies score and suffers much less from industry bias but is still subject to considerable noise. If Kazinic and Valheim's controversies scores were to be used as a replacement of actual ESG scores in portfolio management, the predictive validity of the controversies scores would still not be fully sufficient. The approach can be recommended as a screening tool in financial analysis but cannot replace deeper analysis of a company's actual ESG profile before an investment decision is made.

Innhold

Introduction

Background and objectives

Today, in the world of business and investing, environmental, social, and governance assessments have never mattered more. ESG ratings measure a company's commitment to ethical, sustainable, and governance practices. ESG ratings have developed into a helpful tool as corporate preference leans toward companies returning profits and operating sustainability and ethically. This fact is in line with broader social trends in sustainability and corporate responsibility and the general agreement on the fact that businesses should be evaluated not only based on financial but also environmental and contribution to social well-being aspects (Euronext, 2024).

Bloomberg Intelligence identifies a boom in the global market for ESG investment. ESG assets will exceed $40 trillion in 2030, representing more than 25% of projected global assets under management . Europe leads in ESG investments; it has the largest share of ESG assets under management in the world. By 2030, ESG investments in Europe will exceed $18 trillion by 2030 (Bloomberg, 2024). The main factors include the advanced legislative base and investment environment in Europe, which create the most favourable conditions for integrating ESG principles into business and investment strategy.

ESG investing in the Nordic region and Norway, in particular, is notable on a global scale. In fact, concerning the Nordic countries, domestic institutional investors have been paying attention to ESG factors in their investment process for quite some time. A survey by NN Investment Partners revealed that it became mandatory in the region, with Sweden and Denmark being leaders in addressing ESG concerns through appropriating investments (ESG Investing, 2021). In Norway, ESG investing has been particularly active thanks to the efforts of Norges Bank Investment Management (NBIM), the asset management institution of Government Pension Fund Global. In 2023, several additional steps indicated the fund's commitment to ESG principles. Specifically, regarding active ownership, the fund took part in 115,266 votes at shareholder meetings and conducted 3,298 company meetings on ESG topics and matters, corresponding to 62% of its equity portfolio. Additionally, active divestiture was used through 86 divestures based on companies' failure to address ESG risks, expressing a proactive stance towards addressing these issues (NBIM, 2024).

The growing significance of ESG ratings is evident, as evidenced by the increasing use of these factors by investors and businesses. ESG ratings offer a uniform measure of how businesses perform in relation to ESG considerations, allowing investors to make informed decisions that match their values and risk criteria. These ratings affect investor preferences by highlighting firms that are the top performers in sustainability with increased investor attention. In addition, businesses with high ESG ratings enjoy minimal capital costs and better reputations, and they become more willing to incorporate sustainable policies into their core strategies. MSCI research demonstrates that firms rated well on ESG have lower capital costs than those listed poorly. This is because companies with these ESG scores have a lower risk spiral, leading to minimal expenses of equity and debt (MSCI, 2020). Mckinsey and Company explain that strong ESG performers are more productive and obtain top talent, enhancing their financial performance and decreasing their capital expenditure (McKinsey, 2020). The above plays a crucial part in assisting businesses in establishing sustainable policies.

ESG ratings have become an essential part of investment decisions as well as corporate strategies. Nevertheless, the challenges of their accuracy and transparency are very significant. The above-mentioned challenges are rooted in the methodologies' diversity, subjective expert assessment, and a little developed standardization in ESG rating agencies. Firstly, the existing methodologies used in ESG ratings have this main disadvantage: the software inconsistency of different rating systems. Indeed, different criteria and weights for ESG factors are utilized, and this reason leads to great disparities in results. Thus, it can be the example of this issue: a company received excellent high results from one agency while at the same time showing very poor results from another agency. Investors and companies are confused and cannot determine the ESG assessment based on the main three aspects of it. There is "surprisingly low correlation" between the ratings of even well-known information intermediaries. The average value of the correlation coefficient ranges from 0.3-0.66 which is a primary reason for the lack of coherence and sustainability (Liu, 2022).

Another critical issue is that the majority of ESG rating agencies are transparent regarding methodologies. Many agencies do not disclose detailed information on how data are collected, assessed, and how the ESG scores are ultimately calculated. As a result, it is almost impossible for investors and companies to understand the rationale behind the ratings. In addition, it makes it impossible to compare different ratings meaningfully since each provider uses its own methodologies. According to Harvard Corporate Governance blog highlights that ratings providers rarely provide concrete, systematic evidence to support their claims about the reliability and effectiveness of their ratings (Larcker et al., 2022; Tayan et al., 2022). The variability in methodologies and lack of transparency in ESG ratings rigorously undermine investment options and potential growth. Investors relying on ESG ratings might be not able to fully integrate environmental, social, and governance factors, hampering the allocation of capital and reducing incentives for companies to develop ESG metrics improvement (Liu, 2022). The lack of a universally agreed standard is another issue. It is difficult to compare ESG data because there are no agreed-upon standards. Although all dimensions exist in standard metrics and frameworks are close to the same, rating from different agencies cannot be compared meaningfully. ESG metrics also vary by industry and country, which can worsen the situation (Boffo & Patalano, 2020).

Environmental, social, and governance (ESG) scores refers to quantitative scores assigned to companies, funds, or securities used to assess their performance on multiple sustainability criteria. ESG scores are important in enabling investors and other stakeholders to assess the sustainability practices of companies and understand the possible risks or opportunities related to environmental, social, and governance issues. ESG scores are obtained from data on several issues, including carbon emissions, labour practices, corporate governance, and social responsibility. Leading agencies such as LSEG, MSCI, Sustainalytics, and Bloomberg use differing scales and methodologies to rate their ESG scores, with some using numeric scores while others use letter grades (Arvidsson & Dumay, 2022; MSCI, 2020). The process of calculating ESG scores usually involves data collection from public disclosures or company reports or sometimes directly engaging the company on sustainability practices. ESG scores are essential in helping investors identify firms effectively managing their ESG risks and those with higher risks based on poor management. However, due to the inconsistency in scoring mechanisms in the rating agencies, various studies report inconsistency in ESG scoring, leading to difficulties for investors comparing ESG scores across companies (Berg et al., 2022).

Apart from the conventional ESG scores, ESG controversies scores are a set of metrics rating a company's exposure to and management of incidents that have potential harmful effects on society and the environment. These scores are very important in determining how controversies impact a company's overall ESG performance and reputation (Agnese et al., 2024). ESG controversies are analysed and rated by a number of agencies, including Thomson Reuters and MSCI, which monitor news and other accessible information to identify events and misdeeds associated with companies. As previously mentioned, these may include environmental catastrophes, social misdeeds, and malpractice incidents (Refinitiv, 2022). ESGC scores are adjusted to reflect the severity and frequency of scandals, with more severe scandals leading to lower ratings. Such ratings are important because they allow investors and other stakeholders to determine to what extent the company is involved in relevant misconduct, which may lead to adverse results in terms of reputation and finances. Similarly to ESG scores, the calculation of ESG controversies scores varies from one vendor to another, resulting in different ratings for the same events or discrepancies between the scores given by different agencies (Leaders Arena, 2023).

Given the challenges associated with the current methodologies used to assess ESG and ESG controversies, such as inconsistency and lack of transparency, there is a growing interest in using advanced technologies to improve the accuracy and reliability of these assessments. One of the most promising technologies in this framework is the application of Natural Language Processing (NLP). Natural Language Process is an area of artificial intelligence concerned with the interaction between computers and humans in natural language. When it comes to ESG association, however, NLP involves automatically examining text data gathered from a variety of channels, including news reports, social media mentions, statements, and public disclosures, to extract and quantify information relevant to a company's actions and ESG. The NLP system helps ESG rating agencies examine extensive unstructured data arrays to identify patterns, trends, and sentiments that indicate a company's behavior or controversies (Chowdhury et al., 2023).

ESG controversy scores based on natural language processing (NLP) of media coverage, ticker messages and corporate disclosures are often considered a simple, cheap and broadly applicable alternative to the classical ESG rating systems sold by financial information providers. However, there is insufficient research on their predictive validity. The objective of the research presented in this thesis is to fill this gap.

Theory and previous research

The growing emphasis on Environmental, Social, and Governance (ESG) factors has fundamentally reshaped the landscape of sustainable finance. To introduce the context, I will examine the strategies employed by investors to integrate ESG ratings into their investment processes. This analysis will encompass both established approaches like positive and negative screening, as well as emerging trends like thematic and impact investing. I will assess the effectiveness of these strategies in achieving sustainable outcomes and explore their impact on financial performance. Then, I will review in detail the theoretical underpinnings and methodological approaches that define standard ESG ratings for listed companies. I will present the leading ESG rating systems, exploring their criteria and methodologies. Finally, I will introduce of Natural Language Processing (NLP) and its application in generating ESG controversy scores. By analysing unstructured data sources like news articles, reports, and social media, NLP offers a novel approach to identifying and quantifying ESG-related controversies surrounding companies. This section will not only introduce the concept but also evaluate the effectiveness of NLP techniques in capturing ESG risks, with a focus on their potential to provide real-time insights and a more nuanced understanding of a company's sustainability profile.

ESG strategies in sustainable finance

In 2015, all United Nations member states adopted the 2030 Agenda for Sustainable Development, which includes 17 goals, the so-called Sustainable Development Goals (SDGs). Achievable by 2030, these goals represent a global commitment to progress in areas that are vital to people and the world (United Nations, 2020). Sustainable finance is recognized as a key element in bridging the gap between financial systems and the Sustainable Development Goals (SDGs). The European Commission defines sustainable finance as the incorporation of environmental, social, and governance (ESG) factors into investment decisions and financial services, with the aim of fostering sustainable economic growth, minimizing environmental impact, promoting social justice, and providing financial support (European Commission, n.d.) It represents a forward-thinking approach to financial decision-making that prioritizes long-term economic benefits while simultaneously advocating for environmental preservation and social progress, all underpinned by robust governance practices. PK Ozili further expands on this concept, defining sustainable finance as the consideration of ESG factors in financial decisions (Ozili, 2022). This definition not only encompasses the integration of these factors into investment strategies but also emphasizes the importance of supporting initiatives that positively contribute to the SDGs across economic, environmental, and social dimensions.

Sustainable finance has a long history but has gained momentum in recent decades. Its progress is a gradual development driven by the interaction of various factors. The concept of sustainable finance was sown as early as the 1990s with the emergence of socially responsible investing (SRI) (Lawton, 2023). In addition to financial returns, this approach also focuses on ethical considerations and results in divestments of companies that have a negative impact on society or the environment. The 1990s were a period of gradual growth, marked by the launch of the Domini 400 Social Index (1990), which proved that high financial returns could be achieved alongside social responsibility (Lawton, 2023). In addition, the United Nations Framework Convention on Climate Change (1992) laid the foundation for international environmental action and emphasized the increasing urgency of solving environmental problems.

The Principles for Responsible Investment (PRI) released in 2006 encouraged investors to actively consider ESG factors, which was an important step towards mainstream acceptance. The period since the late 2000s has been characterized by further institutionalization and regulatory framework conditions. The Sustainability Accounting Standards Board (SASB) was established in 2011 to develop industry specific ESG accounting standards to increase transparency and comparability (Lawton, 2023). An important milestone was the establishment of the United Nations Sustainable Development Goals (SDGs) in 2015, which address a wide range of social and environmental challenges and provide a common framework for sustainability efforts. Finally, the EU Sustainable Finance Disclosure Regulation (2021) signals a growing regulatory push for ESG reporting and sustainable investment practices The growth of sustainable investing is primarily driven by the increased financial relevance of environmental, social and governance (ESG) factors, a shift among investors towards demanding financial and non-financial outcomes, and a clear preference for investments that are ethically aligned with the individual values (Edmans & Kacperczyk, 2022).

Sustainable finance includes various investment methods such as sustainable investment, socially responsible investment, impact investment and ESG investment. While the terms vary, at their core they share a focus on environmental, social and governance (ESG) factors. The overall goal of these investment strategies is to improve the ESG performance of companies and portfolios, thereby various stakeholders including customers, employees, investors, regulators, and the public. This collective investment philosophy, often referred to for convenience as "green investing", aims to ensure a long-term positive impact on business practices, while reviewing the impact of such investments on financial risk and return (De Spiegeleer et al., 2020). The growing emphasis on ESG factors is not just a trend, but a fundamental shift towards a more sustainable financial landscape. Sustainable financing has a dual benefit: it can help lenders meet regulatory requirements and expand their product range, while providing borrowers with a wider range of choices. Companies with strong ESG performance benefit from sustainable financing and demonstrate greater operational resilience and sustainability (KPMG, 2022). As sustainable finance becomes more common, it is changing companies' ability to raise capital and affecting the capital allocation of financial institutions (KPMG, 2022). ESG financing options must be available at competitive prices to promote sustainability and incentivize sustainable practices on the part of borrowers, thereby laying the foundation for long-term environmental and social benefits. By integrating ESG considerations into their operations, companies can realize strategic advantages that benefit both funders and recipients (KPMG, 2022). In addition, ESG factors play a key role in transforming operating models to adapt to future regulations in the financial sector, highlighting the key role of ESG in shaping the future of finance. As companies increasingly focus on ESG disclosure, they not only attract investment from financial institutions, but also align with climate risk management strategies, emphasizing the material impact of ESG factors on business activities, growth and financial performance (Haagensen et al., 2023) . This comprehensive approach to ESG integration is critical for companies looking to ensure financial sustainability and drive long-term growth in an increasingly ESG-focused financial environment.

In the realm of sustainable finance, the integration of Environmental, Social, and Governance (ESG) strategies has emerged as a pivotal approach to aligning investment decisions with broader societal and environmental objectives. This shift towards sustainable investment practices reflects a growing consensus on the importance of ESG factors in enhancing corporate performance and mitigating risks. As delineated in the work by Borglund et al. (2021), sustainable investment transcends the traditional focus on financial returns to incorporate ethical, environmental, and social considerations into the investment process.

Negative screening is a foundational ESG strategy, characterized by the omission of sectors, companies, or practices from investment portfolios based on specific ESG criteria. This method aims to avoid investments in industries perceived to be harmful to society or the environment, such as tobacco, firearms, or fossil fuels. The primary advantage of negative screening is its simplicity and direct alignment with ethical values. However, it may limit investment opportunities and could lead to potential underperformance compared to less restrictive portfolios. Additionally, this approach might not encourage companies to improve their practices, as it simply avoids rather than engages with problem areas.(Borglund et al., 2021, p. 321). For example, Norway's pension fund has actively implemented negative or exclusionary screening in its investment strategy, particularly emphasizing environmental sustainability. The fund has decided to exclude companies involved in deforestation or those with serious environmental pollution, companies that manufacture tobacco products and weapons from its investment portfolio (Borglund et al., 2021).

In contrast, positive or best-in-class screening involves the selection of companies that demonstrate leadership in ESG practices within their respective sectors. This method seeks to invest in businesses that not only comply with baseline ESG criteria but also excel in their implementation of sustainable practices. While it rewards companies with top ESG performance and encourages broader sector improvements, its success depends on the availability of reliable and detailed ESG reporting and ratings. However, these can vary greatly in quality and transparency, making the selection process resource-intensive and requiring significant analyst effort (Borglund et al., 2021, p. 320–321).

ESG integration represents a more holistic approach, systematically incorporating ESG factors into traditional financial analysis. Unlike exclusionary screening, this strategy does not exclude investments based on ESG criteria but rather uses ESG data to enhance the financial analysis process. This approach offers a comprehensive view of potential risks and opportunities, potentially leading to better risk-adjusted returns. The Principles for Responsible Investment (PRI), launched in 2006 by the UNEP Finance Initiative and the UN Global Compact, provide a voluntary framework for investors to incorporate environmental, social, and governance (ESG) issues into their decision-making and ownership practices. Specifically, the first principle of the UN PRI calls for the integration of ESG issues into investment analysis and decision-making processes, highlighting the importance of considering these factors to enhance overall investment outcomes (UN, n.d.). However, the effectiveness of ESG integration is dependent on the depth and quality of ESG data integration into financial analysis, which can be both complex and resource intensive as presented earlier on ESG rating theories.

Thematic investing focuses on specific ESG themes expected to gain prominence in the future, such as clean energy, sustainable agriculture, or gender equality. This strategy targets investments in areas that align with ESG criteria and are anticipated to outperform as these themes become more central to economic and social systems (Borglund et al., 2021). While thematic investing enables investors to contribute to advancements in critical ESG areas, it may involve higher risk due to the speculative nature of emerging technologies and markets.

Impact investing focuses on directing capital to companies, organizations, or funds with the explicit goal of achieving measurable social and environmental benefits, alongside financial returns. This investment approach aims to address global challenges such as improving health, reducing poverty, or enhancing access to education through direct capital investment (Borglund et al., 2021, pp. 326–327). The emphasis on achieving these social and environmental outcomes distinguishes impact investing from traditional investments, which typically prioritize financial returns without necessarily considering the broader societal impacts. For example, according to a survey by the Norwegian National Advisory Board for Impact Investing (NorNAB), the impact investing market in Norway is robust, valued at approximately NOK 100 billion. This market encompasses investments aimed at social and environmental changes, highlighting Norway's significant role in European impact investing. Notably, Norwegian impact investors often have a global reach, with 75% holding assets abroad and 80% investing domestically, demonstrating a broad commitment to fostering positive change both within Norway and internationally (NorNAB, 2023)

In conclusion, the diverse array of ESG strategies available to investors reflects a comprehensive approach to integrating ethical considerations into financial decision-making. As these strategies continue to evolve, they offer promising pathways for aligning investments with broader societal values and sustainability goals, underscoring the critical role of ESG in shaping the future of finance.

ESG and financial performance

The relationship between Environmental, Social, and Governance (ESG) factors and financial performance has been framed through various theoretical lenses. One prominent theory is the Stakeholder Theory, which posits that companies that manage relationships with their broad set of stakeholders (not just shareholders) can achieve superior financial performance (Borglund et al., 2021). This theory suggests that ESG practices enhance trust and cooperation from stakeholders such as employees, customers, suppliers, and communities, which in turn leads to improved operational efficiencies and reduced risks. Another relevant theory is the Resource-Based View (RBV), which argues that ESG practices can be viewed as valuable, rare, inimitable, and non-substitutable resources that provide competitive advantages. For instance, sustainable practices can lead to innovations in products and processes that differentiate firms from their competitors (Borglund et al., 2021).

Empirical evidence supporting the positive impact of ESG on financial performance is substantial. Friede et al. (2015) conducted a comprehensive meta-analysis, revealing that approximately 90% of studies found a non-negative relationship between ESG and financial performance, with many indicating positive outcomes. This suggests that ESG initiatives can enhance corporate profitability and shareholder value by improving operational efficiency, enhancing brand reputation, and fostering stakeholder loyalty (Friede et al., 2015). However, not all findings are unanimous in emphasizing the direct impact of ESG components on financial metrics. Further supporting this view, Khan, Serafeim, and Yoon (2015) demonstrated that firms with robust performance on material ESG issues tend to outperform those with poor ESG performance, particularly in terms of profitability and stock price stability. This indicates that effective management of ESG factors is closely linked to financial success (Khan et al., 2015).Additionally, recent studies such as those by Fu and Li (2023) have highlighted the role of digital transformation in amplifying the benefits of ESG. They found that digital technologies enhance the positive impact of ESG on financial performance by improving efficiency and transparency, which are crucial for effective ESG integration (Fu & Li, 2023).

Despite substantial evidence supporting the positive impacts of ESG on financial performance, there remains considerable debate and divergence in findings. Some critics argue that the positive correlations found in many studies may be influenced by factors such as data quality, the specific ESG factors examined, or the methodologies used in these studies. This critique suggests that while ESG may correlate with improved financial performance under certain conditions, these findings are not universally applicable and can vary greatly depending on numerous variables (Atz et al., 2023).This view aligns with the findings from Hwang, Kim, and Jung (2021), who observed that during the COVID-19 pandemic, firms with robust ESG practices experienced less decline in earnings, suggesting that ESG activities can indeed form a critical resource in times of crisis**.**

Moreover, Jamal and Horstad's thesis indicates that individual and combined ESG factors do not significantly affect financial performance metrics such as Return on Equity (ROE). Yet, their analysis does reveal that specific aspects, particularly the social dimension of ESG along with overall ESG compliance, contribute positively to the value of Nordic firms (Jama & Horstad, 2022). This suggests that while ESG as a whole may not always correlate directly with financial metrics like ROE, certain components of ESG, especially those related to social responsibility, can enhance a firm's valuation by improving its reputation and stakeholder relationships. Furthermore, Haagensen et al. found no significant relationship in their thesis about the Impact of ESG Factors on Financial Performance. After analyzing a dataset of 31 aluminium and 173 iron & steel firms over a time period from 2012 to 2022, they found that ESG scores, ESG pillar scores, and ESG category scores have no significant impact on financial performance, suggesting that market valuations do not necessarily reflect the ESG standings of firms within these high-impact industries (Haagensen et al., 2023). This finding underscores the complexity of the ESG-financial performance relationship and highlights that the benefits of ESG may not be uniformly realized across all industries, particularly in sectors with significant environmental impacts.

These divergent findings illustrate the nuanced and complex nature of the relationship between ESG factors and financial performance. They suggest that while ESG can provide substantial benefits in terms of risk mitigation and reputation enhancement, its financial impacts are not always straightforward and can depend heavily on industry characteristics, the specific ESG dimensions emphasized, and the broader economic context. As such, firms and investors should consider these factors when evaluating the potential financial benefits of ESG initiatives. This balanced approach will enable more informed decision-making that considers both the potential benefits and limitations of ESG in enhancing corporate financial performance.

ESG ratings in sustainable finance

The growing importance of sustainable finance and the integration of ESG factors into investment decisions have led to the emergence of various ESG rating systems and methodologies. These rating systems are to provide investors and stakeholders with a standardized and comprehensive assessment of a company's ESG performance, enabling them to make informed decisions that align with their values and investment goals. As the demand for ESG ratings has increased, numerous organizations and service providers have developed their own methodologies and frameworks for evaluating and scoring companies' ESG practices. However, the lack of a universally accepted standard has resulted in significant variations and inconsistencies among these rating systems (Eccles & Stroehle, 2018). This diversity in approaches has raised concerns about the reliability, transparency, and comparability of ESG ratings, potentially hindering their effectiveness in driving sustainable investment decisions. To address these concerns, there have been calls for greater standardization and harmonization of ESG rating methodologies (Kotsantonis & Serafeim, 2019). Regulatory bodies and initiatives, such as the EU Sustainable Finance Disclosure Regulation (SFDR) and the Task Force on Climate-related Financial Disclosures (TCFD), have also supported the use of ESG ratings and disclosures (Boffo & Patalano, 2020). This regulatory support has further legitimized the role of ESG rating systems and encouraged their adoption by market participants. Despite the challenges, the proliferation of ESG rating systems reflects the growing recognition of the importance of incorporating environmental, social, and governance factors into investment strategies. As investors increasingly prioritize sustainable and responsible investing, these rating systems play a crucial role in providing the necessary information and analysis to support informed decision-making (Amel-Zadeh & Serafeim, 2018).

ESG rating systems are becoming increasingly popular with investors and companies alike. The primary purpose of ESG rating systems is to assess the sustainability of a company's operations and investments. By analysing a company's environmental impact, social responsibility, and corporate governance practices, ESG rating systems help investors make informed decisions on where to invest their money. There are several prominent ESG rating systems available, each with its own methodology and criteria. Among the most well-known vendors of ESG scores and ratings are Bloomberg ESG Data, Fitch Ratings, ISS ESG, MSCI, Moody's, Refinitiv, RepRisk, S&P Global Corporate Sustainability Assessment, and Sustainalytics. These vendors use a combination of their own proprietary scoring systems and data from other sources to evaluate companies' ESG performance. In this thesis, data from Refinitiv is utilized for analysing ESG scores, and detailed examples of how Refinitiv evaluates these scores will be presented subsequently. Before delving into Refinitiv's ESG evaluation methodology, a brief overview of the ESG evaluation methodologies employed by some other companies will be provided.

MSCI is one of the largest providers of ESG ratings, scoring roughly 8,500 companies and more than 680,000 fixed income and equity securities globally. MSCI's ESG score is based on a key issues framework that measures risk across 10 categories of environment, social, and governance areas. The environmental categories include climate change, natural capital, pollution and waste, and environmental opportunities. The social categories are human capital, product liability, stakeholder opposition, and social opportunities. The governance categories are corporate governance and corporate behaviour. MSCI assigns a score from 0 to 10 based on an issue's timeliness and probable impact, with higher weights assigned to issues with a greater potential for impact. Companies are then compared to others in the same industry and given a final rating that can range from CCC (Laggard) to AAA (Leader) (Plassholder8).

S&P Global is another major provider of ESG scores, offering assessments for over 7,300 companies worldwide. S&P Global's ESG score is based on a company's performance across 20 different ESG criteria. These criteria are selected based on their materiality to the company's industry and are divided into three pillars: environmental, social, and governance. Each criterion is scored on a scale of 0 to 100, and the scores are then weighted based on their importance to the company's industry. The weighted scores are then aggregated to calculate the overall ESG score. What sets S&P Global apart is its industry specific ESG scores, which allow for more accurate comparisons within a particular sector. This is essential as ESG issues can vary significantly from one industry to another. For example, a company in the oil and gas sector may have a higher environmental impact compared to a company in the technology sector. Therefore, industry specific ESG scores help investors to compare companies more accurately and make more informed decisions. S&P Global's ESG scores are updated annually, and they are available for both public and private companies (Global, 2021).

ESG rating systems offer several potential advantages in promoting transparency, accountability, and informed decision-making in the realm of sustainable investing. Firstly, these rating systems encourage companies to disclose relevant information and improve their environmental, social, and governance practices by assessing their achievements in these areas (Berg et al., 2022). This increased transparency can foster a culture of accountability and drive positive change within organizations. Furthermore, ESG ratings provide investors with valuable insights into a company's sustainability practices, allowing them to make investment choices that align with their values, risk preferences, and long-term goals (Amel-Zadeh & Serafeim, 2018). By considering ESG factors, investors can better identify and manage potential risks associated with environmental, social, and governance issues, which can have significant financial implications for companies (Dunn et al., 2017). Moreover, ESG rating systems contribute to the development of a common language and framework for evaluating and comparing companies' sustainability performance (Kotsantonis & Serafeim, 2019). This standardization facilitates benchmarking and allows investors to identify leaders and laggards within industries, promoting healthy competition and driving continuous improvement in sustainability practices.

Despite their strengths, ESG rating systems face several limitations that need to be addressed. One significant challenge is the lack of standardization in the methodologies and criteria used by different rating agencies(Berg et al., 2022; Dimson et al., 2020). This lack of consistency can lead to divergent ratings for the same company, undermining the reliability and comparability of the ratings. Another limitation is the reliance on the availability and quality of data provided by companies. Incomplete, inconsistent, or unaudited data can affect the accuracy of ESG ratings (Kotsantonis & Serafeim, 2019). Furthermore, the subjective nature of ESG assessments introduces the potential for biases and inconsistencies in the ratings, as different rating agencies may interpret the same information differently (Berg et al., 2022).
Additionally, some rating systems may not adequately account for industry-specific nuances and material issues, making it challenging to compare companies across different sectors (Kotsantonis & Serafeim, 2019). This limitation highlights the need for industry-specific frameworks and tailored approaches to ESG assessments.

Lastly, there are concerns about the potential for greenwashing practices, where companies present misleading or exaggerated information about their sustainability efforts (Berg et al., 2022). ESG ratings based on such information may provide an inaccurate representation of a company's true sustainability performance, undermining the credibility of the rating system.

In their comprehensive analysis, Larcker et al. (2022) critically evaluate the effectiveness of ESG ratings in conveying the non-financial impacts of companies on their stakeholders. The article examine various ESG ratings industries and express significant concerns regarding the substantial methodological variability among these ESG rating providers, which leads to inconsistencies and potential biases in the ratings (Larcker et al., 2022). They highlight the lack of a consistent definition of what ESG ratings are supposed to measure, with some providers focusing on a company's impact on stakeholders and society, while others assess the impact of societal and environmental factors on the company's financial performance (Larcker et al., 2022). They also underscore issues related to the quality and availability of data, noting that incomplete, inconsistent, or unaudited data provided by companies can undermine the reliability of ESG assessments. Furthermore, the authors point out the influence of industry and company size on ratings, suggesting that larger companies or certain industries might receive disproportionately favourable evaluations due to the weighting and aggregation methodologies employed by rating agencies. Their study discusses potential conflicts of interest that may compromise the independence of ratings, as some agencies offer consulting services to the firms they assess, raising concerns about the objectivity of their evaluations (Larcker et al., 2022). Overall, the articles by Larcker et al and others presented above underscores the significant challenges and limitations faced by the ESG rating industry, highlighting the need for improved methodologies, data quality, and transparency to enhance the credibility and effectiveness of these ratings in driving sustainable investment decisions.

The LSEG/Refinitiv ESG scoring methodology

In 2018, Thomson Reuters' finance and risk division was spun off and re-formed as Refinitiv. The company was subsequently acquired by London Stock Exchange Group (LSEG) in 2021 (Thomson Reuters, 2021). As of 2023, London Stock Exchange Group actually phased out the Refinitiv brand name and consolidated its products and services under the broader London Stock Exchange Group (LSEG) umbrella. However, since the rebranding has not been fully completed at the time of writing (flagship products such as Datastream still carry the Refinitiv name and logo), I shall use LSEG/Refinitiv in the remainder of the thesis to avoid confusion.

LSEG/Refinitiv is one of the prominent providers of ESG scores for listed companies, covering over 90% of the global market, with more than 630 ESG indicators dating back to 2002. Its ESG score is based on a company's performance across ten ESG themes, which are further divided into 186 ESG measures. The ESG themes include emissions, environmental innovation, human rights, community, and corporate governance. Each measure is scored on a scale of 0 to 100, and the scores are then weighted based on their importance to the company's industry. Additionally, LSEG also provides an ESG combined score (ESGC), which incorporates the impact of significant ESG controversies, adjusting the overall score based on negative media stories and other relevant events. This score is calculated as a weighted average of the ESG score and the ESG controversies score, with recent controversies given more weight (LSEG Data & Analytics, 2023, p. 4–5). A graphical representation of the LSEG Data & Analytics, ESG score is shown in Figure 6.

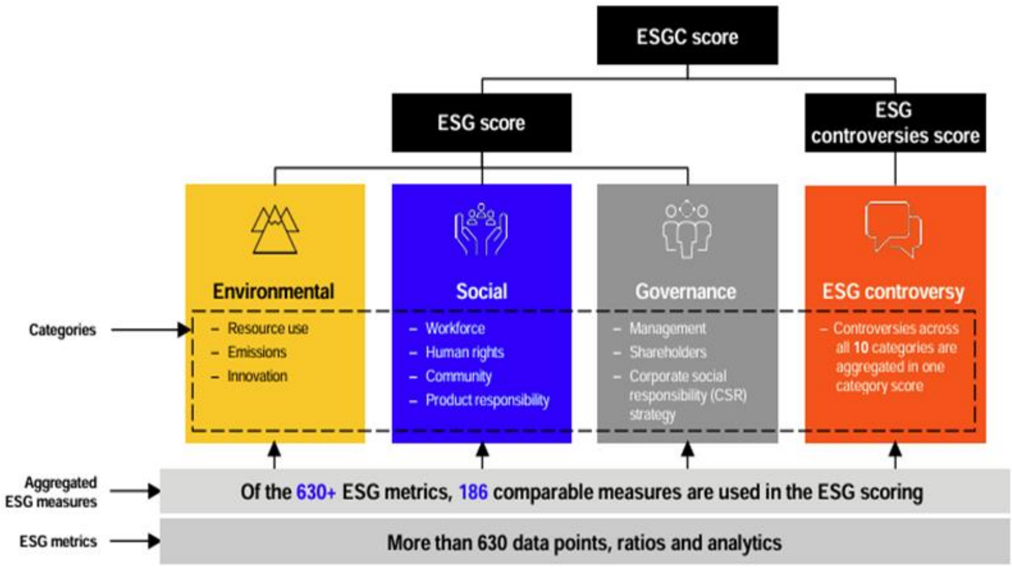

Figure 1. Graphical representation of the LSEG/Refinitiv ESG score (LSEG Data & Analytics, 2023)

Data collection process

LSEG operates one of the largest ESG content collection operations in the world, employing more than 700 specially trained content research analysts. These analysts are responsible for gathering ESG data from a variety of publicly available sources. They leverage local language expertise and operate in diverse locations around the world, adding breadth and depth to the data they collect. The data includes more than 630 ESG indicators, each of which is manually processed for every company in the LSEG ESG universe. The process of data standardization and validation is thorough. Each ESG metric is standardized to ensure that information is comparable across companies and industries. The standardization process includes the use of approximately 400 built-in error checking logic in capture tools during the data entry or preproduction stages. Additionally, during the post-production phase, approximately 300 automated quality check filters are run on the ESG capture tool to further ensure data accuracy and reliability. The ESG database is continuously updated to align with the company's reporting model. These updates may include adding new companies to the database, incorporating the latest fiscal year data, or including new controversial events. Additionally, ESG news and controversies are constantly updated as such events occur and are reported in global media. This ongoing update process ensures that the ESG scores provided by LSEG are up-to-date and reflect the latest available information (LSEG Data & Analytics, 2023, p. 6).

Score structure

The LSEG ESG Score evaluation system is designed to provide a clear, standardized way to assess and compare a company's ESG performance. The score is calculated based on a comprehensive analysis of more than 630 ESG measures, divided into 10 main categories and then summarized into three pillars: environmental, social and governance. Each of these pillars is assigned an individual score that contributes to a company's overall ESG score.
The overall ESG rating is then translated into a rating system ranging from D- to A+. The rating system enables companies to easily compare their ESG performance with competitors and gain a clear understanding of their environmental, social and governance strengths and weaknesses. Ranks are assigned based on a company's percentage ranking within its industry, ensuring rankings reflect relative performance rather than absolute metrics (LSEG Data & Analytics, 2023, p. 9–10).

Score calculation methodology

The LSEG/Refinitiv ESG scoring methodology involves a structured five-step process to evaluate a company's performance across ESG) dimensions. The first step in the scoring methodology involves calculating ESG category scores, which are the basis for assessing a company's ESG performance. This step is divided into processing Boolean and numerical data points. Boolean data points are answered with "yes", "no", or "null", where "yes" usually produces a value of 1 and "no" or "null" produces a value of 0. The polarity of each measure determines whether higher values are considered positive or negative and affects how Boolean data points are converted into numerical values used in percentile score calculations. Numerical data, on the other hand, are actual quantitative metrics reported by companies that are used to rank relative percentile within industry groups Additionally, industry group relevancy is considered, meaning some indicators are excluded from calculations if they are not relevant to a specific industry. This approach ensures that ESG category scores both accurately reflect a company's practices and are comparable to companies in its industry (LSEG Data & Analytics, 2023, p. 11-12).

The ESG Score is calculated using the percentile ranking method, which evaluates a company's performance relative to its peers across 10 categories and ESG controversies. This approach considers how many companies in each category perform worse, equally well, or have no value at all to ensure a comprehensive comparison. The approach aims to minimize the impact of outliers and achieve a balanced and fair assessment of ESG performance.

$$score = \frac{no.of\ companies\ with\ a\ worse\ value + \frac{no.of\ companies\ with\ the\ same\ value\ included\ in\ the\ current\ one}{2}}{no.of\ companies\ with\ a\ value}$$

For environmental and social categories, as well as controversies, the Thomson Reuters Business Classification (TRBC) industry group serves as the benchmark, reflecting the relevance of these issues within similar industries. Governance scores, however, are benchmarked against the country of incorporation to align with the consistency of governance practices within those regions(LSEG Data & Analytics, 2023, p. 11-12).

The second step in the scoring methodology involves the assignment of category scores which utilizes a materiality matrix. This matrix determines the weights of various category weights within each industry group are calculated through a structured process. First, the size weights for all ten-industry specific ESG categories are aggregated into a cumulative total. This total reflects the combined importance of all ESG factors relevant to the industry. The weight of each individual category is then determined by dividing its size weight by this cumulative sum as shown below(LSEG Data & Analytics, 2023, p. 12-15).

$$Category\ weight\ of\ an\ industry\ group = \frac{Magnitude\ weight\ of\ a\ category}{Sum\ of\ magnitudes\ of\ all\ categories}$$

These weights are designed to vary across industries, reflecting specific ESG concerns pertinent to each sector. This ensures that the scores are customized and relevant to different industrial contexts. The scoring system is dynamic, undergoing regular updates to incorporate the latest advancements in corporate ESG disclosures and shifts in industry standards. Such updates are essential to maintain the system's accuracy and relevance in assessing the sustainability performance of companies.

The third step in the LSEG/Refinitiv ESG methodology involves calculating an overall ESG score and individual pillar scores, which are key components in assessing a company's overall ESG performance. In this step, the category weights derived from the materiality matrix (Step 2) are used to aggregate and evaluate the company's performance in relation to the Environmental, Social and Governance (ESG) pillars. ESG pillar scores are derived by aggregating category scores within each pillar. Each pillar score is a relative sum of category weights, with the weights of the environmental and social pillars varying by industry. For the governance pillar, the weighting of each industry remains unchanged. Pillar scores provide a nuanced view of a company's performance across each ESG dimension, allowing stakeholders to identify areas of strength and opportunities for improvement (LSEG Data & Analytics, 2023, p. 16).

The fourth step in the methodology involves the calculation of *controversies scores*, which assesses the impact of ESG-related controversies on a company's overall ESG performance. The ESG Controversies Score is derived from an analysis of 23 identified ESG controversy topics and is examined in the context of the most recently completed financial year to ensure timeliness and relevance. The initial baseline value for all controversy measures is set to zero, meaning there is no controversies. This baseline allows for a simple adjustment mechanism, and any controversies discovered will result in a score drop. Companies that did not encounter any controversies during the assessment period were awarded the maximum score of 100, indicating exemplary ESG conduct (LSEG Data & Analytics, 2023, p. 17).

Like ESG scores, ESG controversy scores are normalised within industry groups to allow for "fairer" comparisons between companies. This kind of benchmarking is often considered critical because it contextualizes industry-specific risks and controversies in standards. In addition, the calculation method also takes into account adjustments for market capitalization distortions. This adjustment is important because larger companies are more likely to have their controversies come under public scrutiny due to greater visibility and wider media coverage, which can have a disproportionate impact on their ESG scores (LSEG Data & Analytics, 2023, p. 17).

The final step in the LSEG ESG scoring methodology involves calculating the ESGC score, a composite metric that combines the ESG score with the ESG controversy score. To determine the ESGC score, two scenarios are considered based on the relationship between the

ESG score and the controversies score. If the controversies score is equal to or greater than the ESG score, the ESGC score is set equal to the ESG score. This situation demonstrates that controversy is so widespread that it obscures ESG performance and has a direct impact on a company's overall sustainability rating. However, if the controversies score is lower than the ESG score, the ESGC score will be calculated as the average of the two scores. This average reflects the moderate impact of controversies on overall ESG performance, suggesting that controversies, while present, do not completely negate a company's positive ESG practices. The score reflects the dual aspects of proactive ESG engagement and reactive controversies management, providing a balanced and dynamic measure of a company's ESG posture.

ESG in unlisted companies

In the usual financial information databases, ESG scores and ESG controversy scores are only available for listed companies. For unlisted companies, the same methodology could in principle be followed that financial information providers such as LSEG/Refinitiv (see above) use for listed companies. This would naturally be a rather expensive approach if applied to many companies. To provide a “screening tool”, Kazinic and Valheim (2020) have developed an alternative approach that utilizes natural language processing of news media content to construct ESG controversy scores similar to the ones offered by Refinitiv ESG for listed companies.

Normally, the unlisted company is not strictly obliged to disclose information, unlike the listed ones. For this reason, such information is often not always relevant or even precise, which complicates obtaining accurate data on ESG factors. The basis for assessing the various ESG-disclosed information is, as I said above, often taken from publicly available information from financial reports and other sources, which is not possible because not all the necessary information is available. Thus, the discrepancy forces us to come up with alternative methods of collecting and analysing the necessary information.

One of them was the application of Natural Language Processing to unstructured data that sources of news articles, social media activity, and other available public resources. This method enabled identifying relevant ESG metrics that are not covered in formal corporate reports. Moreover, NLP is able to determine the sentiment of ESG issues, public trends and controversies concerning environmental, social, and governance principles. It enables a more dynamic and changing evaluation of a private company’s ESG performance. Another option would be direct engagement with unlisted companies. This may be accomplished by arranging tailored surveys and inviting them to participate in interviews. This typically leads to a deep understanding of how companies perceive ESG and how it is integrated into their business processes. Nonetheless, this approach is cost- and time-consuming and might not be easily scaled for a large universe of unlisted companies. Finally, third-party databases and industry benchmarks can be utilized. The former can provide "alternative" data collected from non-traditional sources which is essential to build a full image of a private company ESG profile. The later allows a comparison of unlisted companies' ESG performance with the known standard and best practices.

NLP-based ESG controversies scores

Natural language processing (NLP) is a branch of artificial intelligence that focuses on the interaction between computers and humans through natural language (KDnuggets, 2020). NLP enables machines to understand, interpret, and generate human language, allowing for more natural communication between humans and computers. In recent years, there has been a significant focus on applied NLP research, with a particular emphasis on perfecting cutting-edge techniques for various applications, including sustainability and ESG rating (Arabesque, 2021).

NLP techniques have revolutionized the analysis of media coverage of company-related ESG issues (Schimanski et al., 2024). Using NLP techniques, researchers and analysts can easily and cheaply analyse media coverage, corporate disclosures and ticker messages. Techniques such as sentiment analysis have been instrumental in detecting ESG-related terms and sentiments within textual data, enabling a more comprehensive understanding of the ESG controversies surrounding companies. Both LSEG and (Kazinic & Valheim, 2020), for instance, utilize NLP for sentiment analysis and providing ESG controversies score also for unlisted companies. As the demand for sustainable and ethical investing continues to rise, the integration of NLP in ESG analysis is poised to play a pivotal role in shaping the future of investing.

Utilizing NLP approaches for assessing Social and Governance Controversy Scores has become increasingly prevalent in the realm of ESG evaluation (Perazzoli et al., 2022). NLP algorithms, particularly those based on deep learning, play a crucial role in analyzing social and governance datasets to derive meaningful insights (Lee et al., 2022). These approaches enable the extraction of valuable information from vast amounts of unstructured text data, allowing for a more comprehensive assessment of social and governance controversies within organizations. Furthermore, the incorporation of ESG scores, which measure non-financial complementary information, can improve the accuracy of performance forecasts and risk assessments for investments, thereby promoting sustainable and responsible investment practices (Del Vitto et al., 2023). Overall, the integration of NLP techniques not only enhances the efficiency of ESG analysis but also contributes to more informed and sustainable investment strategies in the corporate landscape.

Potential validity problems

In principle, NLP-based ESG “controversy scores” (in the remainder of the thesis, I shall use this term which LSEG have introduced) could be utilised instead of the traditional rating-based ESG scores in financial analysis and portfolio management. As discussed above, this would make ESG analysis of private equity investments easier and much less costly. However, there is no published research so far about the validity of ESG controversy scores when they are used for such purposes. There are several threats to validity that may cause problems.

Spurious correlation refers to a statistical relationship where two or more events or variables are associated not because one causally affects the other, but due to either coincidence or the presence of a certain third, unseen factor (confounding variable)(Ferson et al., 2002). In the context of predictive modelling, spurious correlations can lead to misleading conclusions about the relationship between variables, such as ESG ratings and NLP-based ESG controversy scores. Spurious correlation typically arises when the variables being analyzed are each correlated with a third variable, which might not be included in the analysis. This third variable can induce or create the illusion of a correlation between the two primary variables. For example, in financial studies, time-series data might show spurious correlations if the data are non-stationary—meaning their statistical properties like mean and variance change over time (Ferson et al., 2002).

As discussed in detail above, the LSEG ESG scores are percentile rank scores that are calculated within industry peer groups, i.e. within-industry rank normalizations. According to Korenok et al. (2021), common forms of scaling and transformations in data processing can induce substantial spurious correlations, resulting in biased parameter estimates (Glasscock et al., 2021). This is particularly relevant in the context of ESG, where scaled or transformed financial and textual data from corporate disclosures might interact in unforeseen ways, highlighting the importance of rigorous statistical testing and validation to avoid drawing erroneous inferences about the relationship between ESG controversies and overall ESG scores. It is essential that researchers and practitioners to employ diagnostic tools to detect spurious correlations effectively. Korenok et al. (2021 emphasize that common diagnostic tools are often ineffective in detecting various scale effects which might mislead about the actual correlations present in the data (Glasscock et al., 2021). Hence, any analysis that aims to predict ESG ratings from NLP-based controversy scores should be accompanied by robust checks for spuriousness, ensuring that any observed relationships are genuine and not due to underlying, unaccounted-for confounders or data processing artifacts.

Simpson's Paradox is a statistical phenomenon where a trend observed in several different groups of data reverses when these groups are combined (Sprenger & Weinberger, 2021). This paradox is pivotal in understanding how aggregated data can reverse relationships apparent in subgroup analyses. Pearl (2013) describes Simpson's paradox as an instance where the relationship between a pair of variables X and Y reverses sign upon the conditioning on a third variable Z, irrespective of Z's values. This phenomenon illustrates the complexity of causal relationships and the risks of drawing conclusions from aggregated data without considering potential confounders (Pearl, 2014). The relevance of Simpson’s paradox in predictive modelling, particularly in sectors like ESG ratings predicted from NLP-based controversy scores, cannot be overstated. For example, an ESG model might show a positive relationship between environmental controversy scores and overall ESG ratings within specific industries such as energy or utilities but could reverse when data from all industries are aggregated. This might occur due to industry-specific factors that, when not accounted for, lead to misleading conclusions about the general effectiveness of ESG practices across all sectors.

Industry sector as a confounding variable

A confounding variable is one that influences both the dependent variable and independent variable, causing a spurious association (Thomas, 2020). In the context of predictive modelling for ESG ratings, the industry sector can serve as a significant confounding variable. Different industries have different characteristics and regulatory environments that can influence both ESG performance and its perception or measurement through controversy scores derived from natural language processing (NLP) techniques. When industry sector is not adequately controlled for, it can introduce bias and affect the accuracy of predictive models, particularly in identifying best-in-class strategies for portfolio management. The failure to account for industry sector as a confounding variable may lead to misinterpretation of results and suboptimal decision-making processes,

Research questions

ESG scores exist in many forms and are sold by many financial information providers. One frustration many analysts complain about is the poor convergent validity of ESG scores from different providers. Differences in underlying indicators, differences in peer group definitions and differences in normalisation approaches have been made responsible for this(Berg et al., 2022). Peer group definitions, including those used in the normalisation of scores, are important because the materiality of ESG issues differs between industries. This does not only concern classical ESG scores based on annual report data and analyst ratings, but also ESG controversies scores based on natural language processing of ticker messages or news media content.

Since ESG controversies scores are often proposed as a cheaper and more broadly applicable alternative to classical ESG scores, this leads to the question whether the predictive validity of ESG controversy scores with respect to classical rating-based ESG scores is invariant across industries. The aim of the research presented in this thesis is to explore this problem. Specifically, the research will address three questions:

- How well can classical rating-based ESG scores be predicted from NLP-based ESG controversy scores?
- Does the predictive relationship differ by industry?
- How consistent are the effects when compared across different approaches for constructing ESG controversy scores and rating-based ESG scores)

Two studies will be conducted. Study 1 will use data for OSX-listed companies, extracted from LSEG/Refinitiv Datastream. Study 2 will focus on unlisted companies, re-analysing a data set developed by Kazinic and Valheim (2020).

Study 1: ESG controversy scores extracted from LSEG/Refinitiv ESG

Method

Asset selection

The asset selection process involved identifying Norwegian companies listed on the Oslo stock exchange (OSE). The asset selection process is a critical step in ensuring the robustness of the study on ESG scores and ESG controversy scores for Norwegian listed companies. Initially, the dataset extracted from LSEG Datastream (for more information, see next section) included over 95 companies, but due to the prevalence of missing ESG controversy score data (i.e., years where there was no media coverage on the controversy topics monitored by LSEG ESG), it was necessary to refine the dataset to maintain the integrity of the analysis. This refinement process resulted in a reduced set of 26 companies with ESG scores and ESG controversy scores in at least one reporting year.

These 26 companies represent a diverse cross-section of industries, including Aluminium, Banks, Food Processing, Integrated Oil & Gas, Oil & Gas Exploration and Production, Oil-Related Services and Equipment, Wireless Telecommunications Services, and Miscellaneous sectors (see Appendix 1 for details). This diversity is crucial as it provides a comprehensive overview of the ESG performance across various segments of the Norwegian market, allowing for a more nuanced analysis of ESG scores in relation to ESG controversies and industry. The inclusion of companies from a range of industries also enables the study to explore the potential differences in ESG performance and controversies across sectors. This is particularly important given that ESG factors can have varying levels of relevance and impact depending on the industry context. For instance, the environmental criteria might be more pertinent to the Oil & Gas sector, while social criteria could be more significant for the Banking sector.

ESG rating data

The ESG rating data, the cornerstone of this study, was extracted from LSEG Datastream. This platform, previously operating under the names Thomson Reuters ESG and Refinitiv ESG, is renowned for its comprehensive, transparent, and objective approach to evaluating companies' performance across environmental, social, and governance (ESG) criteria. The data encompassed a range of scores, including the ESG Score, ESG Controversies Score, and ESG Combined Score, for each company over an up to seven-year period from 2015 to 2022. In addition to these scores, the industry code for each company was retrieved, which is pivotal for segmenting the dataset into relevant sectors. This classification is crucial for the study as it enables a sector-specific analysis, allowing for the examination of ESG performance within the context of industry-specific challenges and norms.

Predictive modeling

The predictive modelling undertaken in this study was methodically executed using JMP Pro Version 17.2. The initial model predicted ESG scores by solely from ESG controversies scores, deliberately excluding industry segmentation to establish a baseline understanding. A second, more nuanced model was specified, incorporating industry-specific considerations to capture the influence of sectoral factors on ESG performance. This two-way approach let us look at a basic model that didn't consider what industry a company was in and compare it with a more detailed model that did. The detailed model was designed to understand how different types of industries might affect a company's ESG scores. This meant we could see if the industry a company was in changed the relationship between its ESG scores and its controversies, compared to just looking at the controversies alone. This comparison helps us see how important it is to think about the industry when we're trying to predict ESG scores.

Results

To assess how well the LSEG ESG controversies score predicted the actual LSEG ESG score in the sample of OSX-listed companies, two models were specified and estimated by ordinary least squares.

Baseline model

The first model, which predicted ESG scores based solely on ESG controversies scores (without controlling for industry), revealed a significant but in absolute terms rather weak relationship between these two variables. Coefficients, standard errors and statistical tests are shown in Table 1. The regression plot is shown in Figure 2. The ESG controversy score was a significant predictor of the ESG score. However, the R^2 value of .07 from this model indicates that, in the pooled sample, a mere 7% of the variance in ESG scores could be explained by controversies alone, pointing to the presence of other influential factors, for example industry-moderated effects, that were not captured by this model.

The Root Mean Square Error (RMSE) measures the average magnitude of the model's prediction errors, indicating that the predicted ESG scores deviated from the actual scores by approximately 14.49 points on average. This level of inaccuracy underscores the model's limited predictive power and suggests that relying on ESG controversies scores to predict ESG performance based on data that are pooled across industries may not yield precise results.

Table 1. Parameter estimates, Model 1 (including only an overall effect of the ESG controversy score)

Term	Estimate	*SE*	*t*	*p*	Lower 95% CI	Upper 95% CI	Std. beta
Intercept	75.80	5.06	14.97	.000	65.63	85.96	.00
ESG Controversies Score	-0.18	0.09	-2.02	.048	-.36	.00	-.27
R^2	.07						

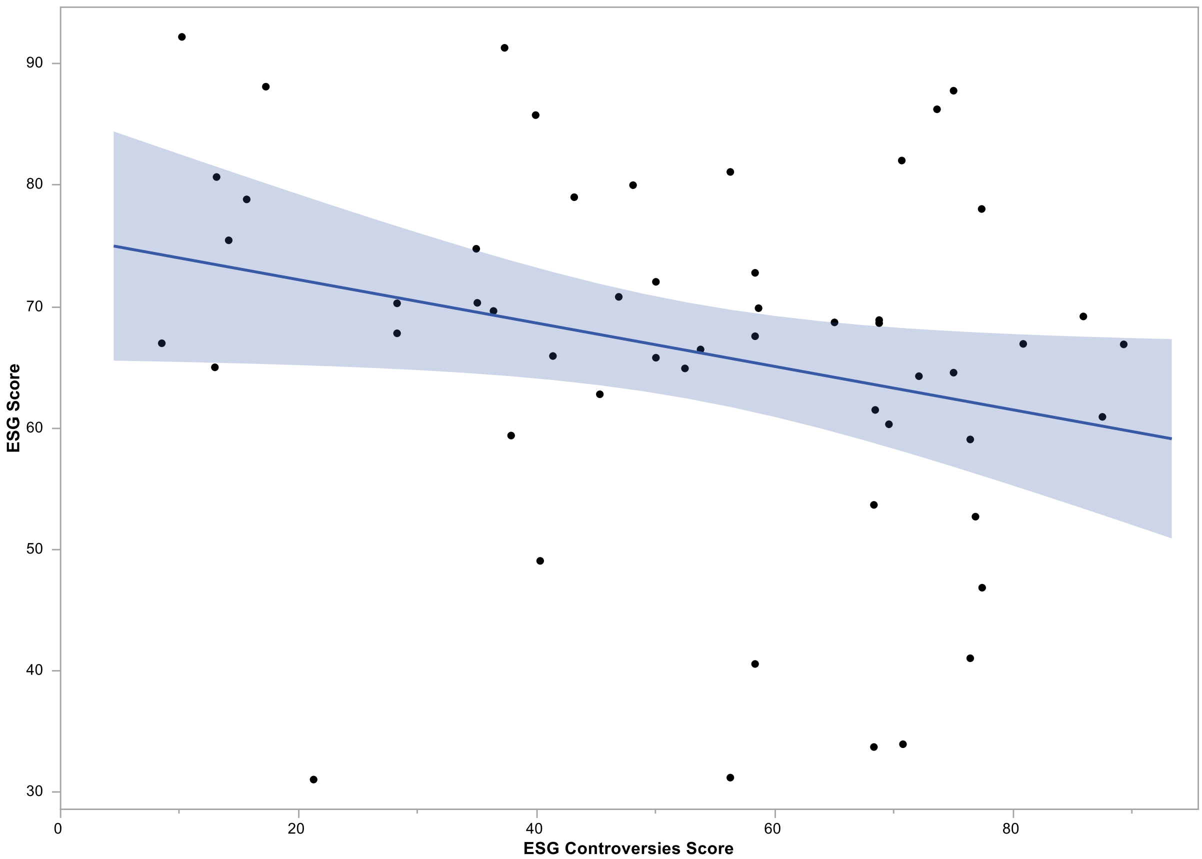

Figure 2. Regression plot, baseline model (Model 1)

Table 2. Parameter estimates, Model 2 (including industry-specific effects of the ESG controversy score)

Term	Estimate	*SE*	*t*	*p*	Lower 95% CI	Upper 95% CI	Std. beta
Intercept	65.37	2.01	32.56	.000	61.30	69.43	.00
Industry: Aluminum	22.86	5.77	3.96	.000	11.18	34.55	.70
Industry: Banks	-9.85	5.42	-1.82	.077	-20.81	1.12	-.30
Industry (regrouped)[Food Processing]	-4.56	5.03	-.91	.370	-14.73	5.62	-.14
Industry (regrouped)[Integrated Oil & Gas]	14.36	4.25	3.38	.002	5.76	22.95	.49
Industry (regrouped)[Miscellaneous]	-10.32	3.46	-2.98	.005	-17.32	-3.31	-.40
Industry (regrouped)[Oil & Gas Exploration and Production]	-15.51	6.76	-2.30	.027	-29.19	-1.83	-.42
Industry (regrouped)[Oil Related Services and Equipment]	3.41	3.71	.92	.363	-4.09	10.91	.12
Industry (regrouped)[Aluminum]:(ESG Controversies Score-52.78)	-.05	.22	-.22	.829	-.50	.40	-.03
Industry (regrouped)[Banks]:(ESG Controversies Score-52.78)	-.40	.23	-1.73	.092	-.87	.07	-.20
Industry (regrouped)[Food Processing]:(ESG Controversies Score-52.78)	.03	.19	.17	.864	-.35	.41	.02
Industry (regrouped)[Integrated Oil & Gas]:(ESG Controversies Score-52.78)	.03	.16	.19	.852	-.30	.36	.02
Industry (regrouped)[Miscellaneous]:(ESG Controversies Score-52.78)	-.10	.17	-.58	.564	-.45	.25	-.06
Industry (regrouped)[Oil & Gas Exploration and Production]:(ESG Controversies Score-52.78)	.18	.36	.48	.631	-.56	.91	.06
Industry (regrouped)[Oil Related Services and Equipment]:(ESG Controversies Score-52.78)	-.10	.26	-.38	.703	-.62	.42	-.04
Industry (regrouped)[Wireless Telecommunications Services]:(ESG Controversies Score-52.78)	.06	.36	.18	.861	-.66	.79	.03
R^2	.60						

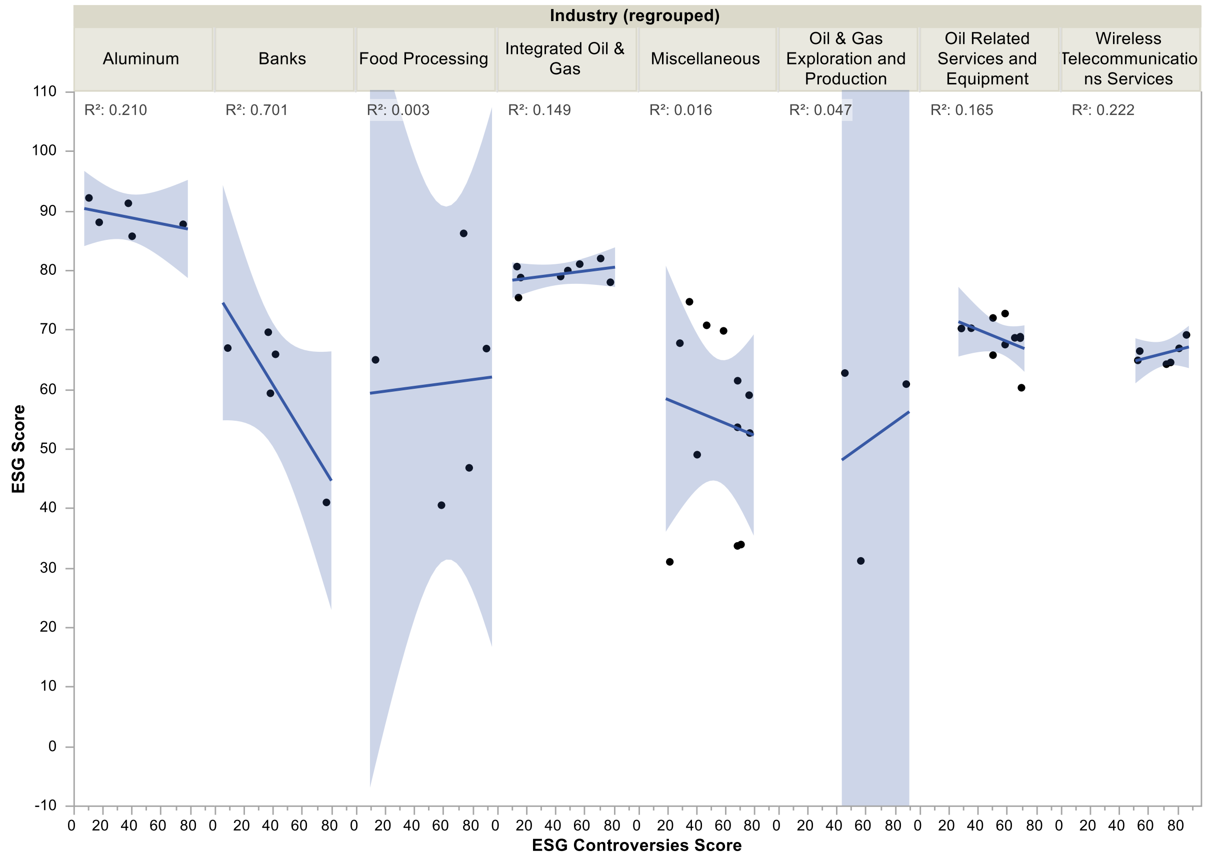

Figure 3. Regression plot, industry-specific model (Model 2)

Model with industry-specific effects

A second model was estimated that allowed the relationship between the ESG controversies score and the actual ESG score to be moderated by industry. Parameter estimates, standard errors and statistical tests are shown in Table 2. The regression plot of the industry-specific effects is shown in Figure 3. The model revealed strongly heterogeneous predictive validities of the ESG controversies scores across sectors, with relationships not only fluctuating in strength but also in direction. The RMSE improved to 11.20, indicating enhanced predictive accuracy, and the R^2 value increased substantially to .60, indicating that 60% of the ESG scores' variability could be explained by the model when industry classification is taken into account. There was strong predictive validity in the banking sector (within-industry R^2 = .70), whereas the there was no relationship in food processing sector (within-industry R^2 = .00).

To gauge the relative size of the global effect of the ESG controversies score, global differences between industries, and industry-specific effects of the ESG controversies score, analysis of variance (ANOVA) was conducted. In the context of this model, the global effect of the ESG controversies score became insignificant (main effect: $F[1,38] = .23$, $p = .63$, effect size $\eta^2 = .00$), global differences between industries were highly significant (main effect: $F[7,38] = 6.18$, $p < .001$, effect size $\eta^2 = .46$), and the industry-specific effects of the ESG controversies score were insignificant (interaction effect: $F[7,38] = .47$, $p = .85$, effect size $\eta^2 = .04$).

The effect sizes suggest that, in this particular sample of OSX-listed companies and in the years for which actual ESG scores where available from LSEG/Refinitiv, global differences between industries were responsible for almost all (92%) of the explainable variation in the ESG scores of the companies, whereas the ESG controversies score only contributed 8% of the explainable variation. If the LSEG/Refinitiv ESG controversies score were to be used as a replacement of the actual ESG score in financial analysis and portfolio management, such a low level of predictive validity would clearly be unacceptable.

Use of Artificial Intelligence in the thesis

In the course of writing this master's thesis, I utilized ChatGPT(KI-Chat (Sikt), developed by OpenAI, as a tool for language editing. ChatGPT assisted in improving the overall language quality by refining sentence structure, grammar, and flow to ensure clarity and readability throughout the thesis.

Study 2: ESG controversy scores constructed for private equity assets

In Study 1, the moderating effects of industry on the predictability of company ESG scores based on NLP analyses of media coverage of the respective companies were investigated using LSEG/refinitiv Datastream data, i.e. in a universe of large companies listed on the OSX stock market. Study 2 will focus on unlisted companies, re-analysing a data set developed by Kazinic and Valheim (2020).

Method

This section employs a methodological framework developed for constructing Environmental, Social, and Governance (ESG) controversy scores for private equity assets. Inspired by the work of Kazinic and Valheim (2020), who used natural language processing (NLP) and predictive modelling to generate ESG ratings for unlisted companies in their Master of Science in Business Administration thesis at NMBU, our approach not only leverages but also adapts their methodologies to the specific context of private equity assets within the Norwegian market. This study extends their approach by initially evaluating the predictability of actual ESG scores from controversies scores without industry segmentation, and then extending the model with industry-specific effects to assess the impact of industry factors on ESG performance, mirroring the analytical approach taken in Study 1.

Asset selection

This section reports on the asset selection methodology as developed and implemented by Kazinic and Valheim (2020) in their study of ESG ratings for unlisted companies. Their method involved collecting and preparing data from diverse online sources, including news articles, reports, and reviews found in online communities. These sources provided a comprehensive set of unstructured textual data, which was essential for the development of their machine learning algorithms.

According to Kazinic and Valheim, the textual data required extensive preprocessing to identify relevant information for statistical analysis, a practice supported by the methodologies described in Netzer et al. (2012). Additionally, they incorporated accounting data from Proff

Forvalt, focusing on a carefully selected group of 102 Norwegian unlisted companies. This combination of textual and financial data was integrated into two distinct datasets, which were then merged to facilitate a regression-based analysis aimed at quantifying ESG ratings.

The criteria for selecting these companies included having a workforce between 15 to 100 employees and being founded within the decade from January 1, 2010, to January 1, 2020. This focus reflects the prevalence and significance of medium-sized companies in the Norwegian market, as noted by Statistics Norway (2020). To further ensure a representative sample and minimize selection bias, the study limited the selection to about six companies per industry, prioritizing those that showed the highest economic growth from 2016 to 2018, based on revenue. This methodical selection resulted in a total of 102 companies from multiple industries, as detailed in Appendix 2.

ESG rating data

In their study, Kazinic and Valheim (2020) established a robust framework for evaluating the ESG performance of unlisted companies, an essential precursor to their analytical assessment. This section leverages the criteria and methodologies they developed, which underpin the ESG ratings used in their research. Before the commencement of data collection, the researchers prepared six ESG criteria, carefully chosen to be applicable universally across all companies, regardless of industry, geographical location, or size. These criteria were grouped into three broad categories: environmental, social, and governance.

Environmental criteria were comprised of 'Product or Service' and 'Negative Publicity'. The former evaluated the environmental impact of the company’s offerings, including their potential for reducing emissions and the environmental risks posed during production. The latter criterion assessed how environmental issues related to the company were reported in the media, focusing on incidents like pollution or destruction of nature.

In the social domain, the criteria focused on examining the company's efforts to ensure employee health, safety, and minimize labor disputes, as well as the adequacy of safety rules and training programs. Furthermore, the relationship between managerial and employee salaries, the proportion of full-time and part-time employees, and the fairness of employment practices were analyzed. The company's involvement in supporting local communities, charities, or encouraging employee volunteer work was also evaluated.

The governance criterion centered on assessing the diversity and structure of the company's board, including the presence of women, separation of the Board Chairman and CEO roles, and whether the company was family-owned.

To provide a target variable for the training of their machine learning model, Kazinic and Valheim independently rated each company to ensure objectivity and reduce potential biases or misunderstandings. Companies were rated on a scale from -5 to 5, with specific guidelines for each score level, allowing for the differentiation of achievement and performance across the ESG criteria, capturing both positive and negative aspects.

After completing the independent ratings, the average score for each company was calculated, resulting in a final ESG score. To ensure consistency and reliability, the inter-rater reliability for each criterion was calculated, ranging from .55 to .84, with the "Wage" criterion being the most consistent.

Media reporting data

Kazinic and Valheim opted for manual retrieval of textual data over automated methods such as web crawling and web scraping. The data were collected by manually copying snippets of text documents, which were then translated into English and compiled into an Excel sheet for further analysis. Data collection was discriminative as it was based on the company's media visibility. For companies with significant media coverage, they limited their data collection to the past three to four years to ensure relevancy. For others, the collected media reports did not exceed ten years, trying to ensure a balance between comprehensive data gathering and current relevance.

They categorized the collected documents as either positive or negative with respect to ESG factors. This classification facilitated a more straightforward analysis, allowing them to identify and highlight specific terms and phrases indicative of positive or negative ESG behaviors. Through this process, they aimed to discern underlying patterns in the text that could inform their predictive modeling. Notably, out of the total text documents collected, approximately 8% were classified as negative, indicating prevalent ESG concerns within the sampled documents.

This comprehensive approach to asset selection, data collection, and manual ESG rating aimed to develop a robust and reliable dataset that would enable accurate forecasting of ESG controversy scores for private equity assets. The methods for extracting, collecting, and evaluating data are graphically represented in Figure 4.

Figure 4. Graphical Representation of the data preparation process (adapted from Kazinic and Valheim, 2020, p. 28)

NLP-based feature engineering

In their study, Kazinic and Valheim (2020) employed NLP techniques for feature engineering. They created a document term matrix (with TF-IDF weighting), which was the primary feature engineering effort executed in their analysis.

ESG controversies score

The closest equivalent to an ESG controversies score in the Kazinic and Valheim (2020) study is the predicted ESG score produced by their machine learning model. The model was a random forest with the TF-IDF weighted entries of the document-term matrix as input features. Note that this type of controversies score is optimally weighted for the prediction of the ESG scores that had been used as training data in the modelling (unlike the Refinitiv/LSES controversies scores, which do not have this property).

Predictive modeling

The predictive modelling approach in this study was the same as in Study 1. The initial model predicted company-level ESG scores exclusively from ESG Kazinic and Valheim's (2020) equivalent of controversy-scores. The second model included industry-specific effects of the controversy score.

Results

Baseline model

In the initial analysis, the baseline predictive model was constructed without considering industry-specific factors. This model included the ESG controversy score as the sole predictor. Parameter estimates, standard errors and statistical tests are shown in Table 3, the regression plot in Figure 5.

Table 3. Parameter estimates, Model 1 (including only an overall effect of the ESG controversy score)

Term	Estimate	*SE*	*t*	*p*	Lower 95% CI	Upper 95% CI	Std. beta
Intercept	-.98	.06	-17.12	.000	-1.09	-.86	.00
ESG Controversies Score	1.85	.04	42.80	.000	1.76	1.93	.79
R^2	0.62						

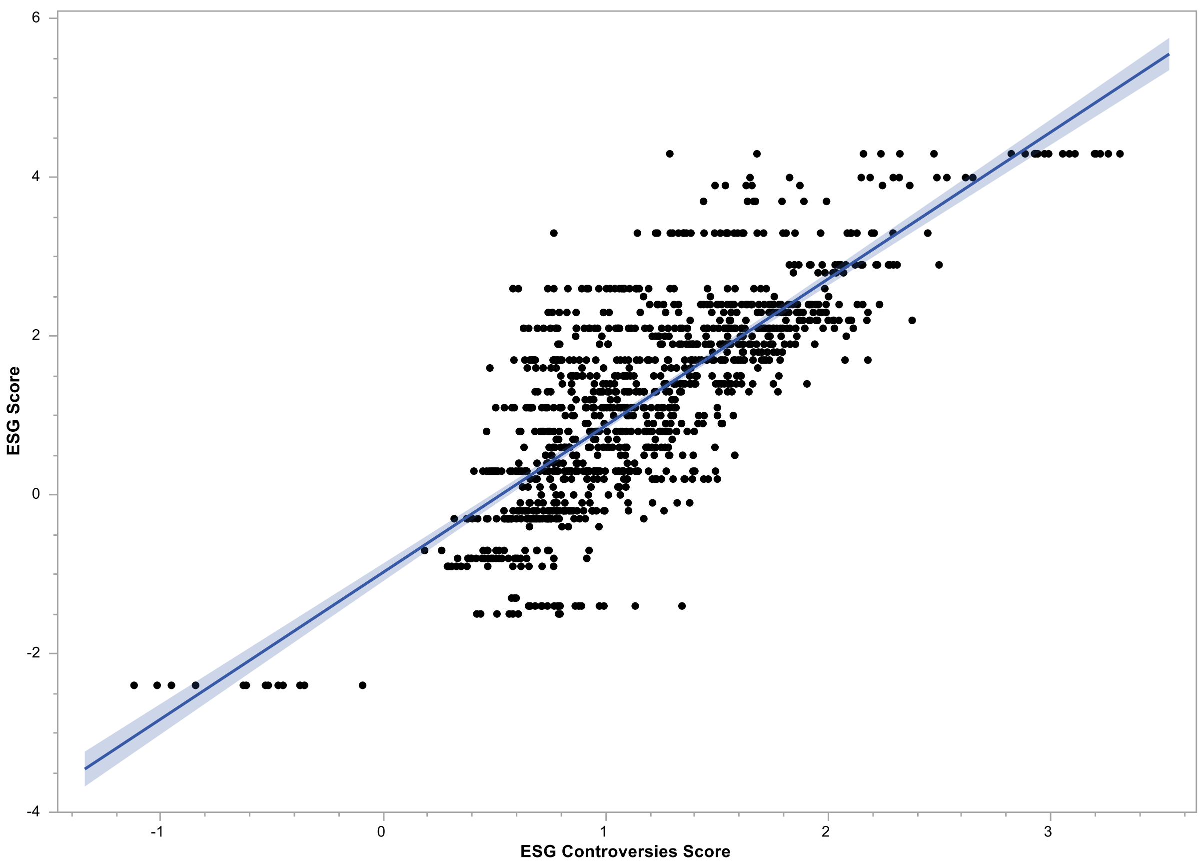

Figure 5. Regression plot, baseline model (Model 1)

The baseline model yielded an R^2 value of .62, which indicates that the controversies score alone accounts for approximately 62% of the variance in the actual ESG score. The RMSE was .81. Compared to the results of Study 1, the baseline model results from Study 2 indicated a much stronger overall relationship between the controversies score and the actual ESG score.

Model with industry-specific effects

A more complex model, Model 2, was then estimated that allowed industry-specific effects in the prediction of ESG scores. Parameter estimates, standard errors and statistical tests are shown in Table 4. The regression plot is displayed in Figure 6. The R^2 value of the model increased to .76, indicating that 76% of the variance in the overall ESG score can be explained by the predictors included in the model. The RMSE dropped to .66, indicating more accurate predictions compared to the baseline model.

Table 4. Parameter estimates, Model 2 (including moderator effects by industry)

Term	Estimate	*SE*	*t*	*p*	Lower 95% CI	Upper 95% CI	Std. beta
Intercept	1.31	.03	46.79	.000	1.26	1.37	.00
Industry[A]	-.02	.13	-.16	.870	-.27	.23	.00
Industry[B]	-.28	.10	-2.67	.008	-.49	-.07	-.07
Industry[C]	-.29	.06	-4.58	.000	-.41	-.16	-.09
Industry[D]	.87	.09	9.34	.000	.69	1.05	.27
Industry[E]	.19	.08	2.31	.021	.03	.35	.05
Industry[F]	-.87	.10	-8.78	.000	-1.06	-.67	-.21
Industry[G]	-.58	.12	-5.07	.000	-.81	-.36	-.17
Industry[H]	-.05	.14	-.38	.702	-.33	.22	-.01
Industry[I]	.21	.11	2.02	.044	.01	.42	.06
Industry[J]	.14	.08	1.66	.097	-.03	.30	.04
Industry[K]	.97	.13	7.65	.000	.72	1.21	.23
Industry[L]	.01	.16	.07	.946	-.30	.32	.00
Industry[M]	.42	.10	4.06	.000	.22	.63	.11
Industry[N]	-.32	.11	-2.83	.005	-.54	-.10	-.07
Industry[Q]	.73	.10	7.14	.000	.53	.93	.17
Industry[R]	-.84	.13	-6.20	.000	-1.10	-.57	-.18
Industry[A]:(ESG Controversies Score-1.19)	1.91	.21	9.24	.000	1.50	2.31	.16
Industry[B]:(ESG Controversies Score-1.19)	1.57	.24	6.59	.000	1.10	2.04	.11
Industry[C]:(ESG Controversies Score-1.19)	1.75	.08	23.02	.000	1.60	1.90	.35
Industry[D]:(ESG Controversies Score-1.19)	.46	.17	2.69	.007	.13	.80	.06

Industry[E]:(ESG Controversies Score-1.19)	.58	.23	2.55	.011	.13	1.02	.04
Industry[F]:(ESG Controversies Score-1.19)	2.63	.39	6.73	.000	1.86	3.40	.11
Industry[G]:(ESG Controversies Score-1.19)	1.75	.23	7.52	.000	1.29	2.21	.20
Industry[H]:(ESG Controversies Score-1.19)	2.35	.34	6.82	.000	1.68	3.03	.17
Industry[I]:(ESG Controversies Score-1.19)	2.01	.22	9.17	.000	1.58	2.44	.21
Industry[J]:(ESG Controversies Score-1.19)	.91	.19	4.89	.000	.55	1.27	.08
Industry[K]:(ESG Controversies Score-1.19)	.98	.20	4.88	.000	.59	1.37	.10
Industry[L]:(ESG Controversies Score-1.19)	1.52	.45	3.38	.001	.64	2.40	.09
Industry[M]:(ESG Controversies Score-1.19)	1.45	.10	14.01	.000	1.25	1.65	.27
Industry[N]:(ESG Controversies Score-1.19)	1.25	.47	2.64	.008	.32	2.17	.04
Industry[Q]:(ESG Controversies Score-1.19)	2.93	.22	13.49	.000	2.50	3.35	.20
Industry[R]:(ESG Controversies Score-1.19)	1.40	.47	2.97	.003	.48	2.33	.05
Industry[S]:(ESG Controversies Score-1.19)	1.88	.24	7.69	.000	1.40	2.36	.17
R^2	0.76						

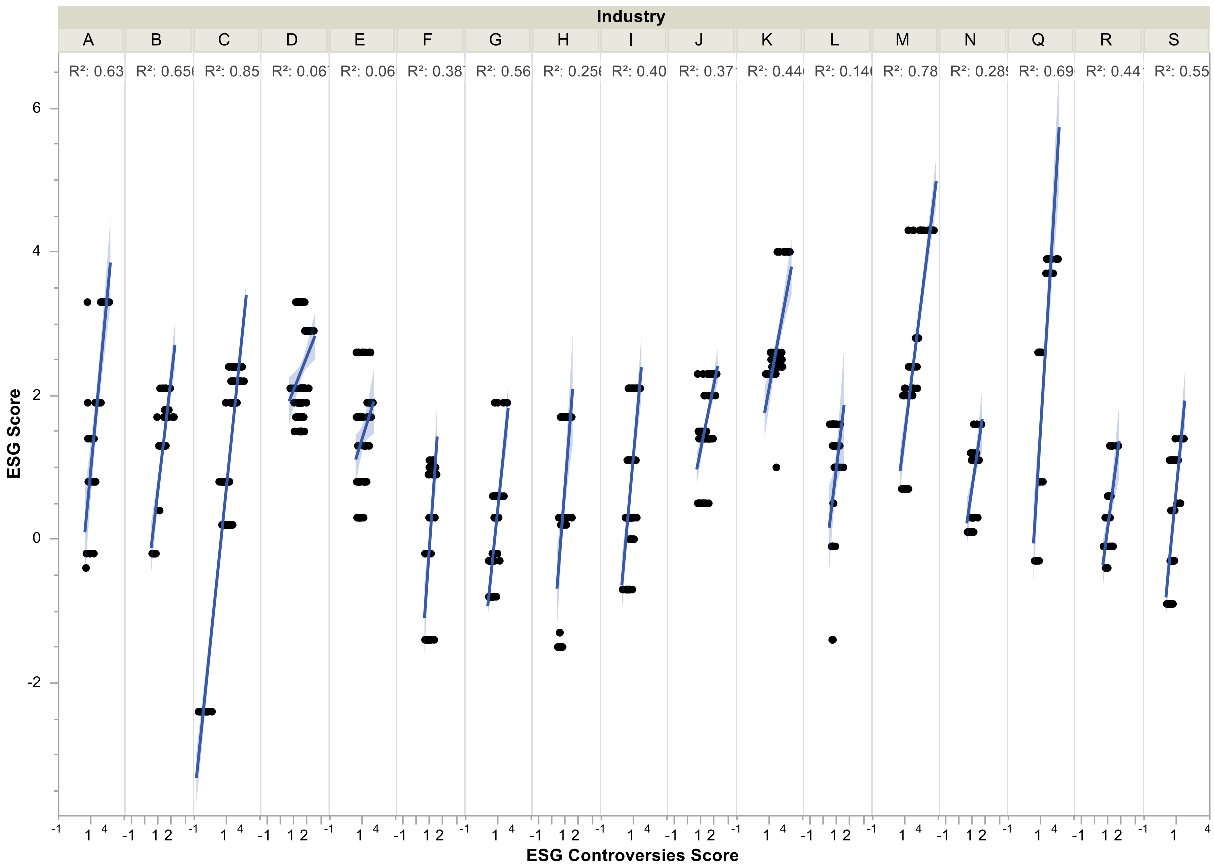

Figure 6. Regression plot, industry-specific model (Model 2)

In Figure 6, each line represents the regression fit for a different industry. The spread of data points around each line varies, indicating different degrees of consistency across industries regarding the prediction of the ESG score from the controversies score. Some industries have data that is densely packed along the regression line; this suggests that consistent, predicted ESG performance is present. Other industries have a more spread out pattern, which suggests greater variability and uncertainty in the ESG results.

In the same way as in Study 1, ANOVA was conducted to gauge the relative size of the global effect of the controversies score, global differences between industries, and industry-specific effects of the controversies score. In the context of this model, the global effect of the controversies was highly insignificant (main effect: $F[1,1079] = 533.30$, $p < .001$, effect size $\eta^2 = .12$), global differences between industries were also highly significant but smaller in size (main effect: $F[16,1079] = 24.51$, $p < .001$, effect size $\eta^2 = .09$), and industry-specific effects of the controversies score were highly significant as well but again smaller in size (interaction effect: $F[16,1079] = 9.32$, $p < .001$, effect size $\eta^2 = .03$).

The effect sizes indicate that, in the Kazinic and Valheim (2020) data of unlisted companies, the controversies score could predict for half of the explainable variation (50%) in actual ESG scores, whereas global industry differences accounted for 36% and industry-specific relationships between the controversies score and the ESG score for 14%. This suggests that the controversies scoring method developed by Kazinic and Valheim is substantially better than what was found in Study 1 for the LSEG/Refinitiv ESG controversies score, and suffers much less from industry bias, but is still subject to considerable noise. If Kazinic and Valheim's controversies scores were to be used as a replacement of actual ESG scores in financial analysis and portfolio management, the predictive validity of the controversies scores would still not be fully acceptable. This corroborates the recommendation by Kazinic and Valheim (2020, p. 58) that their approach can be very useful as a screening tool but cannot replace deeper analysis of a company's actual ESG profile.

General discussion

The exploration of the predictability of ESG scores from NLP-based controversies scores, as conducted in Studies 1 and 2, provides a nuanced understanding of the relationship between controversy scores and ESG performance across different industries. These studies, utilizing datasets from both listed and unlisted companies, offer insights into how NLP-based controversy scores can serve as predictors of ESG scores, and how this predictability varies with the inclusion of industry-specific effects.

Predictability of ESG scores from NLP-based controversy scores: key results

In Study 1, the focus was on Norwegian companies listed on the Oslo Stock Exchange (OSE), with ESG controversy scores extracted from Refinitiv ESG. The initial model, which did not account for industry effects, revealed a relatively weak relationship between ESG controversy scores and ESG scores, with an R^2 value of .07. This indicates that, in the pooled sample, only 7% of the variance in ESG scores could be explained by controversy scores alone. This highlights the limited predictive power of controversy scores without considering industry nuances, suggesting the presence of other influential factors not captured by this model. A subsequent model incorporated industry-specific effects, leading to a significant improvement in predictability. The R^2 value increased to .60, indicating that 60% of the variance in ESG scores could be explained when industry classification was considered. This model demonstrated strong predictive validity in certain sectors, such as the banking sector, which had a within-industry R^2 value of .70, whereas no significant relationship was found in the food processing sector. This substantial improvement underscores the importance of contextual industry factors in ESG score predictability. The variance in results across different industries suggests that industry characteristics heavily influence the relationship between controversy scores and ESG ratings.

Study 2 extended the analysis to unlisted companies, re-analyzing a dataset developed by Kazinic and Valheim (2020). Similar to Study 1, the initial model without industry segmentation yielded an R^2 value of .62, suggesting that controversies scores alone accounted for 62% of the variance in the actual ESG score. This model, while providing a reasonable level of predictive power, indicated that incorporating industry-specific factors could still moderate the model's performance. The introduction of industry-specific moderators in the predictive model significantly improved its explanatory power. The R^2 value increased to .76, demonstrating that 76% of the variance in ESG scores could be explained by the model when industry-specific factors were considered. Both studies reveal that industry-specific effects are pivotal in enhancing the predictability of ESG scores from controversy scores. The notable difference in R^2 values with and without industry effects in Study 1 illustrates how industry characteristics can modulate the impact of controversy scores. The overall predictive power in both studies suggests that while NLP-based controversy scores can provide significant insights into ESG performance, their validity is dependent on industry context.

The difference in R^2 values between the two studies, particularly in the baseline models, may be attributed to the methodological approach in constructing the controversies scores. In Study 2, the controversy scores were optimally weighted based on the predictive power of the NLP features, which likely contributed to the overall better predictability compared to Study 1, where standard ESG controversy scores from LSEG/Refinitiv were used. This finding aligns with the theoretical perspective that emphasizes the importance of tailored approaches in ESG assessment, considering the specific characteristics and challenges of different industries and asset classes.

Size of differences between industries

The differences in the predictive relationships between NLP-based ESG controversy scores and actual ESG ratings across industries varied significantly between the two studies. These variations offer critical insights into how industry characteristics influence the application and effectiveness of NLP techniques in ESG assessments. In Study 1, which focused on Norwegian companies listed on the Oslo Stock Exchange and utilized ESG scores and ESG controversies scores extracted from LSEG/Refinitiv Datastream, the differences between industries were found to be much larger, with relationships even exhibiting different signs: certain industries showed positive relationships between controversies scores and ESG ratings, while others displayed negative relationships. This shows that the same controversies score might indicate different levels of ESG compliance depending on the industry context. For instance, the banking sector showed strong predictive validity, with a within-industry R^2 value of .70, suggesting a significant relationship between ESG controversy scores and ESG scores within this industry. Conversely, no significant relationship was found in the food processing sector, highlighting the stark differences in predictability across industries. The substantial variability observed in Study 1 underscores the complexity of ESG performance across different sectors. The differences in the industry-specific means as well as in the signs and magnitudes of the relationships between ESG controversy scores and ESG scores suggest that industry-specific factors play a crucial role in determining ESG performance. This variability could be attributed to the unique environmental, social, and governance challenges and public sensitivity to ESG issues.

Study 2, examining ESG controversy scores for private equity assets, found smaller industry differences than anticipated. Although the slopes of the relationships varied in magnitude, their direction remained consistent across industries. The smaller differences could be attributed to the methodology used to construct the ESG controversy scores, which may have been more sensitive to capturing industry-specific nuances. Additionally, the focus on private equity assets, which may operate under different dynamics compared to publicly listed companies, could contribute to the observed uniformity in predictive relationships across industries.

The role of within-industry normalization of ESG scores

One reason for the generally much lower predictability in Study 1 may be the percentile rank normalisation underlying the Refinitiv ESG scores. The normalisation makes the scores ordinal, i.e. the sizes of differences between any two pairs of ESG scores are not metrically comparable. This may have diluted the strength of the relationships, and theoretically it may even have created spurious effects (when very small differences were rank-normalised into large differences) or hidden effects that were substantial (when very large differences were rank-normalised into small differences). Ordinal data, by its nature, simplifies the underlying metrics into categories that do not necessarily hold the same value across different contexts. This simplification can obscure genuine variations and nuances in data, which are critical for accurate prediction and analysis (Eccles & Stroehle, 2018). Furthermore, inconsistencies may arise when comparing companies from various sectors due to the normalization practices within industries. This becomes a real challenge in a global market where stakeholders and investors seek to evaluate ESG performance across different industry boundaries. The normalization approach assumes that the significance of ESG issues remains consistent within an industry but differs across industries. However, this is not always the case, especially in today's complex and interconnected global markets.

Unfortunately, the Refinitiv system cannot create ESG scores that are not normalized; this is due to the way they are constructed. One way to reduce the industry-dependence would be to normalise the scores for the complete asset universe, for example by using all companies for which ESG scores exist in the underlying database to construct the percentile scores. This would not solve the problem of ordinality, but it would at least ensure that all companies are ranked with respect to the same global reference distribution.

Optimal weighting of NLP-based features

A key difference between Study 1 (where the standard controversy score from Refinitiv was used) and Study 2 (where a new NLP-based controversy score had been constructed) is that the controversy score in Study 2 was optimally weighted in terms of the predictive power of the NLP features. This is likely to be responsible for the overall better predictability in Study 2. The better predictability observed in Study 2 can be attributed to the strategic application of NLP that picked out and weighted the features that tell us the most about ESG controversies. As a result, this approach not only improves the accuracy of predictions; It also reveals the specific aspects of ESG controversies that really matter, providing a better insight into the specific elements of ESG performance across sectors.

Generalizability of the results to other geographies and asset classes

The effectiveness of ESG scores' predictability from NLP based controversy scores demonstrated in these two studies of my thesis presents potential applicability of these findings across different geographies and asset classes. Reflecting on the results obtained from Norwegian stocks and private equity assets, it is necessary to consider the broader implications and the feasibility of extending this research framework to other contexts. These studies focus on the Norwegian market, which, while providing a reliable data set for analysis, represents a relatively specific geographical context. Expanding this research to a larger, more diverse geographic space, such as the New York Stock Exchange (NYSE) or the London Stock Exchange (LSE), provides an interesting perspective.

Regarding other asset classes such as corporate bonds, government bonds or real estate, this is difficult to say. In theory, extending the approach used in Study 2, this should be possible. But since these assets are relatively rarely covered in the news media, it is unlikely that sufficient training data exist to develop original controversy score models. And ESG scores for validation do not currently exist either. Future research is needed here, for example by following the similar data collection and scoring steps as the ones Kazinic and Valheim (2020) used for developing their method for private equity.

Conclusion and recommendations

The research presented in this thesis addressed three questions: (1) how well can classical rating-based ESG scores be predicted from NLP-based ESG controversy scores, (2) does the predictive relationship differ by industry, (3) how consistent are the effects when compared across different approaches for constructing ESG controversy scores and rating-based ESG scores. Two studies were conducted. Study 1 used data for OSX-listed companies, extracted from LSEG/Refinitiv Datastream. Study 2 focused on unlisted companies, re-analysing a data set developed by Kazinic and Valheim (2020).

The results of Study 1 suggest that the LSEG/Refinitiv ESG controversies score has no substantial relationship with the actual LSEG/Refinitiv ESG score; all explainable variation was due to global differences between industries. In financial analysis and portfolio management, such a low level of predictive validity would clearly be unacceptable. Hence, using the LSEG/Refinitiv ESG controversies score as a replacement for classical, rating-based ESG scores cannot be recommended.

The results of Study 2 indicate that the controversies scoring method developed by Kazinic and Valheim (2020) is substantially better than the LSEG/Refinitiv ESG controversies score and suffers much less from industry bias, but is still subject to considerable noise. If Kazinic and Valheim's controversies scores were to be used as a replacement of actual ESG scores in portfolio management, the predictive validity of the controversies scores would still not be fully sufficient. The approach can be recommended as a screening tool in financial analysis, but cannot replace deeper analysis of a company's actual ESG profile before an investment decision is made.

.

References

Agnese, P., Cerciello, M., Oriani, R., & Taddeo, S. (2024). ESG controversies and profitability in the European banking sector. *Finance Research Letters*, *61*, 105042. https://doi.org/10.1016/j.frl.2024.105042

Amel-Zadeh, A., & Serafeim, G. (2018). *Why and How Investors Use ESG Information: Evidence from a Global Survey* (SSRN Scholarly Paper 2925310). https://doi.org/10.2139/ssrn.2925310

Arabesque. (2021). ESG, News and the power of Natural Language Processing. *Arabesque*. https://www.arabesque.com/de/2021/10/18/esg-news-and-the-power-of-natural-language-processing/

Arvidsson, S., & Dumay, J. (2022). Corporate ESG reporting quantity, quality and performance: Where to now for environmental policy and practice? *Business Strategy and the Environment*, *31*(3), 1091–1110. https://doi.org/10.1002/bse.2937

Atz, U., Van Holt, T., Liu, Z. Z., & Bruno, C. C. (2023). Does sustainability generate better financial performance? Review, meta-analysis, and propositions. *Journal of Sustainable Finance & Investment*, *13*(1), 802–825. https://doi.org/10.1080/20430795.2022.2106934

Berg, F., Kölbel, J. F., & Rigobon, R. (2022). Aggregate Confusion: The Divergence of ESG Ratings*. *Review of Finance*, *26*(6), 1315–1344. https://doi.org/10.1093/rof/rfac033

Bloomberg. (2024). Global ESG assets predicted, Bloomberg LP. *Bloomberg L.P.* https://www.bloomberg.com/company/press/global-esg-assets-predicted-to-hit-40-trillion-by-2030-despite-challenging-environment-forecasts-bloomberg-intelligence/

Boffo, R., & Patalano, R. (2020). *ESG Investing: Practices, Progress and Challenges*.

Borglund, T., De Geer, H., Sweet, S., Frostenson, M., Grafström, M., Lerpold, L., Nordbrand, S., Sjöström, E., & Windell, K. (2021). *CSR and sustainable Business* (Second edition). Sanoma Utbildning.

Chowdhury, M. A. F., Abdullah, M., Azad, Md. A. K., Sulong, Z., & Islam, M. N. (2023). Environmental, social and governance (ESG) rating prediction using machine learning approaches. *Annals of Operations Research*. https://doi.org/10.1007/s10479-023-05633-7

De Spiegeleer, J., Höcht, S., Jakubowski, D., Reyners, S., & Schoutens, W. (2020). *ESG: A New Dimension in Portfolio Allocation* (SSRN Scholarly Paper 3712857). https://doi.org/10.2139/ssrn.3712857

Del Vitto, A., Marazzina, D., & Stocco, D. (2023). ESG ratings explainability through machine learning techniques. *Annals of Operations Research*. https://doi.org/10.1007/s10479-023-05514-z

Dimson, E., Marsh, P., & Staunton, M. (2020). Divergent ESG Ratings. *The Journal of Portfolio Management*, *47*(1), 75–87. https://doi.org/10.3905/jpm.2020.1.175

Dunn, J., Fitzgibbons, S., & Pomorski, L. (2017). *ASSESSING RISK THROUGH ENVIRONMENTAL, SOCIAL AND GOVERNANCE EXPOSURES*. https://www.semanticscholar.org/paper/ASSESSING-RISK-THROUGH-ENVIRONMENTAL,-SOCIAL-AND-Dunn-Fitzgibbons/de71d75565bf2bb338afef9f5711cef742134014

Eccles, R. G., & Stroehle, J. C. (2018). *Exploring Social Origins in the Construction of ESG Measures* (SSRN Scholarly Paper 3212685). https://doi.org/10.2139/ssrn.3212685

Edmans, A., & Kacperczyk, M. (2022). Sustainable Finance. *Review of Finance*, *26*(6), 1309–1313. https://doi.org/10.1093/rof/rfac069

ESG Investing. (2021). *Nordics are pioneering the future of ESG | ESG Investing: Sustainability News, Events & Awards*. https://www.esginvesting.co.uk/the-nordics-are-pioneering-the-future-of-esg/

Euronext. (2024). *Why Your Organisation's ESG Rating Is Important? - Euronext*. https://www.corporateservices.euronext.com/blog/esg-rating

European Commission. (n.d.). *Overview of sustainable finance*. Retrieved April 28, 2024, from https://finance.ec.europa.eu/sustainable-finance/overview-sustainable-finance_en

Ferson, W. E., Sarkissian, S., & Simin, T. T. (2002). *Spurious Regressions in Financial Economics?* (SSRN Scholarly Paper 328695). https://papers.ssrn.com/abstract=328695

Friede, G., Busch, T., & Bassen, A. (2015). *ESG and Financial Performance: Aggregated Evidence from More than 2000 Empirical Studies* (SSRN Scholarly Paper 2699610). https://papers.ssrn.com/abstract=2699610

Fu, T., & Li, J. (2023). An empirical analysis of the impact of ESG on financial performance: The moderating role of digital transformation. *Frontiers in Environmental Science*, *11*. https://doi.org/10.3389/fenvs.2023.1256052

Glasscock, R., Korenok, O., & Dorminey, J. (2021). Spurious Correlation Due to Scaling. *Journal of Accounting, Auditing & Finance*, *39*, 0148558X2110637. https://doi.org/10.1177/0148558X211063704

Global, S. (2021). *S&P Global makes over 9,000 ESG Scores publicly available to help increase transparency of corporate sustainability performance*. https://www.prnewswire.com/news-releases/sp-global-makes-over-9-000-esg-scores-publicly-available-to-help-increase-transparency-of-corporate-sustainability-performance-301228818.html

Haagensen, A., Ashraf, J. M., & Nilssen, T. A. (2023). *The Impact of ESG Factors on Financial Performance: Evidence from the Aluminum and Iron & Steel Industries* [Master's thesis, NTNU]. https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/3102834/no.ntnu:inspera:146714823:152917245.pdf?isAllowed=y&sequence=1

Jama, K., & Horstad, A. (2022). *ESG Scores and Firm Performance in the Nordics* [Master's Thesis, NHH,Norwegian school of Economics]. https://openaccess.nhh.no/nhh-xmlui/handle/11250/3015274

Kazinic, S., & Valheim, A. S. (2020). *Method for generating ESG ratings for unlisted companies based on NLP and predictive modelling* [Master's Thesis]. NMBU.

KDnuggets. (2020). *Spotting Controversy with NLP*. KDnuggets. https://www.kdnuggets.com/spotting-controversy-with-nlp

Khan, M., Serafeim, G., & Yoon, A. (2015). *Corporate Sustainability: First Evidence on Materiality*. https://dash.harvard.edu/handle/1/14369106

Kotsantonis, S., & Serafeim, G. (2019). Four Things No One Will Tell You About ESG Data. *Journal of Applied Corporate Finance*, *31*(2), 50–58. https://doi.org/10.1111/jacf.12346

KPMG. (2022, May 9). *Sustainable finance*. KPMG. https://kpmg.com/dp/en/home/industries/environmental-social-governance/sustainable-finance.html

Larcker, D. F., Pomorski, L., Tayan, B., & Watts, E. M. (2022). *ESG Ratings: A Compass without Direction* (SSRN Scholarly Paper 4179647). https://papers.ssrn.com/abstract=4179647

Lawton, G. (2023). *Timeline and History of ESG Investing, Rules and Practices*. Sustainability and ESG. https://www.techtarget.com/sustainability/feature/A-timeline-and-history-of-ESG-investing-rules-and-practices

Leaders Arena, E. (2023). *Why are ESG controversies so controversial & what can companies do about them?* Leadersarena. https://www.leadersarena.global/single-post/why-are-esg-controversies-so-controversial-what-can-companies-do-about-them

Lee, O., Joo, H., Choi, H., & Cheon, M. (2022). Proposing an Integrated Approach to Analyzing ESG Data via Machine Learning and Deep Learning Algorithms. *Sustainability*, *14*(14), Article 14. https://doi.org/10.3390/su14148745

Liu, M. (2022). Quantitative ESG disclosure and divergence of ESG ratings. *Frontiers in Psychology*, *13*, 936798. https://doi.org/10.3389/fpsyg.2022.936798

LSEG Data & Analytics. (2023). *Environmental, social and governance scores from LSEG*. https://www.lseg.com/en/data-analytics/sustainable-finance/esg-scores

McKinsey. (2020). *Why ESG scores are here to stay | McKinsey*. https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/why-esg-is-here-to-stay#/

MSCI. (2020). *ESG and the cost of capital*. https://www.msci.com/www/blog-posts/esg-and-the-cost-of-capital/01726513589

NBIM. (2024). *Responsible investment 2023*. Norges Bank Investment Management. https://www.nbim.no/en/publications/reports/2023/responsible-investment-2023/

NorNAB. (2023). *Landmark survey estimates Norwegian impact investing market at NOK 100 billion*. https://www.nornab.no/nyheter-blogg/the-state-of-impact-investing-in-norway

Norway's Pension Fund Vows To Purge Holdings That Drive Deforestation. (n.d.). *Ecosystem Marketplace*. Retrieved May 12, 2024, from https://www.ecosystemmarketplace.com/articles/norway-s-pension-fund-vows-to-br-purge-holdings-that-drive-deforestation/

Ozili, P. K. (2022). *Theories of Sustainable Finance* (SSRN Scholarly Paper 4055371). https://papers.ssrn.com/abstract=4055371

Pearl, J. (2014). Comment: Understanding Simpson's Paradox. *The American Statistician*, *68*(1), 8–13. https://doi.org/10.1080/00031305.2014.876829

Perazzoli, S., Joshi, A., Ajayan, S., & Neto, J. (2022). Evaluating Environmental, Social, and Governance (ESG) from a Systemic Perspective: An Analysis Supported by Natural Language Processing. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4244534

Refinitiv. (2022). *Environmental, Social and Governance scores from Refinitiv*.

Schimanski, T., Reding, A., Reding, N., Bingler, J., Kraus, M., & Leippold, M. (2024). Bridging the gap in ESG measurement: Using NLP to quantify environmental, social, and governance communication. *Finance Research Letters*, *61*, 104979. https://doi.org/10.1016/j.frl.2024.104979

Sprenger, J., & Weinberger, N. (2021). Simpson's Paradox. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2021/entries/paradox-simpson/

Tayan, B., Larcker, D., Watts, E., & Pomorski, L. (2022, August 24). ESG Ratings: A Compass without Direction. *The Harvard Law School Forum on Corporate Governance*. https://corpgov.law.harvard.edu/2022/08/24/esg-ratings-a-compass-without-direction/

Thomas, L. (2020, May 29). *Confounding Variables | Definition, Examples & Controls*. Scribbr. https://www.scribbr.com/methodology/confounding-variables/

UN. (n.d.). *Integrate the Principles for Responsible Investment | UN Global Compact*. Retrieved May 12, 2024, from https://unglobalcompact.org/take-action/action/responsible-investment

United Nations. (2020). The Sustainable Development Agenda. *United Nations Sustainable Development*. https://www.un.org/sustainabledevelopment/development-agenda/

Appendix

A1: OSX-listed companies used as assets in Study 1

ADEVINTA

AKASTOR

AKER

AKER BP

AKER SOLUTIONS

AUTOSTORE HOLDINGS

B2 IMPACT

DNB BANK

ELKEM

EQUINOR

GAMING INNOVATION GROUP

GRIEG SEAFOOD

MOWI

NORSK HYDRO

ODFJELL DRILLING

ORKLA

PROSAFE

REC SILICON

SALMAR

SPAREBANK 1 SMN ORDS

STOREBRAND

TELENOR

TGS

TOMRA SYSTEMS

WALLENIUS WILHELMSEN

YARA INTERNATIONAL

A2: Unlisted companies used as assets in Study 2

AGDER MASSETRANSPORT AS

AGDER STEIN & GRUNNARBEID AS

AKERSHUS MILJØ AS

AKILLES SALONG AS

ALBA CATERING AS

AQUABYTE AS A

BERGEN TANNKLINIKK AS

BERGNESET PUKK & GRUS AS

BLOMSTERENG LANDBRUK & MASKIN AS

BLÅFJELL AS

BRAGE FINANS AS

BRYGGERI 13, TROMSØ MIKROBRYGGERI AS

BRØDR. BØCKMANN AS

C-FEED AS

CHOSEN AS

CORVUS ENERGY AS

CSP AS

DEEP C SOLUTIONS AS

DISRUPTIVE TECHNOLOGIES RESEARCH AS

DOKTOR MOBIL AS

DRIFT SØR AS

DYRANUT FJELLSTOVA AS

DYREKASSEN AS

ENJOY ØKSNES AS

EXELA TECHNOLOGIES AS

FAGERHØY FJELLSTUE AS

FALCK ALARM BY VERISURE AS

FEM SMÅ STUER AS

FISKEBRYGGA LILLESAND AS

FREEDOME NORGE AS

GIAX PRODUKSJON AS

GRAPHCORE AS

GRAVITY SPORTS LARVIK AS

GRIFF AVIATION AS

GRÜNERLØKKA FLYTTEBYRÅ AS

GUDBRANDSDAL ENERGI AS F 43.210

H3 ARENA AS

HALFWAVE AS

HERMOD TEIGEN AS

HGN ELEKTRO AS

HJEM+ AS

HOLMATUN AS

HURUM NETT AS

HØYT OG LAVT TRONDHEIM AS

IAM INSURANCE AS

ICELAND MAT AS

INDIE AS

INZPIRE.ME AS

JÆREN PRIVAT OMSORG AS

KLM TERMO AS

KROGSVEEN NYBYGG AS

KVS TECHNOLOGIES AS

LANDSKAPSSERVICE AS

LASHES BY KASJA AS

LIPRO KOMPETANSE AS

LOFOTEN SEA PRODUCTS AS

MALO TRANSPORT AS

MEDITERRANEAN GRILL AS

MESTRINGSHUSENE AS

MHM ENTREPRENØR & SERVICE AS

MIDTKRAFT NETT AS

MILJØXPRESS AS

MOVING MAMAS AS

NETTBIL AS

NORD-SALTEN KRAFT AS

NORDLAND RENSEFISK AS

NORSERVICE AS

NORSK OMBRUK AS

NOVA SEA SERVICE AS

OKBAS TRANSPORT AS

PAYR AS L

PEOPLEHUB AS

R2S ELEKTRO AS

REV OCEAN AS

RIR TRANSPORT AS

RISKPOINT AS

ROGALAND ALARMTJENESTER AS

SAFEDRIVE AS

SEKKEN GRØNT AS

SENIOR AS

SJØGLØTT OMSORGSSENTER AS

SKOGLI-AS

SLETTÅS VANN- OG AVLØPSTEKNIKK AS

SNØHETTA SIKKERHET AS N

SPOND AS Q

SR-ENTREPRENØR AS

SUBSEAPARTNER AS

SUP & STUP AS

SVEISING-BYGG AS

SØR AURDAL ENERGI AS

TESTON AS R 90.035

TROLLFJORD NETT AS

TROMS BYGGFORVALTNING AS

TYTLANDSVIK AQUA AS

VAION AS

VALIDÉ AS

VALSØYA OPPLEVELSER AS

VERONA PIZZA OG CAFE AS

VESTAS NORWAY AS

WECLEAN NORGE AS

WELLESLEY PETROLEUM AS

X-VAC AS

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway