UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

ESG Rating Divergence and their Predictive Power for Carbon Emissions: *Evidence from the STOXX Europe 600 Companies*

Mai Onuki
Robert Marc Dovidat

Supervisor: Åsa Löfgren

Master's Thesis in Economics & Finance, 30 hec

Spring 2025

Graduate School, School of Business, Economics and Law, University of Gothenburg, Sweden

Abstract

This thesis investigates the divergence of environmental, social, and governance (ESG) ratings and the relationship between ESG ratings and corporate carbon emissions. Using two firm-level datasets covering 478 non-financial companies listed in the STOXX Europe 600 as of February 2025, we apply pairwise Spearman rank correlation analysis and two-way fixed-effects (TWFE) regression models. The empirical findings suggest significant divergence across ESG ratings. However, we do not observe a statistically significant negative relationship between ESG ratings and firms' future carbon emissions or intensity, which contradicts our hypothesis. This study contributes to the literature by offering a large-scale analysis of the European market. Furthermore, by directly comparing the predictive power of two ESG ratings with respect to corporate carbon emissions, we potentially provide important insights for sustainability-oriented investors regarding which ESG rating is a reliable indicator of genuine environmental performance.

Key words: ESG, raters, divergence, CO_2 emissions, carbon intensity, signaling

Contents

List of Figures

List of Tables

1 Introduction

Over the last decade, environmental, social, and governance (ESG) considerations have become a central topic in academic and professional discourse. Investors increasingly integrate ESG factors into their decision-making, driven by both financial and non-financial long-term values (OECD, 2020). This shift is urgent for the mitigation of accelerating climate change, where finance plays a crucial role in facilitating the transition to a carbon-neutral economy (International Labour Organization, 2022). ESG metrics are intended to provide a firm's business practices and performance with information on non-financial dimensions, driving a transition towards more responsible and sustainable finance (Sciarelli et al., 2021).

However, despite their growing importance, ESG ratings have one big issue: inconsistency across rating agencies (raters). Prior studies have shown that ESG scores from different raters often diverge significantly, potentially misleading stakeholders (Berg et al., 2022; Billio et al., 2021). This divergence not only complicates investment decisions but also undermines the credibility of ESG ratings as signals of corporate sustainability.

Beyond inconsistency, a further question arises: Do ESG scores serve their intended purpose? That is, do higher ESG scores actually predict better environmental outcomes? If ESG ratings are to act as a meaningful proxy for sustainability, one would expect that firms with higher ESG ratings emit less carbon dioxide (CO_2) in the following years. This expectation follows the idea that ESG scores not only reflect current environmental performance but also act as signals to investors. If perceived as credible, high ESG scores may attract sustainable investors and capital, which in turn allows and incentivizes firms to invest in initiatives to further reduce carbon emissions. Yet existing evidence is mixed: some studies find a negative relationship between ESG and firms' environmental performance, while others observe no significant effect, or even a reversed link.

This thesis aims to address these two issues. First, we investigate the divergence in ESG ratings of European corporations across major raters. We use a single-year dataset of ESG scores from five distinct raters: Refinitiv, Bloomberg, MSCI, Sustainalytics, and S&P Global. Conducting pairwise Spearman rank correlation analysis, we compare scores assigned to the same company by different raters. Second, we assess whether ESG ratings have any predictive power for corporate carbon performance in the future. We employ a panel dataset containing European firm-level ESG ratings from two raters: Refinitiv and Bloomberg, total CO_2 emissions, carbon intensity, and control variables over the period 2015 to 2023. Using two-way fixed-effects (TWFE) regression models, we test whether higher ESG scores in a given year will lead to lower carbon emissions or improved carbon intensity in the following year, controlling for firm and time fixed-effects (FE).

Our findings suggest that ESG scores vary significantly across raters. Pairwise correlations vary widely, from as low as 0.06 to 0.60 in absolute values. This confirms the prior research (e.g., Berg et al., 2022), which indicates ESG inconsistency among raters. However, we find no empirical support for the assumption that higher ESG ratings are associated with improved future carbon performance. Refinitiv ESG scores are positively and significantly associated with carbon emissions and intensity, contradicting our expectations. Bloomberg scores show the expected negative relationship with CO_2 emissions and a positive effect on carbon intensity, but neither relationship is statistically significant. Taken together, these findings suggest that ESG ratings may not capture firms' future carbon outcomes, limiting their effectiveness as a simplified decision-making tool for sustainable investors.

This thesis makes several important contributions. First, by focusing on firms in the European market, we fill the gap for geographic locations. While earlier studies report significant negative effects of ESG ratings on CO_2 emissions in China, we find no such impact for European firms. This suggests that the relationship between ESG ratings and firms' carbon performance may be influenced by regional regulatory structures or reporting standards. Second, our analysis directly compares the predictive power of different ESG ratings on firms' carbon outcomes. By assessing how different ESG raters predict actual environmental outcomes with a focus on firm-level carbon emissions, we provide investors a better understanding of which scores, if any, are reliable in a CO_2 context.

The rest of this thesis is organized as follows: Section 2 reviews relevant literature. Section 3 introduces our theoretical framework and poses the hypotheses to be tested. Section 4 describes the data and econometric methodology employed in our analysis. Section 5 presents the results, including robustness checks. Section 6 discusses our empirical findings. Finally, section 7 concludes and suggests future research possibilities.

2 Literature Review

As ESG factors increasingly influence corporate strategy and investment decisions, a growing body of literature has sought to define ESG, assess its measurement through ratings, and evaluate its implications for financial and environmental outcomes. This literature review is structured around the following topics: the conceptual foundations of ESG and ESG ratings (section 2.1), the divergence among ESG ratings (section 2.2), the relationship between ESG ratings and financial performance (section 2.3), and finally, the relationship between ESG ratings and environmental performance (section 2.4). These discussions provide the groundwork for identifying unresolved questions that this thesis aims to address.

2.1 ESG

Shareholders, investors, employees, and regulatory bodies are demanding greater transparency and actions on sustainability-related issues, driven by the growing urgency of environmental and social challenges (Chytis et al., 2024). Despite its widespread use, the concept of ESG still lacks a universally accepted definition. However, Matos (2020) defines ESG and its three pillars as follows. The environmental (E) pillar captures a company's impact on the natural ecosystem, including pollution levels, waste management, and resource consumption. The social (S) pillar evaluates how a company interacts with its stakeholders, particularly its employees, customers, and society at large. It encompasses aspects such as labor conditions, workplace safety, and community engagement. The governance (G) pillar refers to the structures ensuring that managements act in the long-term interests of shareholders, covering issues such as board composition, shareholder rights, and ethical business practices (Matos, 2020).

As ESG has gained popularity, it has become the dominant framework for assessing corporate sustainability. According to Bloomberg (2024), global ESG assets exceeded $30 trillion in 2022. Despite recent setbacks, including political resistance in the United States under current President Donald Trump, such as the withdrawal from the Paris Climate Agreement (Federal Register, 2025) or an executive order limiting state ESG policies (The White House, 2025), and the omnibus proposal of the European Union (European Commission, 2025) which aims to reduce the burden of multiple climate-related regulations, the overall trend remains positive. Bloomberg (2024) predicts that global ESG assets will surpass $40 trillion by 2030.

To quantify ESG performance, various raters have developed ESG ratings, combining sustainability-related data into a single score. Prominent rating providers include MSCI, Sustainalytics, S&P Global, Refinitiv, and Moody's ESG (Berg et al., 2022). These ratings rely on multiple data sources, such as corporate disclosures, stakeholder reports, and industry-specific characteristics. Companies typically self-report sustainability data through reports on carbon emissions, governance structures, and employee welfare. In addition, ESG raters incorporate external stakeholder information, including government databases, NGO reports, and social media sentiment. The business model, geographic location, and industry sector also play roles in determining its ESG score (OECD, 2025).

ESG ratings serve as crucial information intermediaries for investors, financial institutions, and regulators. They are also widely used in academic research and corporate decision-making (Cregan et al., 2024a). High ESG scores are often associated with lower exposure to financial risks, which can reduce the volatility of stock prices and the probability of default (Friede et al., 2015; Whelan et al., n.d.). Additionally, investors use ESG scores to filter companies based on sustainability performance (Bolton & Kacperczyk, 2021).

2.2 ESG Rating Divergence

Despite their widespread use, ESG ratings remain controversial and are highly debated in academia. A key concern is the lack of consistency among the raters. Unlike credit ratings, which typically exhibit near-perfect correlations among raters (0.99 according to Berg et al., 2022), ESG ratings vary significantly. Berg et al. (2022) investigate the U.S. market and find correlations between the major ESG raters, ranging from 0.38 to 0.71, with a mean of only 0.6. The authors argue that this divergence is primarily due to differences in measurement approaches, as well as differences in the scope and weighting of ESG components.

Further supporting this view, Billio et al. (2021) analyze the ESG scores of nine leading raters and document substantial heterogeneity. They argue that there is no universally accepted definition of ESG, and that the specific standards and indicators used to evaluate environmental, social, and governance dimensions differ widely. Moreover, Capizzi et al. (2021) emphasize that differences in the way the three pillars are weighted (for example, placing greater emphasis on governance in one model and on the environment in another) further contribute to divergence of the scores. A recent OECD (2025) study examines more than 2,000 ESG metrics used by eight major raters and identifies gaps in the coverage of ESG metrics and inconsistencies in how metrics are defined and applied. The report also highlights that for each pillar (E, S, and G), the data is often drawn from distinct data sources, including company disclosures, stakeholder feedback, and private indicators, adding another layer of disagreement and making it less transparent.

This lack of consistency has critical implications for both investors and researchers. According to Billio et al. (2021), fragmented ESG ratings lead to inconsistent investment universes, making it difficult to compare or benchmark sustainability performance across firms. The absence of a shared framework increases uncertainty for both institutional and private investors trying to evaluate what truly represents a sustainable investment. Considering the growing reliance on ESG ratings to guide sustainable investment decisions, it is crucial to assess how well these ratings reflect actual environmental performance. In particular, it is important to examine whether different rating agencies vary in how strongly their scores are associated with concrete environmental outcomes such as firm-level carbon emissions. We aim to contribute to filling this gap by examining different ESG ratings and their impact on carbon emissions. That is, we empirically compare multiple ESG ratings and we test their ability to predict firm-level CO_2 emissions, offering insight into the informativeness and reliability of ESG ratings as environmental performance indicators. Despite the widespread use of ESG ratings and ongoing concerns about rating divergence, few studies systematically compare rating providers using the same firms and outcome metrics.

In addition to the divergence among raters, ESG ratings also exhibit systematic patterns. Larger firms tend to receive higher ESG scores, primarily due to their ability to produce more detailed and frequent disclosures (Drempetic et al., 2020). The OECD (2025) has also analyzed that some of the drivers of ESG scores are correlated with a company's size and market capitalization. Companies with higher ESG scores are, on average, larger in terms of market capitalization than those with lower scores. This phenomenon, referred to as the size effect by Cregan et al. (2024a), implies that the ratings are partially driven by the volume of disclosure rather than the actual performance. Similarly, European companies generally score higher. Doyle (2018) and Gyönyörová et al. (2023) argue that stricter regulatory disclosure requirements in Europe might be the reason for this pattern. Lee (2021) supports this interpretation, arguing that ESG raters rely heavily on self-reported data, which influences scores in favor of firms that are either more transparent or subject to more stringent reporting obligations. These inconsistencies raise questions about whether ESG ratings reflect actual firm behavior or primarily act as a signaling mechanism influenced by disclosure incentives.

2.3 ESG Ratings and Financial Performance

The relationship between ESG ratings and financial performance is highly debated in the academic literature. Wang et al. (2025) and Chen and Xie (2022) find that ESG disclosures lower financing costs and attract (institutional) investors. In addition, Velte (2017) finds a positive effect of ESG on the return on assets (ROA).

However, other authors (e.g., Atan et al., 2018) cannot find significant impacts. Atan et al. (2018) even find that ESG might increase the weighted average cost of capital (WACC). These conflicting results might arise from differences in methodological approaches, locations, and raters (as discussed before). For example, Atan et al. (2018) analyze Malaysian firms while Wang et al. (2025) and Chen and Xie (2022) analyze Chinese firms, suggesting that the effect on financial performance may not be generalized to all markets or regions. Furthermore, ESG rating uncertainties can reduce investor demand for green stocks, and therefore, the impact of ESG ratings on asset pricing remains ambiguous (Avramov et al., 2022).

Several meta-analyses attempt to resolve these contradictions. Analyses by Friede et al. (2015) and Whelan et al. (n.d.) find that most studies report a non-negative relationship between ESG and financial performance. This suggests that although the overall trend leans positive, industry-, region-, and methodology-specific factors play a significant role in determining the financial impact of ESG.

2.4 ESG Ratings and Environmental Performance

The link between ESG ratings and environmental performance is even less conclusive than the one with financial performance. A majority of studies examining the relationship between ESG ratings and carbon emissions are conducted in the Chinese market (Cong et al., 2022; Li & Xu, 2024; Xie et al., 2024). These studies generally support a negative correlation between ESG scores and emissions, suggesting that ESG investments can reduce environmental harm. However, the use of Chinese-specific ESG indices (e.g., China Securities ESG Index, SynTao Green Finance ESG) raises concerns about the generalizability of these findings beyond the Chinese context.

In contrast, studies conducted in non-Chinese settings often fail to confirm such relationships. For example, Cregan et al. (2024a) and Cregan et al. (2024b) examine ESG ratings in high-emission sectors, the airline and steel industries, respectively, and find no significant link between ESG scores and carbon emissions. These findings suggest that even firms with high ESG ratings may still be major polluters. Similarly, Treepongkaruna et al. (2024) argue that such patterns support the greenwashing hypothesis, in which companies achieve high ESG ratings without meaningful reductions in actual emissions.

These findings highlight the need to investigate the role of ESG ratings in the European market. While prior research has acknowledged that ESG ratings may vary across regions, few studies have conducted a detailed analysis focused specifically on Europe. This study addresses that gap by analyzing companies of the STOXX 600 index [1] to assess how different ESG rating providers capture firm-level carbon emissions. Using a consistent set of European firms enables a controlled comparison across ESG ratings while minimizing regional confounding effects.

[1] The STOXX Europe 600 is a broad measure of the European equity market. With 600 fixed components, it offers broad and diversified coverage across 17 countries and 11 industries within Europe's developed economies, representing nearly 90% of the investable market (STOXX, n.d.-a).

3 Theory and Hypotheses

As ESG factors play a fundamental role in investment decisions, it is essential to explore the theoretical frameworks that explain how ESG ratings influence investor behavior and firms' performance. This section introduces two key concepts: signaling theory (section 3.1) and investor behavior (section 3.2), to guide the hypotheses development regarding the ESG divergence and the relationship between ESG scores and corporate sustainability outcomes, specifically carbon emissions.

3.1 Signaling Theory

Signaling theory, initially introduced by Spence (1973), is an important framework for understanding corporate behavior in the context of ESG ratings. Originally developed to explain how individuals signal their productivity in the job market, signaling theory addresses situations with asymmetric information, where one party has more information about its attributes than the other. In Spence's model, job applicants know their own productivity while employers do not. Hence, to assess the potential quality of employees, employers use education as a signal of individuals' capabilities, mitigating information asymmetries. Individuals, in turn, strategically invest in education to shape how they are perceived in the labor market. Spence (1973) explains that the decision to obtain a signal depends on a cost-benefit trade-off between signaling costs and expected returns: individuals weigh the cost of acquiring education against the offered wages.

This theoretical framework further extends beyond job markets to broader situations where asymmetric information exists, including corporate sustainability and ESG ratings. In financial markets, firms use ESG scores to signal a credible commitment to environmental and social responsibility, with the aim of attracting sustainability-oriented investors (Qian & Liu, 2024). ESG ratings serve as standardized assessments of non-financial performance, helping to reduce information asymmetry between firms and investors (Li & Xu, 2024). As a result, firms with higher ESG scores are generally perceived as more environmentally responsible, less exposed to long-term risks, and potentially more financially attractive (Amel-Zadeh & Serafeim, 2018; Bauer et al., 2021; Lashgari, 2024). This investor perception creates incentives for firms to undertake real environmental improvements, such as investing in renewable energy, energy efficiency, or green innovation, as a means to enhance or maintain their ESG scores. These actions can be seen as costly signals, reinforcing the credibility of the firm's commitment and increasing the likelihood of favorable capital flows.

However, unlike education, where individuals choose their own signals, companies cannot directly assign themselves an ESG score. While signals are alterable and therefore potentially subject to manipulation by job applicants (Spence, 1973), ESG ratings are determined by third-party raters, meaning that firms cannot fully control the signals they send.

Still, this lack of direct control over ESG ratings does not eliminate the potential for greenwashing. Greenwashing refers to a deliberate use of misleading information by companies to deceive stakeholders about their environmental performance (De Freitas Netto et al., 2020). A key issue here is that ESG ratings heavily rely on input-based and qualitative metrics. According to OECD (2025), less than one-third of the metrics primarily rely on output-based metrics, while 68 percent are input-based. ESG metrics focus more on self-reported actions rather than the outcomes of these policies and activities. Additionally, ESG performance is mostly (72 percent) assessed using qualitative rather than quantitative metrics (OECD, 2025). Qualitative metrics may not always provide a reliable proxy of a firm's actual ESG performance, but rather evaluate the existence of the activities or policies. This reliance on input-based and qualitative metrics could incentivize companies to have 'tick-boxing' approaches over actual effectiveness (OECD, 2025). Moreover, by selecting what information to disclose, firms may influence their scores without improving their performance.

3.2 Investor Behavior

While signaling theory explains how firms utilize ESG scores to signal their sustainability efforts, the impacts of these signals depend on how investors perceive and respond to them. Therefore, understanding investors' behavior is crucial to exploring the mechanism of how ESG scores affect corporate carbon performance. Behavioral finance suggests that investors are driven by both financial and social factors such as larger financial returns, risk mitigation, and social preferences when making investment decisions (Amel-Zadeh & Serafeim, 2018; Bauer et al., 2021; Lashgari, 2024).

Some investors engage in sustainable investments because they expect them to outperform traditional investments (Bauer et al., 2021). Amel-Zadeh and Serafeim (2018) suggest that investors use ESG information since it is financially material [1] to investment performance, and interestingly, they show that non-ESG specialist individuals consider ESG information for materiality reasons more compared to specialists. Others perceive sustainable investments as having lower downside risk, since they mitigate the financial risks associated with environmental damages (Lashgari, 2024). Bauer et al. (2021) explain that investors tend to dislike uncertainty and seek investments that reduce potential risks. A survey by the OECD (2020) also shows that institutional investors primarily focus on the financial returns and risk management advantages of ESG investing.

Beyond financial considerations, some investors are motivated by social preferences, meaning that they are willing to sacrifice financial returns in order to increase social welfare (Bauer et al., 2021). Bauer et al. (2021) find that many investors are willing to engage in sustainable investing even when they face negative return expectations or are uncertain about the return, due to strong social preferences in investment decisions. A study by Pástor et al. (2021) also suggests that agents are willing to pay more for greener firms and expect to gain lower returns, yet do not experience disutility, as they derive non-financial satisfaction from holding green investments.

As described above, investors' interest in sustainable investments is driven by a combination of financial motivations and non-financial preferences, such as ethical values or environmental concerns, or either. However, the decision-making process in sustainable investment is complex and costly, requiring investors to evaluate corporate sustainability performance across multiple ESG factors, firms, and industries. Kolling and Busch (2025) argue that, under the existence of asymmetric information among market participants, investors are unable to directly observe companies' sustainability performance and have to put much effort into gathering information. This process can be challenging given that investors may lack information or expertise to analyze companies' sustainable performance.

Simon (1955) introduced the concept of 'bounded rationality', which suggests that individuals are limited in their ability to process all available information and compute op-

[1] According to the U.S. Supreme Court (1976), information is considered 'material' if a reasonable investor would likely see it as important when making investment decisions.

timal decisions due to cognitive constraints. In practice, investors may lack the expertise, resources, or time, potentially preventing them from processing all available information about a firm's sustainability performance. As a result, many investors rely on heuristics, i.e., processes that simplify and reduce mental effort (Shah & Oppenheimer, 2008).

In line with this, Löfgren and Nordblom (2020) argue that individuals tend to rely on heuristics because utility maximization is cognitively demanding. In the context of ESG investing, this suggests that many investors who are in favor of sustainable investments simplify their decision-making process by using ESG scores as a proxy for overall sustainability. Instead of conducting a detailed analysis of underlying ESG factors, they may rely on the score alone, assuming it adequately reflects a company's environmental performance. Although this heuristic approach does not represent fully rational or attentive evaluation, it allows investors to make quicker and more manageable decisions under conditions of limited information and cognitive resources. This can result in mistakes (Löfgren & Nordblom, 2020) if ESG ratings do not accurately reflect a company's performance and therefore compromise investors' intentions of investing in sustainable companies. The use of ESG ratings as decision-making shortcuts highlights the cognitive constraint faced by investors.

Additionally, the divergence of ESG ratings adds further complexity. As discussed in section 2, ESG ratings often vary among raters due to different scopes, measurements, and weights (e.g., Berg et al., 2022). This variation can lead to inconsistencies in how investors interpret a company's sustainability performance. As Darnall and Aragón-Correa (2014) argue that trust is essential for signals to be effective, if ESG scores, as signals of sustainability, lack credibility for investors, they may fail to address information asymmetry. The lack of consistency further complicates the decision-making process and presents a challenge for investors who rely on ESG scores to guide their decisions (Berg et al., 2022).

3.3 Hypotheses

Given the theoretical foundations outlined above, this study addresses two key questions. First, do ESG ratings vary significantly across raters? Second, do higher ESG ratings predict better corporate environmental performance in the future, particularly in terms of carbon emissions?

As prior studies suggest, different ESG raters often assign inconsistent scores to the same firms (e.g., Berg et al., 2022). This lack of alignment can lead to confusion and uncertainty, particularly for investors who seek to make informed decisions based on sustainability criteria. From the perspective of signaling theory, such inefficiency may misallocate capital flow, weakening the effectiveness of ESG ratings as investment signals. Based on this, we propose the following hypothesis:

Hypothesis 1. ***ESG ratings diverge significantly across raters.***

This divergence raises a further question: Despite their inconsistencies, do ESG ratings still provide meaningful information about corporate environmental performance? Which ESG score predicts the performance the most accurately? CO_2 emissions are among the most widely used indicators of environmental impact, given the growing importance of climate concerns (Li & Xu, 2024). In line with the concepts of signaling theory and investor behavior, ESG ratings influence carbon emissions through investors' investment decisions. If ESG ratings are perceived as reliable and effective signals of sustainability, companies with higher ESG scores may attract greater investors' interest and capital inflows. These financial advantages can, in turn, enable firms to invest further in sustainable initiatives, which result in additional reductions in future carbon emissions. ESG ratings, therefore, may not only reflect current environmental performance but also incentivize firms to reduce total emissions in order to attract sustainability-focused investors. Accordingly, we would expect a negative association between ESG ratings and carbon emissions and carbon intensity: higher-rated firms should emit less (per unit of output) in the future, both because ESG scores reflect environmental performance and because they may facilitate further environmental improvement. Previous literature confirms this relationship, e.g., Tang et al. (2024) find that the behavior of sustainable investors can promote the green innovation of firms. This mechanism is illustrated in Figure 1.

However, it is important to note that this negative relationship may not always hold and can be violated in the presence of reverse causality or greenwashing. One assumption about the relationship between ESG ratings and carbon emissions is that firms with lower CO_2 emissions should receive higher ESG ratings, holding other components constant, since CO_2 emissions serve as an input into ESG scoring methodologies. In the case of reversed causality, we might observe an unexpected positive relationship between ESG

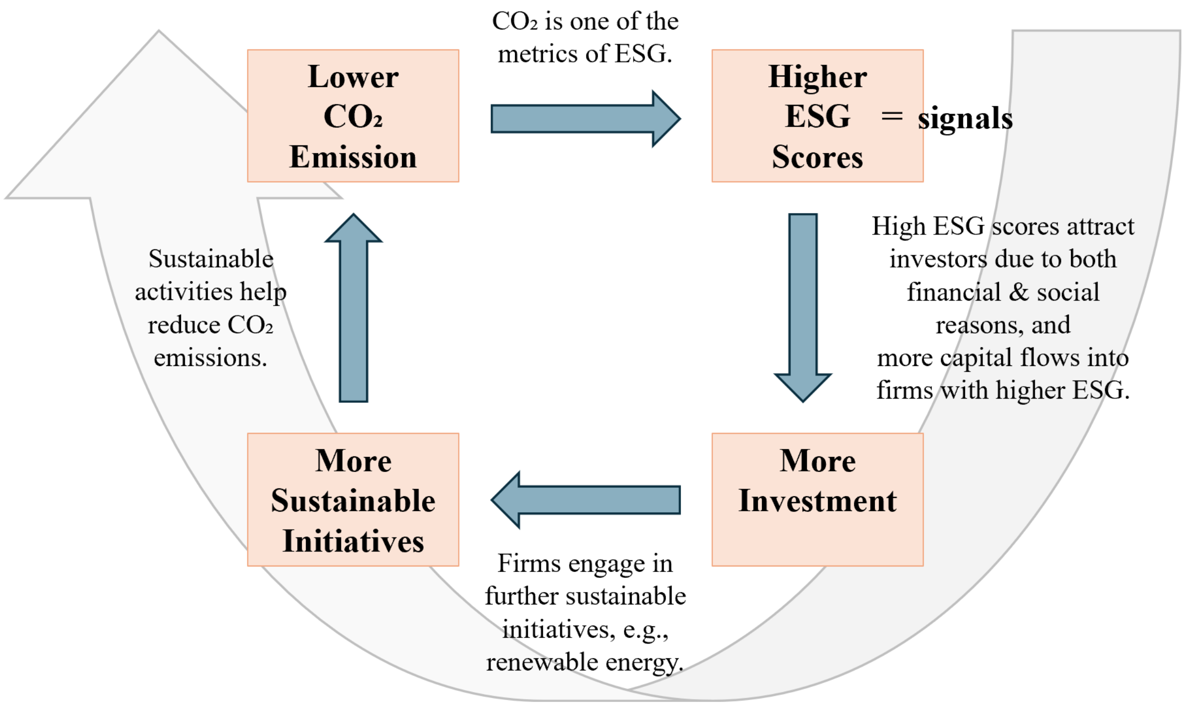

Figure 1: Relationship between ESG Ratings and CO_2 Emissions

ratings and carbon emissions as illustrated in Table 10 in the Appendix. Or, as discussed above, firms may strategically highlight their superficial green initiatives that actually do not have real sustainable impacts, which could result in receiving high ESG scores while still having large emissions. Despite these concerns, based on earlier literature and theories, we propose the following hypothesis:

Hypothesis 2. ***Firms with higher ESG ratings have lower carbon emissions and carbon intensity in the future.***

4 Data and Methodology

This section outlines the data and methodology used to investigate our two hypotheses. It is structured as follows: Section 4.1 describes data construction and key variable selection, with summary statistics of our datasets. Section 4.2 explains the econometric methodology and the model specifications.

4.1 Data and Descriptive Statistics

This study utilizes two datasets corresponding to each of the two hypotheses. The first is a single-year dataset consisting of ESG scores from five raters for non-financial firms listed in the European STOXX 600 as of February 2025. The second is a novel panel dataset of the same firms, covering the period from 2015 to 2023. All variables are sourced from two primary databases: LSEG Workspace and Bloomberg Terminal.

4.1.1 Data Collection

We followed a multi-step process to construct the datasets: We began by downloading firm-level data for over 1500 European firms from LSEG Workspace, including Refinitiv ESG score, CO_2 emissions, and control variables for years from 2014 to 2024. Next, we collected Bloomberg ESG scores for the same period, while ESG scores from Sustainalytics, S&P Global, and MSCI were downloaded for a single year via Bloomberg Terminal. Additionally, a list of STOXX EUR 600 firms as of February 2025 was obtained from the STOXX selection list (STOXX, n.d.-b). Then, we excluded financial firms from the STOXX list based on the Global Industry Classification Standard (GICS) (MSCI, 2024), resulting in 478 non-financial firms. Last, these 478 firms were then matched and merged with the datasets, both using Python and manually.

This process produced two datasets: (i) Dataset 1: a single-year dataset for five ESG ratings used to analyze ESG score divergence, and (ii) Dataset 2: a panel dataset used to examine the relationship between ESG scores and carbon emissions.

Data cleaning and processing were performed using Python (for data merging, computing, and plotting), Stata (for further cleaning and visualization of tables and plots), and manual procedures (for matching) where needed. Some variables, such as carbon intensity (Li & Xu, 2024), Tobin's Q (Qian & Liu, 2024), tangible asset ratio, firm size (Treepongkaruna et al., 2024), and revenue growth rate, were derived using formulas from previous literature.

4.1.2 Dataset 1: ESG Divergence

To test Hypothesis 1, we investigate ESG ratings from five distinct raters: Refinitiv, Bloomberg, Sustainalytics, S&P Global, and MSCI for the 478 non-financial firms.

Refinitiv and Bloomberg scores were extracted for the most recent year available. For MSCI, Sustainalytics, and S&P Global (accessed via Bloomberg Terminal), the exact year of the ratings could not be determined, therefore, we assume them to be the most recent available. Due to variations in data availability, our dataset is not fully balanced, as some ratings are missing for certain firms.

Table 1 shows the definitions and descriptive statistics for the five ESG ratings (Dataset 1). Refinitiv, Sustainalytics, and S&P Global provide scores on a 0-100 scale, while Bloomberg uses a 0-10 scale (Bloomberg, 2023; LSEG Data & Analytics, 2024; Morningstar Sustainalytics, n.d.; S&P Global, 2025). MSCI reports categorical ratings with seven discrete levels: AAA, AA, A, BBB, BB, B, and CCC, ordered from best to worst (MSCI, n.d.). It is important to note that, unlike other ratings, Sustainalytics follows a reverse scale, where higher scores reflect greater risk due to ESG factors (Morningstar Sustainalytics, n.d.). Descriptive statistics indicate considerable variation across ratings. The mean values span from 4.67 (Bloomberg, 0-10 scale) to 83.05 (S&P Global, 0-100 scale), while Refinitiv shows an intermediate average of 69.94 (0-100 scale). Bloomberg exhibits a lower maximum score (7.67) compared to S&P Global and MSCI, for which the highest values reach the maximum of the range. Sustainalytics has a relatively narrower range between 4.51 and 40.09 with an inverted scale, likely due to its risk-based structure. These differences suggest that it may be more challenging for firms to attain high scores from Bloomberg, compared to other ratings. Taken together, these variations highlight the differences among raters.

Table 1: Summary of ESG Ratings for Dataset 1

ESG Ratings	**Definition**	**Range**	**Obs.**	**Mean**	**Std. Dev.**	**Min**	**Max**
Refinitiv	Measures a company's relative ESG performance, commitment, and effectiveness, based on company-reported data. Higher scores mean better performance.	0-100	470	69.94	13.24	16.36	94.67
Bloomberg	Measures a company's management of financially material ESG issues. Higher scores indicate a better management of material issues.	0-10	460	4.67	1.16	0.72	7.67
Sustainalytics	Assesses the extent of a company's economic value at risk due to ESG factors by evaluating its exposure to and management of material ESG issues. Scores above 40 indicate severe ESG risk.	100-0	458	19.09	6.84	4.51	40.09
S&P Global	Measures a company's performance on and management of material ESG risks, opportunities, and impacts. An ESG score of 100 represents the maximum.	0-100	455	83.05	15.31	11	100
MSCI	Measures a company's resilience to long-term, industry-specific ESG risks using a rules-based methodology. A score of AAA indicates the best, while CCC represents the worst.	CCC-AAA	321	4.68	1.19	1	6

Notes: For each range, values toward the right end represent stronger ESG performance. MSCI scores are converted into numerical measures, with 0 indicating CCC (the lowest) and 6 being AAA (the highest).
Sources: LSEG Data & Analytics (2024), Bloomberg (2023), Morningstar Sustainalytics (n.d.), S&P Global (2025), & MSCI (n.d.)

4.1.3 Dataset 2: ESG and Carbon Emissions

To test Hypothesis 2, we use a panel dataset consisting of the same 478 non-financial European firms from 2015 to 2023, resulting in 4302 observations in total. Note that Dataset 2 is also unbalanced, as some years, ratings, or control variables are missing for certain firms, due to variations in data availability. This study focuses on the European stock market in order to bridge the geographical gap in existing literature and to minimize variations in firm characteristics and regulations. Financial firms, including banks, financial services, consumer finance, capital markets, and insurance firms, as classified by the GICS by MSCI (2024), are excluded to align with previous studies (e.g., Li & Xu, 2024). The chosen time frame is motivated by data availability, as there are limited Bloomberg ESG scores available before 2015 and from 2023 onwards (see Figure 4 in the Appendix).

The main variables of interest are the dependent variables: carbon emissions (CO_2) and carbon intensity, and the independent variables: the Refinitiv ESG score and the Bloomberg ESG score.

Our dependent variables, carbon emissions and carbon intensity, are used as a proxy for environmental performance. As CO_2 pollution is one of the main driving factors for climate change (Crowley, 2000), it is a crucial factor in determining the sustainability performance of a company. CO_2 in our study is the total carbon dioxide and carbon dioxide equivalents emitted by the company each year in tonnes. Carbon intensity is created as the ratio of CO_2 emissions to revenue in million euros, referring to Li and Xu (2024). Using both the carbon intensity and the raw emissions allows us to assess not only the absolute impact of a company but also its emission efficiency, enabling more meaningful comparisons across firms of different sizes.

Two ESG ratings, from Refinitiv and Bloomberg, serve as the independent variables, essentially representing the overall company scores based on the environmental, social, and governance pillars. While it would be valuable to include additional ratings, limited data accessibility constrained our selection. According to Li and Polychronopoulos (2020), both Refinitiv and Bloomberg collect data from public sources without offering any value-adding scoring. This means that they are similar in the data collecting process, but also differ in scoring methodology, making the ESG scores more comparable in the analysis. The variations in two ESG ratings are shown in Figure 5 in the Appendix.

In addition, we use a variety of control variables to limit the impact of other influencing variables on the regression outcomes. Building on existing research (Li & Xu, 2024; Tang et al., 2024; Velte, 2017; Xie et al., 2024; Yang et al., 2024), we include the following controls in our study: leverage (debt-asset ratio), return on average total assets (ROA) as an indicator of profitability, Tobin's Q for firm values, the natural logarithm of total assets to account for firm size, revenue growth, and tangible asset ratio. We also control for governance measures such as board size, independent board member ratio, and CEO-Chairman duality (whether the CEO simultaneously chairs the board or the chairman of the board has been the CEO of the company). These financial and governance indicators are often considered to have an influence on firms' emission levels (Li & Xu, 2024). Chairperson independence, which is used in several previous studies, presents similar information to the CEO-Chairman duality. We have chosen the latter due to its greater data availability, as shown in Figure 4 in the Appendix. Additionally, to align with earlier studies mentioned, we initially intended to include research and development (R&D) expense, as Xu et al. (2021) argue that increased R&D investments in green innovations can improve carbon emission performance. However, we decided to drop it due to the high number of missing observations (see Figure 4). Environmental regulation was also considered important because emissions may be far lower under strict regulations, however, we omitted it due to data availability constraints.

Table 2 presents the descriptions and summary statistics for the panel dataset (Dataset 2). CO_2 emissions exhibit great variation, with a mean of 3.97 million metric tons, ranging from as low as 0.31 to a maximum of 195 million tonnes. This reflects variation in emission levels across firms. Carbon intensity also shows a substantial variation, varying from 0.003 tonnes of CO_2 per million euros of revenue to 9,882.8. The average ESG scores of Refinitiv and Bloomberg are 67.18 and 3.937, corresponding to 67.18% and 39.37% of their respective best scores on average across years, highlighting significant differences between their scores. Additionally, Refinitiv scores span a broad range from 3.91 to 95.58, while Bloomberg has a narrower range between 0.47 and 7.67. This again suggests that Refinitiv applies a wider distribution of scores, whereas it may be more challenging for firms to attain high scores from Bloomberg.

Table 2: Summary of Variables for Dataset 2

Variables	**Definition**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Independent Variables						
CO_2	Total carbon dioxide (CO_2) and CO_2 equivalents emission in tonnes.	3632	3966970	1.50e+07	0.31	1.95e+08
Carbon Intensity	Ratio of corporate carbon emissions to total revenue (in million euros): $\frac{CO_2 \text{ Emissions}}{\text{Total Revenue (in €)}} \times 1000000$.	3620	215.565	640.245	0.003	9882.796
Dependent Variables						
Refinitiv ESG	Overall company score based on the self-reported information in the E, S, and corporate G pillars.	3872	67.184	16.324	3.910	95.583
Bloomberg ESG	Score that evaluates the company's aggregated E, S, and G performance, based on Bloomberg's view of ESG financial materiality.	3406	3.937	1.314	0.47	7.67
Control Variables						
Leverage	Ratio of total debt to total assets: $\frac{\text{Total Debt}}{\text{Total Assets}}$.	4126	0.264	0.151	0	1.304
ROA	Return on average total assets, showing how efficient a company is at using its assets to generate earnings.	4136	0.066	0.119	-0.535	2.368
Tobin's Q	Ratio of market capitalization to total assets: $\frac{\text{Market Capitalization}}{\text{Total Assets}}$.	4058	1.857	3.793	0.002	80.617
Tangible Ratio	Ratio of total fixed assets: $\frac{\text{Fixed Assets}}{\text{Total Assets}}$.	4157	0.458	0.231	0.002	0.994
Firm Size	Natural logarithm of total assets (in euros).	4177	22.822	1.518	16.304	27.121
Growth Rate	Growth rate of total consolidated revenue of a company.	4140	0.103	0.381	-4.079	11.953
Board Size	Total number of board members at the end of the fiscal year.	3869	10.851	3.672	3	30
Independent Board	Percentage of independent board members.	3870	62.690	23.334	0	100
CEO Duality	Dummy, =1 if the CEO is also the chairman, 0 otherwise.	3875	0.239	0.427	0	1

Sources: LSEG Workspace & Bloomberg Terminal

Table 3 reports the pairwise correlation coefficients among all variables included in our panel dataset (Dataset 2) used to test Hypothesis 2. CO_2 emissions and carbon intensity are highly correlated (0.67), which is expected as carbon intensity is derived as the ratio of emissions to total revenue. Two ESG ratings, Refinitiv and Bloomberg, have a moderate correlation of 0.45, indicating some consistency in scoring methodologies, while also suggesting that they capture ESG performance differently. Moreover, we do not observe strong correlations between ESG ratings and financial variables such as leverage, ROA, and Tobin's Q, which implies that ESG scores capture dimensions of corporate behavior that are not solely explained by financial performance, as expected.

Table 3: Correlation Matrix for All Variables in Dataset 2

	co2	ci	ref	blo	lev	roa	tobq	tang	firsiz	grow	bosiz	indep	dual
co2	1.000												
ci	0.666	1.000											
ref	0.155	0.0205	1.000										
blo	0.153	0.080	0.448	1.000									
lev	-0.038	0.011	0.083	0.112	1.000								
roa	-0.078	-0.082	-0.138	-0.059	-0.181	1.000							
tobq	-0.089	-0.090	-0.173	-0.094	-0.142	0.784	1.000						
tang	0.136	0.169	0.081	0.207	0.356	-0.153	-0.171	1.000					
firsiz	0.343	0.154	0.529	0.306	0.189	-0.334	-0.387	0.235	1.000				
grow	0.005	-0.031	-0.069	0.045	-0.016	0.056	0.028	-0.001	-0.020	1.000			
bosiz	0.132	0.108	0.294	0.064	0.097	-0.153	-0.160	0.057	0.526	-0.039	1.000		
indep	0.052	0.049	0.268	0.308	0.070	-0.004	-0.015	0.032	0.074	0.021	-0.225	1.000	
dual	0.015	-0.053	0.032	-0.085	-0.015	-0.023	0.0002	-0.045	0.088	-0.004	0.142	-0.176	1.000

Notes: Variables are presented as follows: CO_2 emissions: *co2*, carbon intensity: *ci*, Refinitiv ESG: *ref*, Bloomberg ESG: *blo*, leverage: *lev*, ROA: *roa*, Tobin's Q: *tobq*, tangible ratio: *tang*, firm size: *firsiz*, growth rate: *grow*, board size: *bosiz*, independent board: *indep*, CEO duality: *dual*

Figure 2 illustrates the trends of average CO_2 emissions and carbon intensity over time. The left graph shows a consistent decline in CO_2 emissions, with a shallow decrease from 2020 onwards, potentially influenced by the COVID-19 pandemic. The right panel shows a downward trend in carbon intensity, suggesting efficiency improvements over time.

4.2 Methodology

This study employs two distinct methodologies to test two hypotheses. We first describe the methodology used to test Hypothesis 1 regarding the divergence of ESG ratings. We then explain the approach applied to test Hypothesis 2, which examines the predictive power of ESG ratings on future carbon emissions.

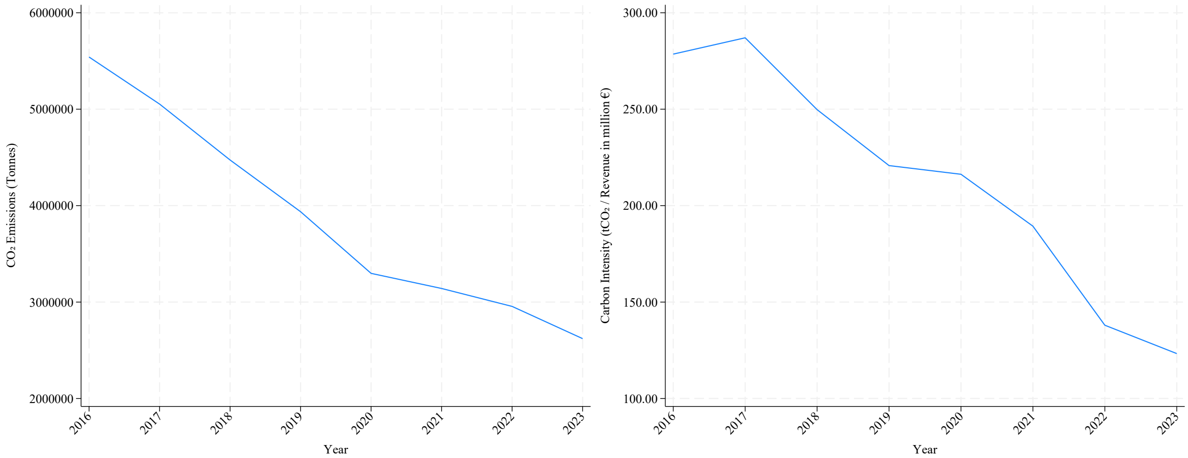

Figure 2: Average CO_2 Emissions and Carbon Intensity Over Time

Notes: Trends of CO_2 and carbon intensity for the period 2014-2024 (based on the full dataset) are presented in Figure 6 in the Appendix.

4.2.1 Methodology for Hypothesis 1: ESG Divergence

To test Hypothesis 1, we begin by conducting a correlation analysis of the ESG ratings. As discussed in section 2, credit ratings typically exhibit near-perfect correlations across providers (with a correlation of 0.99 reported by Berg et al., 2022), suggesting almost identical methodologies applied. In contrast, lower correlations among ESG ratings would indicate a great divergence in how different raters evaluate the same firms.

To analyze the degree of divergence among the raters, we compute pairwise Spearman rank correlations. This approach is appropriate because all ESG ratings in our dataset are at least ordinal in nature (Hauke & Kossowski, 2011). We report both the correlation coefficients and the corresponding significance levels to analyze the strength and the statistical relevance of the relationships.

According to conventional benchmarks by Cohen (1988, pp. 79-80), a correlation of around 0.10 indicates a small relationship, 0.30 a medium relationship, and 0.50 a large relationship. However, since ESG ratings should measure the same underlying construct, we argue that even a correlation as high as 0.50 should be considered relatively low in our context.

To further support the analysis, we visualize all ratings in a scatterplot to illustrate patterns of agreement and divergence across raters more intuitively.

4.2.2 Methodology for Hypothesis 2: ESG and Carbon Emissions

To test Hypothesis 2, we employ a TWFE model. Here, we analyze the effect of ESG ratings on CO_2 emissions and carbon intensity. This approach has multiple advantages over pooled Ordinary Least Squares (OLS), random-effects (RE), and one-way FE models. A TWFE model controls for both individual and time factors that are not observable, that is, the model controls for unobservable heterogeneity (Yang et al., 2024). Time FE are included to account for unobserved shocks or macroeconomic trends, such as changes in climate policy, regulatory pressure, or pandemics that affect all firms in a given year. Many unobservable individual firm characteristics can affect the ESG ratings, for example, firm culture, reputation concerns, or historical environmental incidents. Therefore, the TWFE model is a suitable choice for our analysis. Similar models are employed widely in previous studies for their robustness in panel data settings (Cregan et al., 2024a, 2024b; Qian & Liu, 2024; Xie et al., 2024; Yang et al., 2024).

TWFE models are based on certain assumptions. One of the key assumptions is the assumption of strict exogeneity, $\mathbb{E}(\varepsilon_{it} \mid X_{i1}, X_{i2}, \ldots, X_{iT}) = 0$. This means that the error term must be uncorrelated with all past, present, and future values of the independent variable (Stock & Watson, 2020, pp. 374-376). This assumption is expected to hold in this context, as ESG ratings are externally assessed by rating agencies and are therefore not directly influenced by firm-specific time-varying unobserved factors captured in the error term.

Another assumption of the TWFE model states that observations across entities are independently and identically distributed (i.i.d.). That is, the variables for one firm are assumed to be statistically independent of those for another firm but distributed identically (Stock & Watson, 2020, pp. 374-376). In our case, this assumption may be partially violated, as all firms are located in Europe and may therefore not be independent of each other, particularly within the same sector or country. Additionally, firms from different industries may follow different distributions. However, given the relatively large sample size and the broad sectoral diversification of the STOXX Europe 600 firms, we consider this assumption to be sufficiently addressed in our context. Nonetheless, we cluster standard errors at the firm level to allow for arbitrary correlation and heteroskedasticity within the firm clusters. This ensures that standard errors are not underestimated, which reduces the risk of incorrectly rejecting a null hypothesis due to inflated test statistics (Stock & Watson, 2020, p. 376).

Additionally, the sample should not contain extreme outliers and must not exhibit perfect multicollinearity among the independent variables (Stock & Watson, 2020, pp. 375-376). As shown in Table 2, CO_2 emissions display a wide range, suggesting the presence of potential high outliers. However, Figure 7 in the Appendix indicates that these values are concentrated in highly polluting sectors such as materials, utilities, and energy. In the robustness checks, we include an interaction term for high-pollution industries to assess whether these potential outliers systematically bias our results. Furthermore, the descriptive statistics show that correlations among the explanatory variables are moderate, suggesting that perfect multicollinearity is not an issue in our data.

Finally, we run a Hausman test to decide whether the TWFE model is a more suitable choice than the RE model. While the RE estimator is more efficient under the assumption that the unobserved individual effects are uncorrelated with the explanatory variables (ESG ratings), it becomes inconsistent and biased if this assumption is violated. We expect this assumption to be violated because unobserved firm characteristics, such as corporate culture, long-term sustainability strategies, or geographical location, are likely to influence both the ESG rating and the firm's carbon emissions. The TWFE model allows for correlation between the independent variables and the unobserved (fixed) factors. The Hausman test formally tests whether such a correlation exists. If it does, the TWFE model is preferred as it provides consistent estimates. (Hausman, 1978)

Other empirical strategies could, in principle, help address endogeneity concerns in our setting. Firstly, employing a staggered Difference-in-Differences (DiD) approach, as applied by Li and Xu (2024), would allow us to analyze whether the first introduction of a new ESG rating for a company predicts lower carbon emissions in the future. One could even compare which introduction of which rating has the greater effect. However, this approach is not feasible in our study. Almost all firms in our sample already receive ESG ratings from both providers throughout the observation period, leaving no control group. Furthermore, using a different set of companies, such as smaller firms or those outside Europe, as a control group would be problematic because these companies differ fundamentally in size, sector, and regulatory environment. Therefore, the parallel trends assumption would likely be violated, making valid (staggered) DiD estimation not feasible in our setting. Additionally, comparing the introduction effects of the two ratings is highly problematic. For some firms, both ESG ratings were introduced in the same year, while the ratings were introduced in different years for others. In the latter case, the introduction of the first ESG rating could already influence the firm's CO_2 emissions, making it difficult to isolate the independent effect of the second ESG rating. Therefore, we consider a staggered DiD as not feasible for our study, given data availability.

Secondly, as discussed in section 3, endogeneity concerns due to reverse causality remain a potential issue. Although we will lag the ESG rating by one period in the regressions presented below to mitigate concerns that carbon emissions influence ESG scores, this may not fully solve the problem. Companies that reduce their carbon emissions might also become more willing to increase ESG disclosure, which could lead to higher ESG scores (Qian & Liu, 2024). Therefore, instrumental variable (IV) approaches are often used to address reverse causality, as shown by Qian and Liu (2024), who use the number of shares held by ESG funds and the market value of ESG holdings as instruments. While theoretically attractive, this method is not applicable in our case. First, we do not have access to detailed data on ESG fund ownership at the firm level for the STOXX 600 companies, and building the dataset ourselves is beyond the scope of this thesis. Second, even if such data were accessible, concerns about the validity of the instruments would remain. In European markets, ESG-focused funds may invest directly based on observed sustainability performance, such as lower carbon emissions, which would violate the exclusion restriction necessary for an IV strategy. Before applying this strategy, a thorough understanding of how ESG funds operate in Europe would be required. Therefore, although staggered DiD and IV approaches could theoretically strengthen the thesis, they are not feasible in our case, mainly due to data limitations. Hence, we use a TWFE model.

Equations (1) and (2) show the first two TWFE regression models we use to test Hypothesis 2.

$$CO_{2,i,t} = \beta_0 + \beta_1 ESG_{Ref,i,t-1} + \beta_2 Controls_{i,t} + \mu_i + \lambda_t + \epsilon_{i,t} \tag{1}$$

$$CO_{2,i,t} = \beta_0 + \beta_1 ESG_{Blo,i,t-1} + \beta_2 Controls_{i,t} + \mu_i + \lambda_t + \epsilon_{i,t} \tag{2}$$

$CO_{2,i,t}$ denotes the dependent variables, that is, CO_2 emissions for firm i in year t. $ESG_{Ref,i,t\text{-}1}$ and $ESG_{Blo,i,t\text{-}1}$ represent the independent variables, that is, the ESG ratings (Refinitiv and Bloomberg, respectively) for firm i in year $t\text{-}1$. We lag the ESG scores by one period, following Li and Xu (2024), to account for the time lag between ESG disclosure and its reflection in CO_2 emissions, as well as data availability. Since the mechanism could potentially take several years to show, we also test for longer lags in the robustness tests. β_1 is the estimate of interest, showing the direction and degree of association between ESG scores and subsequent carbon emissions (controlling for time and firm-invariant factors). $Controls_{i,t}$ represents the control variables mentioned earlier for firm i in year t. μ_i captures the time-invariant individual fixed-effects for firm i while λ_t denotes the time fixed-effects in year t. Finally, $\epsilon_{i,t}$ is the error term for firm i in year t. If the β_1 coefficients in these two regressions are negative and statistically significant, we can conclude that Hypothesis 2 holds. Furthermore, differences in the statistical significance or direction of the coefficients may indicate that the two ESG ratings vary in their ability to predict future carbon emissions.

The next two regressions, equation (3) and (4), only differ from the first two regressions in their dependent variable. Here, we use carbon intensity (CI) as the dependent variable, since we are also interested in the effect of the ESG ratings on carbon intensity. This adds an extra layer to our analysis because carbon intensity shows how efficiently firms use carbon emissions to generate revenue.

$$CI_{i,t} = \beta_0 + \beta_1 ESG_{Ref,i,t-1} + \beta_2 Controls_{i,t} + \mu_i + \lambda_t + \epsilon_{i,t} \tag{3}$$

$$CI_{i,t} = \beta_0 + \beta_1 ESG_{Blo,i,t-1} + \beta_2 Controls_{i,t} + \mu_i + \lambda_t + \epsilon_{i,t} \tag{4}$$

5 Results and Analysis

This section shows and discusses the results we obtain from our analysis, structured as follows. First, in section 5.1, we show our main empirical results to test our hypotheses. Next, in section 5.2, we conclude the section by presenting some robustness checks.

5.1 Empirical Results

This section presents the results of our study. We begin by presenting the results to test Hypothesis 1: the divergence of the ESG ratings. Following this, we state the results to test Hypothesis 2: whether ESG ratings can predict future CO_2 emissions.

5.1.1 Results for Hypothesis 1: ESG Divergence

Table 4 presents the correlation matrix of ESG ratings from five distinct raters used to test our first hypothesis: Refinitiv, Bloomberg, Sustainalytics, S&P Global, and MSCI. The coefficient sizes vary greatly among raters. The correlations range from -0.063 (Bloomberg-Sustainalytics) to as high as 0.600 (Refinitiv-S&P Global). Refinitiv and Bloomberg exhibit a moderate positive correlation of 0.488, suggesting some similarity in their scoring. All correlations are statistically significant at the 1% level, except for the pairs Sustainalytics-Refinitiv and Sustainalytics-Bloomberg. This suggests at least a moderate relationship among the respective raters, yet still lower than the correlation among credit rating agencies. Sustainalytics has a negative correlation coefficient with all other raters. We expect that this negative correlation is due to the measurement difference, as Sustainalytics measures risk-based ESG, where 0 indicates the lowest risk, opposite to the direction of other ratings.

Figure 3 visualizes this heterogeneity by plotting the values of ratings from Refinitiv, Bloomberg, S&P Global, and MSCI against Sustainalytics as a benchmark on the x-axis. There is no consistent pattern or strong trend observed, which implies a lack of standardized measurement across ratings. This suggests that ESG scores are highly dependent on raters, and heterogeneity can create confusion for investors who favor sustainable investments, as discussed in section 3.

Table 4: Spearman Rank Correlation Matrix for ESG Ratings in Dataset 1

	Refinitiv	Bloomberg	Sustainalytics	S&P Global	MSCI
Refinitiv	1.000				
Bloomberg	0.488***	1.000			
Sustainalytics	-0.080	-0.063	1.000		
S&P Global	0.600***	0.398***	-0.283***	1.000	
MSCI	0.213***	0.276***	-0.330***	0.294***	1.000

Notes: Correlation coefficients with significance. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

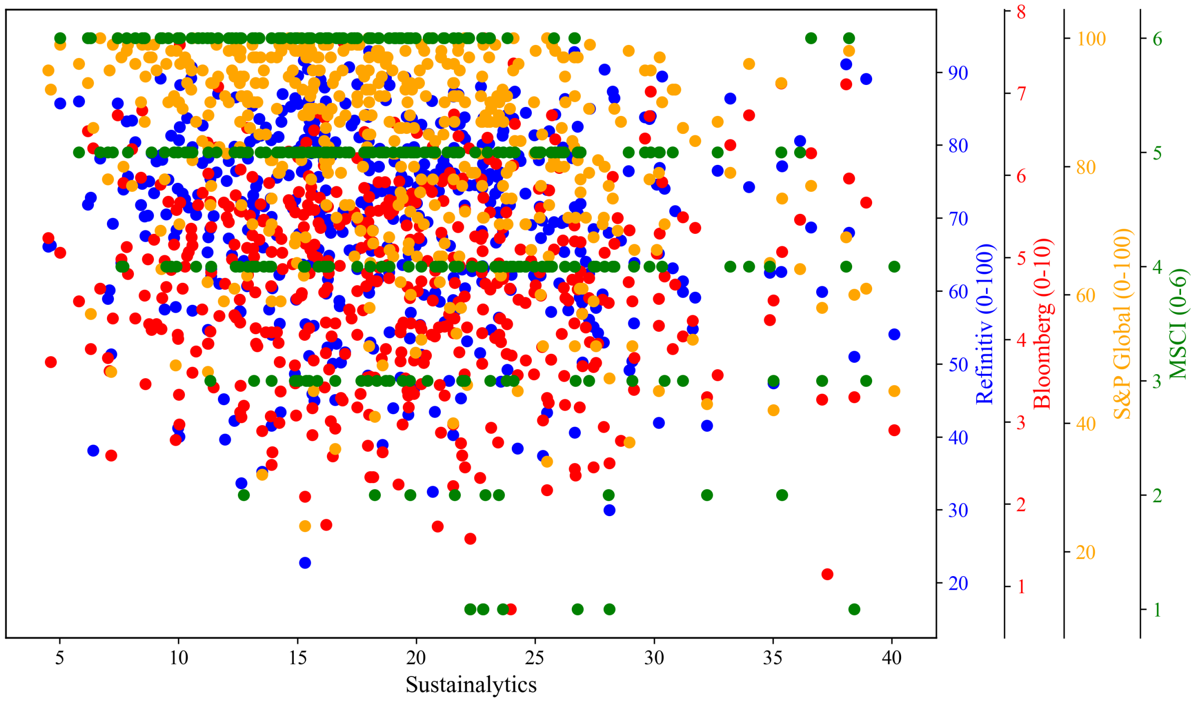

Figure 3: Divergence of ESG Ratings with Sustainalytics on the X-Axis

Notes: Figure 3 illustrates the divergence of ESG ratings. The x-axis shows the value of Sustainalytics ESG as a benchmark for each company (n=458), while the y-axis indicates four other ratings. MSCI ESG ratings are converted into numerical values, from 0 (CCC) to 6 (AAA), with 6 indicating the best performance. MSCI has discrete values, showing up as horizontal lines.

Both the correlation coefficients and the scatterplot strongly support Hypothesis 1, indicating that the ratings significantly diverge, which underlines the lack of standardization in the ESG market.

5.1.2 Results for Hypothesis 2: ESG and Carbon Emissions

Regressions (1) to (4) aim to test Hypothesis 2 of this thesis. The results are shown in Table 5. Regressions (1) and (2) examine the relationship between total CO_2 emissions and the ESG ratings from Refinitiv and Bloomberg, respectively. The results show that the Refinitiv ESG rating has a statistically significant effect on CO_2 emissions (p=0.013). However, the relationship is positive, that is, a one-point increase in the Refinitiv ESG rating in the previous year is associated with a 56629-tonne increase in CO_2 emitted by a firm, cet. par. This implies that firms with higher Refinitiv ESG scores have greater carbon emissions, which contradicts the expected negative relationship in our Hypothesis 2. In contrast, the Bloomberg ESG rating shows a negative relationship, however, it is statistically insignificant (p=0.907).

Regressions (3) and (4) investigate the relationship between carbon intensity and the Refinitiv and Bloomberg ESG ratings, respectively. The findings are similar to those of Regressions (1) and (2). As shown in Table 5, the Refinitiv ESG rating has a statistically significant effect on carbon intensity, with a p-value of 0.015. But again, the relationship is positive: a one-point increase in the lagged Refinitiv ESG rating predicts a 2.75-point increase in the carbon intensity, holding other variables constant. That means that companies with a one-unit higher Refinitiv ESG score are predicted to produce 2.75 tonnes more CO_2 per million euros of revenue. Although the effect size is small, it again contradicts the assumption that higher ESG performance relates to less environmental harm. The Bloomberg ESG rating on carbon intensity remains statistically insignificant (p=0.378), though the results also suggest a positive relationship. The effect size is again small at 7.31.

Taken together, the results provide no evidence in support of Hypothesis 2, namely that higher ESG ratings are effective predictors for reduced CO_2 emissions or improved carbon intensity in the future. Later in section 5.2, we will also test for longer lags to assess whether the mechanism takes several years to show.

5.2 Robustness Checks

To ensure the reliability of our findings, we conduct several robustness checks of the estimated effect of ESG ratings on carbon emissions. These include the Hausman test to

Table 5: Regression Results

	(1) CO_2	(2) CO_2	(3) CarbonIntensity	(4) CarbonIntensity
RefinitivESG (t-1)	56629.133**		2.745**	
	(22797.86)		(1.119)	
BloombergESG (t-1)		-18973.81		7.306
		(162819.1)		(8.287)
Leverage	154209.4	-561408.3	98.028	107.432
	(1701997)	(1674270)	(89.156)	(92.99)
ROA	-2122996	-2346690	-362.625**	-352.345**
	(1804257)	(2029099)	(144.739)	(146.620)
Tobin's Q	13438.36	40922.71	0.633	3.524
	(14827.32)	(41123.91)	(1.318)	(2.904)
Tangible Ratio	2520748	3073846	371.370***	368.633***
	(1547630)	(1588318)	(135.013)	(133.987)
Firm Size	1132810***	1593064***	-30.053	-9.652
	(331809.5)	(383732.8)	(29.221)	(27.01)
Growth Rate	1065766	1005450	-135.729***	-147.14***
	(844920)	(920829.8)	(34.468)	(35.687)
Board Size	341511.2	338149.2	13.907**	12.347**
	(218215.3)	(241906.2)	(5.398)	(5.388)
Independent Board	2373.082	3662.116	0.759	0.367
	(8941.243)	(8807.989)	(0.688)	(0.607)
CEO Duality	861859.3	924300.6	3.787	2.35
	(1147638)	(1293673)	(19.231)	(20.753)
Constant	-2.97e+07***	-3.684e+07***	0.001	1.27e-04
	(8051146)	(9279804)	(0.001)	(0.001)
Observations	3185	2966	3185	2966
R-Squared	0.0670	0.0585	0.1018	0.0979

Notes: Independent variables are Refinitiv (*Refinitiv ESG (t-1)*) and Bloomberg ESG (*Bloomberg ESG (t-1)*), with a one-year lag. All regressions employ a TWFE model, including firm-FE and year dummies. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

test the suitability of the TWFE model we apply, a sensitivity check to control variables, alternative lag structures, the inclusion of both Refinitiv and Bloomberg ESG ratings, and sector-wise analysis. The results are presented in Tables 6 to 9.

We first run the Hausman test for the base Regressions (1) and (2) to evaluate whether the TWFE model is more appropriate than the RE model, as described earlier in section 4.2.2. For the Refinitiv ESG model, the test provides a chi-squared statistic of 152.99 with a p-value below 0.001, strongly rejecting the null hypothesis that the RE estimator is efficient. This confirms the consistency and suitability of the TWFE model (Hausman, 1978). Similarly, the Bloomberg model shows a chi-squared statistic of 81.44 with the 1% significance level (p<0.001), which again rejects the null hypothesis and further supports the employment of the TWFE approach.

Second, a sensitivity check for control variables is conducted to assess the stability of the ESG coefficients when accounting for other factors. Starting from a model that includes only each ESG score, we run a series of regressions by adding control variables stepwise, and the results are reported in Table 6. For Refinitiv ESG, the coefficient remains positive and statistically significant across all model specifications, indicating that the relationship is robust to the inclusion of controls. Although its significance slightly declines as more variables are added, R-squared gradually increases, which suggests improved explanatory power. In contrast, the Bloomberg ESG score remains statistically insignificant throughout, implying a weaker and less robust relationship with CO_2 emissions.

Table 6: Sensitivity Check: Effect of ESG on CO_2 with Controls

	Refinitiv				**Bloomberg**			
Control	**Coef.**	**Std.Err.**	$\mathbf{R^2}$	**Obs.**	**Coef.**	**Std.Err.**	$\mathbf{R^2}$	**Obs.**
ESG	60397.39***	20511.84	0.0409	3239	13080.69	158704.8	0.0305	3016
+ lev	62111.46***	21663.11	0.0421	3208	22622.63	163232.5	0.0314	2987
+ roa	62107.51***	21570.60	0.0426	3200	26445.94	162345.3	0.0320	2980
+ tobq	61844.20***	21652.71	0.0422	3190	18495.86	162786.0	0.0315	2971
+ tang	59160.58***	22251.09	0.0436	3186	828.92	160619.2	0.0337	2967
+ firsiz	50027.97**	22063.26	0.0500	3186	-37720.96	157165.0	0.0431	2967
+ grow	55159.52***	21998.58	0.0567	3185	-17115.38	161431.9	0.0491	2966
+ bosiz	55888.04***	22067.53	0.0651	3185	-4695.40	162370.5	0.0565	2966
+ indep	55687.66**	22928.23	0.0651	3185	-8663.77	164199.3	0.0566	2966
+ dual[a]	56629.13**	22797.86	0.0670	3185	-18973.81	162819.1	0.0585	2966

Notes: Dependent variable is CO_2 Emissions. All regressions employ a TWFE model, including firm-FE and year dummies. The ESG variables are lagged by one year. [a] Refers to Regressions (1) and (2) in our analysis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Third, we examine whether our results are sensitive to the timing of the ESG effect by testing alternative lags. While our main model employs a one-year lag (*t-1*), we additionally run regressions with two- (*t-2*) and three-year (*t-3*) lags. The results in Table 7 show that across all lag structures, the estimated coefficients for Refinitiv ESG remain positive and statistically significant at the 5% level, indicating a stable and persistent relationship. The magnitude of the effect decreases with longer lags, suggesting a diminishing influence of ESG performance on carbon emissions over time. Although R-squared slightly increases at the three-year lag, the number of observations drops due to data limitations, which makes longer lags less feasible in the main analysis. On the other hand, Bloomberg ESG remains statistically insignificant across all lag lengths, consistent with earlier findings and further highlighting its limited explanatory power.

Table 7: Effect of ESG on CO_2 Emissions with Different Lags

	Refinitiv				**Bloomberg**			
Lag	**Coef.**	**Std.Err.**	$\mathbf{R^2}$	**Obs.**	**Coef.**	**Std.Err.**	$\mathbf{R^2}$	**Obs.**
t-1[a]	56629.13**	22797.86	0.067	3185	-18973.81	162819.1	0.0585	2966
t-2	44879.45**	18335.12	0.063	2824	72077.82	141768.7	0.0563	2629
t-3	38861.19**	16025.17	0.073	2410	23984.38	164376.9	0.0674	2237

Notes: Dependent variable is CO_2 Emissions. All regressions employ a TWFE model, including firm-FE and year dummies. [a] Refers to Regressions (1) and (2) in our analysis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Fourth, we run an extra regression including both Refinitiv and Bloomberg ESG ratings to assess (i) whether the two ratings have independent effects on carbon emissions, and (ii) whether the relationship between ESG scores and CO_2 emissions changes when both ratings are considered simultaneously. While our main models estimate the effects of Refinitiv and Bloomberg ESG scores separately, this specification includes both variables in a single regression. The results are presented in Table 8. When used independently (Regression (1)), the Refinitiv ESG score shows a positive and statistically significant effect at the 5% level (p=0.013). When both ESG scores are included simultaneously, the Refinitiv coefficient remains positive but becomes less significant with a p-value of 0.071, indicating some degree of independent explanatory power. The Bloomberg ESG score, which presents a negative and statistically insignificant result when used alone (Regression (2)), becomes highly significant (p<0.001) in the combined model. This change may reflect multicollinearity or overlap between the two ESG metrics. Despite the increased statistical significance of the Bloomberg score, the overall explanatory power (R-squared) of the model decreases, suggesting that including both scores may introduce collinearity concerns rather than adding meaningful explanatory value.

Last, we examine whether ESG ratings are more or less predictive of future carbon performance in high-polluting sectors, which are major contributors to the total carbon emissions (see Figure 7 in the Appendix). Following the GICS industry sectors (MSCI,

Table 8: Effect of ESG Scores on CO_2 Emissions when Both Ratings Are Included

Model	**Variable**	**Coefficient**	**Std.Err.**	$\mathbf{R^2}$	**Obs.**
Refinitiv[a]	RefinitivESG (t-1)	56629.13**	22797.86	0.0670	3185
Bloomberg[b]	BloombergESG (t-1)	-18973.81	162819.1	0.0585	2966
Both ESG Ratings	RefinitivESG (t-1)	36742.95*	20277.44	0.0416	2934
	BloombergESG (t-1)	-873003.4***	233262.7	0.0416	2934

Notes: Dependent variable is CO_2 Emissions. All models employ a TWFE model, including firm-FE and year dummies. The ESG variables are lagged by one year. [a,b] Refers to Regressions (1) and (2) in our analysis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2024), we create a dummy for high-polluting industries HP_i, which takes on the value 1 if firm i belongs to a high-polluting sector: utilities, materials, and energy, and equals 0 otherwise. This helps account for potential outliers concentrated in carbon-intensive industries and allows us to assess whether the main results are driven by these sectors or reflect broader patterns across the sample.

Regressions (5) to (8) introduce interaction terms between ESG ratings and industry pollution level to test whether the effect of ESG ratings on CO_2 emissions or carbon intensity varies between high- and low-polluting industries. Since firms do not change sectors over time in our sample, we omit the time component (t) of the term and the industry dummy variable in the specifications.

$$CO_{2,i,t} = \beta_0 + \beta_1 ESG_{Ref,i,t-1} + \beta_2 \left(ESG_{Ref,i,t-1} \times HP_i\right) + \beta_3 Controls_{i,t} + \mu_i + \lambda_t + \epsilon_{i,t} \quad (5)$$

$$CO_{2,i,t} = \beta_0 + \beta_1 ESG_{Blo,i,t-1} + \beta_2 \left(ESG_{Blo,i,t-1} \times HP_i\right) + \beta_3 Controls_{i,t} + \mu_i + \lambda_t + \epsilon_{i,t} \quad (6)$$

$$CI_{i,t} = \beta_0 + \beta_1 ESG_{Ref,i,t-1} + \beta_2 \left(ESG_{Ref,i,t-1} \times HP_i\right) + \beta_3 Controls_{i,t} + \mu_i + \lambda_t + \epsilon_{i,t} \quad (7)$$

$$CI_{i,t} = \beta_0 + \beta_1 ESG_{Blo,i,t-1} + \beta_2 \left(ESG_{Blo,i,t-1} \times HP_i\right) + \beta_3 Controls_{i,t} + \mu_i + \lambda_t + \epsilon_{i,t} \quad (8)$$

The regression results are presented in Table 9. It indicates that Refinitiv ESG ratings are significant at the 1% level with both CO_2 emissions (p=0.003) and carbon intensity (p=0.001) in low-polluting sectors. However, coefficients are positive: a one-point increase in the Refinitiv ESG is associated with an increase in 58,190 tonnes of total CO_2 emissions and 3.19 tonnes in carbon intensity. These findings suggest that ESG improvements in low-polluting sectors, as measured by Refinitiv, are not necessarily associated with emission reductions and may even indicate increased emissions in the future. The interaction terms with high-polluting sectors are statistically insignificant with p-values of 0.851 and 0.238, implying no meaningful difference in the effect of Refinitiv ESG ratings between high- and low-polluting sectors. This suggests that the positive relationship found in our regression results (Regressions (1) & (3)) is primarily driven by low-polluting sectors, and that Refinitiv ESG does not significantly differentiate across industries.

In contrast, Bloomberg ESG scores show statistically significant and negative interaction effects in high-polluting sectors, indicating carbon emission reductions: a one-point increase in the Bloomberg rating predicts a reduction of 2.41 million tonnes of CO_2 emissions (p=0.002) and 124.63 tonnes of carbon emissions per million euros of revenue (p<0.001). This implies that Bloomberg ESG scores may capture or reflect emission reduction efforts in industries where emission risks are intensive. In low-polluting sectors, Bloomberg ESG ratings are positively but only marginally associated with CO_2 emissions (p=0.061), and positively significantly associated with higher carbon intensity (p=0.004): a one-point increase in Bloomberg ESG corresponds to an increase of 378,440 tonnes in carbon emissions and 29.244 in carbon intensity (tonnes CO_2 per million euros of revenue).

Table 9: Regression Results with Interaction Terms

	(5) CO_2	(6) CO_2	(7) CarbonIntensity	(8) CarbonIntensity
RefinitivESG (t-1)	58190.26***		3.191***	
	(19575.13)		(0.925)	
BloombergESG (t-1)		378440*		29.244***
		(201567)		(10.049)
Refinitiv (t-1) x HP	-12593.53		-3.598	
	(67025.51)		(3.043)	
Bloomberg (t-1) x HP		-2787367***		-153.873***
		(909976.9)		(37.928)
Observations	3185	2966	3185	2966
R-Squared	0.0670	0.1030	0.1029	0.1345

Notes: Independent variables are Refinitiv (*Refinitiv ESG (t-1)*) and Bloomberg ESG (*Bloomberg ESG (t-1)*), with a one-year lag. All models employ a TWFE model, including firm-FE and year dummies. HP refers to a dummy for high-polluting sectors: materials, utilities, and energy. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

These findings offer important insights into the regression results (see Table 5). While lagged Refinitiv ESG scores show a consistent and statistically significant positive association with future emissions, their effects do not vary notably across sectors. This indicates that potential outliers concentrated in carbon-intensive industries do not systematically bias our results for Refinitiv ESG. In contrast, Bloomberg ESG scores, which appear statistically insignificant in Regressions (2) and (4), show significant effects in both low- and high-polluting sectors once the sectoral interactions are accounted for. This highlights the importance of controlling for industry heterogeneity and suggests that the influence of Bloomberg ratings on CO_2 emissions may be sensitive to a sector's pollution intensity.

6 Discussion

The following section reflects on our main findings presented in section 5. We conduct a critical discussion of the results and highlight potential limitations.

The empirical results presented in section 5 reveal several important findings. Our first hypothesis states that ESG ratings significantly diverge across providers. The results presented in section 5.1.1 support this hypothesis, showing that ESG scores assigned to companies in the STOXX Europe 600 index differ considerably among the five distinct ESG ratings. This finding is consistent with earlier research (e.g., Berg et al., 2022) who find correlations as low as 0.38, highlighting methodological inconsistencies across raters. By focusing on a broad European sample, our analysis suggests that the divergence across raters is not subject to a specific sector, but is present across the whole European market. The substantial disagreement between ratings poses fundamental challenges. Rather than reducing information asymmetries, ESG ratings may introduce additional uncertainty and therefore undermine their intended role in guiding sustainable investment decisions.

Our second hypothesis states that higher ESG ratings would predict lower future CO_2 emissions. However, the results do not support this assumption. Contrary to our expectations, the Refinitiv ESG score has a positive relationship with both total CO_2 emissions and carbon intensity, with both relationships being statistically significant. In contrast, the Bloomberg ESG rating does not show a statistically significant relationship with either emissions measure. Interestingly, the sign of the coefficients diverges. The Bloomberg ESG score correlates negatively with total emissions, but positively with carbon intensity. Although these effects are not statistically significant, the inconsistency reinforces doubts about the ability of ESG scores to reliably predict environmental outcomes. Overall, the results suggest that higher ESG ratings may not predict better environmental performance in the future. One possible explanation is greenwashing. Firms may enhance their ESG scores by increasing the volume and quality of disclosures, without implementing substantive emission reduction measures. This would be consistent with prior research showing that ESG ratings often rely on self-reported and input-based data, rather than verified environmental outcomes (e.g., Lee, 2021; OECD, 2025).

Interestingly, our findings contradict prior studies from the Chinese market, for example, Li and Xu (2024) suggest that ESG ratings are able to predict carbon emissions. The difference might be due to the different methodologies in ESG ratings. As mentioned earlier, the OECD (2025) states that, for example, the weighting of environmental factors can differ greatly across raters. Therefore, we conclude that the predictive power of ESG ratings cannot be generalized across raters.

Furthermore, our results show that sustainable investors may face serious challenges when using ESG ratings to guide their investments. Different ESG ratings diverge greatly and, at least the two ESG ratings tested, struggle to predict future carbon emissions. This implies that ESG ratings cannot be used as consistent indicators of environmental sustainability. For investors seeking to reduce the cognitive demand of evaluating complex sustainability data, ESG ratings offer very limited assistance under the current conditions. This is particularly concerning because Amel-Zadeh and Serafeim (2018) show that non-ESG specialists consider ESG information even for materiality reasons.

Our study has several limitations. First, we are only able to include two ESG ratings in the analysis to test the second hypothesis due to limited data availability. A broader comparison that includes other prominent rating providers, such as Sustainalytics, S&P Global, or MSCI, would provide a more complete picture of ESG rating divergence. It would also allow us to evaluate which ratings, if any, are most effective in predicting future CO_2 performance.

Second, some control variables that could improve the results of our study are not included because they were not available to us. For example, R&D expenditures would have been a valuable control since we expect them to be correlated with future carbon emissions. Furthermore, including R&D expense would allow for better comparison with previous studies that have included it (e.g., Li & Xu, 2024).

Third, we acknowledge a potential limitation in our predictive analysis. Firms may selectively respond to specific ESG ratings. While our fixed-effects model accounts for firm-specific characteristics and time trends, it cannot capture the possibility that companies prioritize certain ratings over others when making sustainability-related decisions. For example, one firm with a high Bloomberg ESG score and a low Refinitiv score might reduce its CO_2 emissions in response to the Bloomberg ESG rating, while another firm might respond to the Refinitiv ESG instead. In such cases, changes in emissions can be better predicted by the ESG rating that the firm actively tracks, yet this information is unobservable in our dataset. This selective responsiveness could reduce the predictive power of the ratings we include, and it complicates the interpretation of observed relationships. It also limits our ability to assess whether firms with a high ESG rating and high emissions levels actually do greenwashing or simply react to different rating providers.

Finally, while this study focuses on the European STOXX 600 firms, the results may not be fully generalizable to other markets. As we show in section 2, ESG ratings tend to differ across regions. While most studies in the Chinese market find a significant negative relationship between ESG ratings and carbon performance, studies in other regions tend to struggle to confirm the connection. This shows that one has to be careful to extend our findings to other regions.

7 Concluding Remarks

In this thesis, we addressed two research questions. First, we examined the degree of divergence among the five prominent ESG rating providers. Second, we investigated the extent to which ESG ratings predict future corporate carbon emissions, using panel data from 478 non-financial European firms between 2015 and 2023. We find substantial divergence across ESG scores and limited predictive power regarding environmental performance.

Specifically, the analysis confirms significant differences among the five raters we include. Furthermore, higher ESG ratings from Refinitiv and Bloomberg do not reliably predict lower future CO_2 emissions or carbon intensity. These findings raise concerns about the credibility of ESG scores as indicators of real environmental performance and point to the risk of greenwashing. They also highlight the limitations of ESG ratings as simplified decision-making tools for sustainable investors.

The results suggest that ESG ratings, in their current form, lack sufficient standardization and are not adequate to serve as robust environmental performance signals. This has important implications for policymakers. First, the raters need to be more transparent about what they measure and how the ratings are created. Second, ESG frameworks should shift toward outcome-based environmental indicators (e.g., actual emissions data) rather than relying predominantly on qualitative or input-based measures (OECD, 2025). Without changes, investors may continue to allocate investments based on inconsistent or misleading ESG signals, which undermines their sustainability objectives.

This study calls for future research. First, future studies could include additional ESG ratings, e.g., S&P Global or MSCI, to study which ratings are most reliable in estimating future corporate carbon performance. Second, future studies could explore other environmental dimensions (e.g., water usage, biodiversity) or expand the scope to social outcomes, such as diversity and workplace safety. Third, it would be valuable to examine whether firms selectively respond to specific ESG ratings, which we were unable to study with our quantitative dataset. For this, qualitative data, such as interviews, would be needed. Lastly, future research could assess whether regulations, such as the EU Taxonomy, can improve the consistency and effectiveness of ESG ratings over time.

References

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

Atan, R., Alam, M. M., Said, J., & Zamri, M. (2018). The impacts of environmental, social, and governance factors on firm performance: Panel study of malaysian companies. *Management of Environmental Quality: An International Journal*, *29*(2), 182–194. https://doi.org/10.1108/MEQ-03-2017-0033

Avramov, D., Cheng, S., Lioui, A., & Tarelli, A. (2022). Sustainable investing with ESG rating uncertainty. *Journal of Financial Economics*, *145*(2), 642–664. https://doi.org/10.1016/j.jfineco.2021.09.009

Bauer, R., Ruof, T., & Smeets, P. (2021). Get Real! Individuals Prefer More Sustainable Investments (S. Van Nieuwerburgh, Ed.). *The Review of Financial Studies*, *34*(8), 3976–4043. https://doi.org/10.1093/rfs/hhab037

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

Billio, M., Costola, M., Hristova, I., Latino, C., & Pelizzon, L. (2021). ESG ratings: (dis)agreement and performance. *Corporate Social Responsibility and Environmental Management*, *28*(5), 1426–1445. https://doi.org/10.1002/csr.2177

Bloomberg. (2023, November). *Bloomberg ESG Scores Overview & FAQ* (tech. rep.). Bloomberg. https://hr.bloombergadria.com/data/files/Pitanja%20i%20odgovori%20o%20Bloomberg%20ESG%20Scoreu.pdf

Bloomberg. (2024, February 8). *Global ESG assets predicted to hit $40 trillion by 2030, despite challenging environment, forecasts bloomberg intelligence*. Retrieved April 22, 2025, from https://www.bloomberg.com/company/press/global-esg-assets-predicted-to-hit-40-trillion-by-2030-despite-challenging-environment-forecasts-bloomberg-intelligence/

Bolton, P., & Kacperczyk, M. (2021). Do investors care about carbon risk? *Journal of Financial Economics*, *142*(2), 517–549. https://doi.org/10.1016/j.jfineco.2021.05.008

Capizzi, V., Gioia, E., Giudici, G., & Tenca, F. (2021). THE DIVERGENCE OF ESG RATINGS: AN ANALYSIS OF ITALIAN LISTED COMPANIES. *Journal of Financial Management, Markets and Institutions*, *09*(2), 2150006. https://doi.org/10.1142/S2282717X21500067

Chen, Z., & Xie, G. (2022). ESG disclosure and financial performance: Moderating role of ESG investors. *International Review of Financial Analysis*, *83*, 102291. https://doi.org/10.1016/j.irfa.2022.102291

Chytis, E., Eriotis, N., & Mitroulia, M. (2024). ESG in business research: A bibliometric analysis. *Journal of Risk and Financial Management*, *17*(10), 460. https://doi.org/10.3390/jrfm17100460

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Cong, Y., Zhu, C., Hou, Y., Tian, S., & Cai, X. (2022). Does ESG investment reduce carbon emissions in china? *Frontiers in Environmental Science*, *10*, 977049. https://doi.org/10.3389/fenvs.2022.977049

Cregan, C., Kelly, J. A., & Clinch, J. P. (2024a). Are environmental, social and governance (ESG) ratings reliable indicators of emissions outcomes? a case study of the airline industry. *Corporate Social Responsibility and Environmental Management*, *31*(2), 909–928. https://doi.org/10.1002/csr.2608

Cregan, C., Kelly, J. A., & Clinch, J. P. (2024b). Corporate climate scores, carbon emissions, and investment in decarbonisation in the steel industry. what do ratings tell us? *Environmental and Sustainability Indicators*, *24*, 100481. https://doi.org/10.1016/j.indic.2024.100481

Crowley, T. J. (2000). Causes of climate change over the past 1000 years. *Science*, *289*(5477), 270–277. https://doi.org/10.1126/science.289.5477.270

Darnall, N., & Aragón-Correa, J. A. (2014). Can Ecolabels Influence Firms' Sustainability Strategy and Stakeholder Behavior? *Organization & Environment*, *27*(4), 319–327. https://doi.org/10.1177/1086026614562963

De Freitas Netto, S. V., Sobral, M. F. F., Ribeiro, A. R. B., & Soares, G. R. D. L. (2020). Concepts and forms of greenwashing: A systematic review. *Environmental Sciences Europe*, *32*(1), 19. https://doi.org/10.1186/s12302-020-0300-3

Doyle, T. M. (2018). Ratings that don't rate: The subjective world of ESG ratings agencies. *American Council for Capital Formation*. https://accf.org/wp-content/uploads/2018/07/071818_ACCF_RatingsESGReport_PUB01.pdf

Drempetic, S., Klein, C., & Zwergel, B. (2020). The influence of firm size on the ESG score: Corporate sustainability ratings under review [Publisher: Springer]. *Journal of Business Ethics*, *167*(2), 333–360. Retrieved April 22, 2025, from http://www.jstor.org/stable/45386660

European Commission. (2025, February 26). *Commission simplifies rules on sustainability and EU investments, delivering over €6 billion in administrative relief* [European commission - european commission]. Retrieved April 22, 2025, from https://ec.europa.eu/commission/presscorner/detail/en/https:%5C/%5C/ec.europa.eu%5C/commission%5C/presscorner%5C/detail%5C/en%5C/ip_25_614

Federal Register. (2025, January 30). *Putting america first in international environmental agreements* [Federal register]. Retrieved April 22, 2025, from https://www.federalregister.gov/documents/2025/01/30/2025-02010/putting-america-first-in-international-environmental-agreements

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, *5*(4), 210–233. https://doi.org/10.1080/20430795.2015.1118917

Gyönyörová, L., Stachoň, M., & Stašek, D. (2023). Esg ratings: Relevant information or misleading clue? evidence from the s&p global 1200. *Journal of Sustainable Finance & Investment*, *13*(2), 1075–1109. https://doi.org/10.1080/20430795.2021.1922062

Hauke, J., & Kossowski, T. (2011). Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. *QUAGEO*, *30*(2), 87–93. https://doi.org/10.2478/v10117-011-0021-1

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *46*(6), 1251. https://doi.org/10.2307/1913827

International Labour Organization. (2022, May). *G20 Sustainable Finance Working Group Input Paper: Finance for a Just Transition and the Role of Transition Finance. Input paper prepared for the G20 Sustainable Finance Working Group under the Indonesian Presidency* (tech. rep.). International Labour Organization (ILO). Retrieved May 19, 2025, from https://www.ilo.org/sites/default/files/wcmsp5/groups/public/%40dgreports/%40dcomm/documents/publication/wcms_848640.pdf

Kolling, C., & Busch, T. (2025). Sustainability labels matter. *Finance Research Letters*, *73*, 106609. https://doi.org/10.1016/j.frl.2024.106609

Lashgari, M. (2024). Sustainable Investments: Return and Risk. *Journal of Applied Business and Economics*, *26*(4). https://doi.org/10.33423/jabe.v26i4.7265

Lee, L.-E. (2021). What does ESG investing really mean? measuring materiality. *Pension Research Council Symposium*. https://pensionresearchcouncil.wharton.upenn.edu/wp-content/uploads/2021/04/LeeCombined_OSM-4.9.21.pdf

Li, F., & Polychronopoulos, A. (2020, January). *What a Difference an ESG Ratings Provider Makes!* https://www.researchaffiliates.com/insights/esg

Li, J., & Xu, X. (2024). Can ESG rating reduce corporate carbon emissions? – An empirical study from Chinese listed companies. *Journal of Cleaner Production*, *434*, 140226. https://doi.org/10.1016/j.jclepro.2023.140226

Löfgren, Å., & Nordblom, K. (2020). A theoretical framework of decision making explaining the mechanisms of nudging. *Journal of Economic Behavior & Organization*, *174*, 1–12. https://doi.org/10.1016/j.jebo.2020.03.021

LSEG Data & Analytics. (2024, October). *Environmental, Social and Governance scores from LSEG* (tech. rep.). LSEG Data & Analytics. https://www.lseg.com/content/dam/data-analytics/en_us/documents/methodology/lseg-esg-scores-methodology.pdf?esg=Qube+Holdings+Ltd

Matos, P. (2020). ESG and responsible institutional investing around the world: A critical review. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3668998

Morningstar Sustainalytics. (n.d.). *ESG Risk Ratings* (tech. rep.). Morningstar Sustainalytics. https://connect.sustainalytics.com/hubfs/INV/ESG%20Risk%20Ratings/Sustainalytics%20ESG%20Risk%20Ratings%20Brochure.pdf

MSCI. (n.d.). *MSCI Sustainblity Solutions* (tech. rep.). MSCI. https://www.msci.com/documents/1296102/23164137/Sustainability_Solutions_Brochure_Overview_FINAL_082024.pdf/2f6f9052-448c-34a5-e31c-385a345987c9?t=1736373971342

MSCI. (2024, August). *GLOBAL INDUSTRY CLASSIFICATION STANDARD (GICS) METHODOLOGY* (tech. rep.). MSCI. https://www.msci.com/indexes/documents/methodology/1_MSCI_Global_Industry_Classification_Standard_GICS_Methodology_20240801.pdf

OECD. (2020, September). *ESG Investing: Practices, Progress and Challenges*. https://doi.org/10.1787/b4f71091-en

OECD. (2025, February). *Behind ESG ratings: Unpacking sustainability metrics*. https://doi.org/10.1787/3f055f0c-en

Pástor, Ľ., Stambaugh, R. F., & Taylor, L. A. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics*, *142*(2), 550–571. https://doi.org/10.1016/j.jfineco.2020.12.011

Qian, Y., & Liu, Y. (2024). Improve carbon emission efficiency: What role does the ESG initiatives play? *Journal of Environmental Management*, *367*, 122016. https://doi.org/10.1016/j.jenvman.2024.122016

Higher ESG Ratings improve Carbon Emission Efficiency. Study was being done in China. Listed Companies. What is CEE?

Sciarelli, M., Cosimato, S., Landi, G., & Iandolo, F. (2021). Socially responsible investment strategies for the transition towards sustainable development: The importance of integrating and communicating ESG. *The TQM Journal*, *33*(7), 39–56. https://doi.org/10.1108/TQM-08-2020-0180

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, *134*(2), 207–222. https://doi.org/10.1037/0033-2909.134.2.207

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *69*(1), 99. https://doi.org/10.2307/1884852

S&P Global. (2025, March). *S&P Global ESG Scores Methodology* (tech. rep.). S&P Global. https://portal.s1.spglobal.com/survey/documents/spglobal_esg_scores_methodology.pdf

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, *87*(3), 355. https://doi.org/10.2307/1882010

Stock, J. H., & Watson, M. W. (2020). *Introduction to econometrics* (4th ed.) [Global Edition]. Pearson Education Limited.

STOXX. (n.d.-a). Index Products - STOXX Europe 600. Retrieved June 9, 2025, from https://stoxx.com/index/sxxp/

STOXX. (n.d.-b). Selection Lists - STOXX. Retrieved April 4, 2025, from https://www.stoxx.com/web/stoxxcom/selection-lists

Tang, H., Tong, M., & Chen, Y. (2024). Green investor behavior and corporate green innovation: Evidence from Chinese listed companies. *Journal of Environmental Management*, *366*, 121691. https://doi.org/10.1016/j.jenvman.2024.121691

The White House. (2025, April). *Executive order on protecting american energy from state overreach* [Presidential Actions, The White House]. Retrieved May 22, 2025, from https://www.whitehouse.gov/presidential-actions/2025/04/protecting-american-energy-from-state-overreach/

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

U.S. Supreme Court. (1976, June 14). *TSC Industries, Inc. v. Northway, Inc., 426 U.S. 438 (1976)*. Retrieved May 7, 2025, from https://supreme.justia.com/cases/federal/us/426/438/

Velte, P. (2017). Does ESG performance have an impact on financial performance? evidence from germany. *Journal of Global Responsibility*, *8*(2), 169–178. https://doi.org/10.1108/JGR-11-2016-0029

Wang, M., Liu, L., & Liang, T. (2025). Does the ESG disclosure quality affect financial performance: Empirical evidence from chinese energy-listed companies. *Finance Research Letters*, *77*, 107118. https://doi.org/10.1016/j.frl.2025.107118

Whelan, T., Atz, U., & Clark, C. (n.d.). ESG AND FINANCIAL PERFORMANCE:

Xie, H., Qin, Z., & Li, J. (2024). ESG performance and corporate carbon emission intensity: Based on panel data analysis of a-share listed companies. *Frontiers in Environmental Science*, *12*, 1483237. https://doi.org/10.3389/fenvs.2024.1483237

Xu, L., Fan, M., Yang, L., & Shao, S. (2021). Heterogeneous green innovations and carbon emission performance: Evidence at China's city level. *Energy Economics*, *99*, 105269. https://doi.org/10.1016/j.eneco.2021.105269

Yang, F., Guo, J., Kong, X., Wang, C., & Wang, Z. (2024). Impact of ESG ratings on carbon emission reduction performance in context of green development—evidence from chinese manufacturing listed enterprises. *Business Process Management Journal*. https://doi.org/10.1108/BPMJ-04-2024-0253

A Appendix

Table 10: Example of Reverse Causality Between ESG Ratings and CO_2 Emissions

Year	**CO_2 Emissions**	**ESG Rating**
t-1	Low	
t	Low	**High**
t+1	**High**	High

Notes: Table 10 illustrates how reverse causality can lead to a misleading positive relationship between ESG ratings and future CO_2 emissions. Suppose a firm has low CO_2 emissions in year *t-1* and *t*. In the presence of reversed causality, where ESG scores are assigned based on past emissions rather than affecting them, i.e., if lower carbon emissions in the given year lead to higher ESG ratings in the subsequent year, this results in a high ESG rating in years *t* and *t+1*, respectively. If the firm then increases its CO_2 emissions in year *t+1*, in this case, a high ESG rating in *t* is followed by high emissions in *t+1*. This indicates a positive relationship between current ESG ratings and future carbon emissions, which contradicts our Hypothesis 1. Such a pattern may arise if companies do not sustain emission reductions after initially receiving high ESG scores.

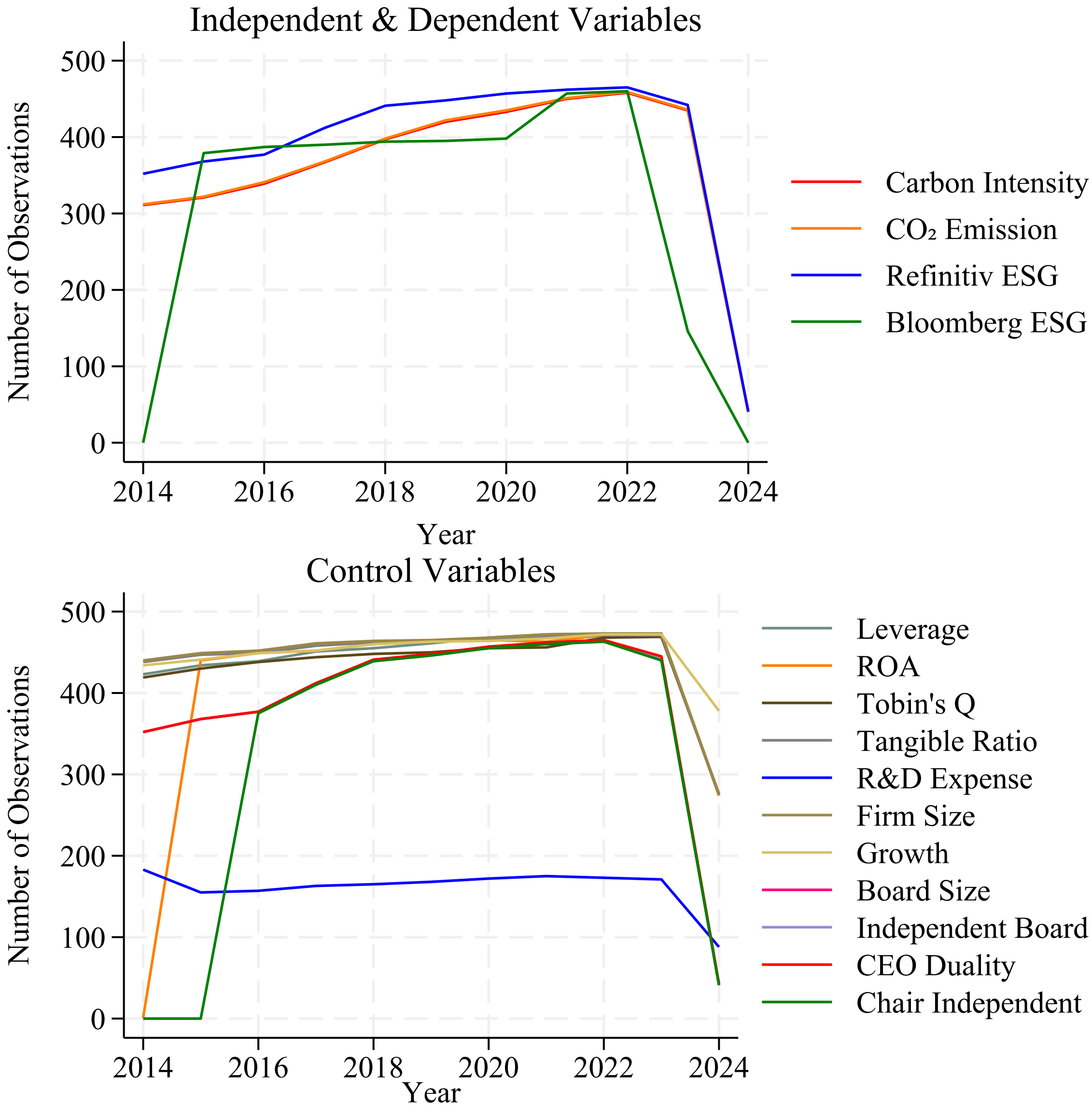

Figure 4: Number of Observations Over Time

Notes: Figure 4 illustrates the number of observations for each variable over time. In the upper graph, which shows our independent and dependent variables, we observe a limited number of observations for Bloomberg ESG scores (represented by the green line) prior to 2015 and from 2023 onwards. Due to this data limitation, the time frame of our study includes the ESG scores for the years 2015 to 2022. Furthermore, the number of observations for research and development (R&D) expense, drawn in the blue line in the lower graph, is significantly less than other variables. Consequently, we decided to exclude this variable from our analysis. CEO-Chairman duality (*CEO Duality*) and chairperson independence dummy (*Chair Independent*) variables essentially explain the same information, both indicating whether the firm's chairperson is independent or not. For example, when CEO duality is true (=1), that means that the chairperson independence variable is false. As shown in the lower graph, the chairperson's independence (green line) lacks observations in 2015. Therefore, we opted to include CEO duality in our analysis instead of chairperson independence, due to the sample size.

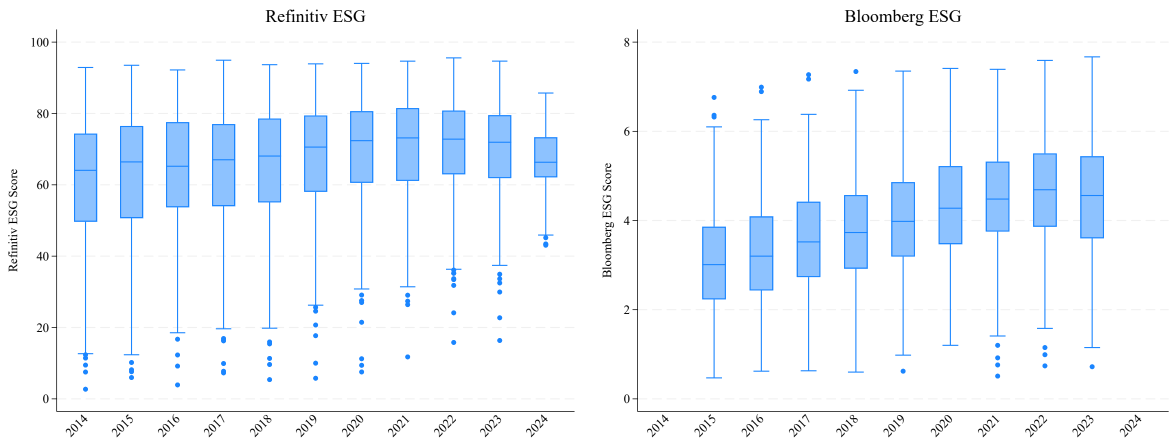

Figure 5: Variation in Refinitiv & Bloomberg ESG Scores Over Time

Notes: Figure 5 illustrates the variations over time in our primary ESG scores utilized in Dataset 2: Refinitiv and Bloomberg. Both ESG scores show gradual improvements over the years, except for 2024. For Refinitiv ESG (left graph), we observe a convergence in variation over time. For Bloomberg (right graph), the ranges remain relatively similar across years with a slight increase each year. Bloomberg revised its ESG rating measure in 2022, which could have caused notable differences before and after the change. However, no significant shift in trend is observed from 2019 to 2020, even when Bloomberg changed its scoring methodology. Therefore, we use the Bloomberg ESG score consistently throughout the period selected for our study, i.e., from 2015 to 2022.

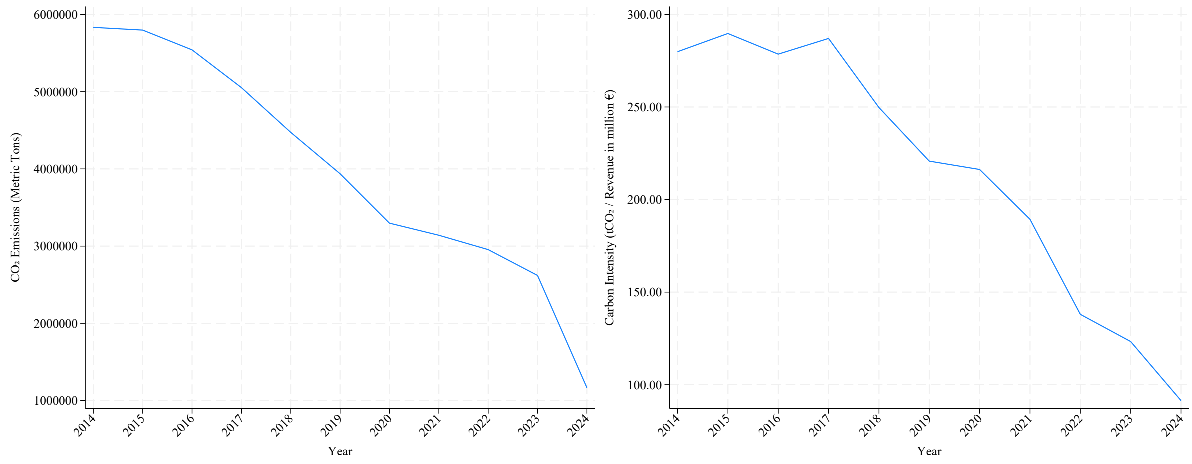

Figure 6: Average CO_2 Emissions & Carbon Intensity Over Time

Notes: Figure 6 illustrates the trends of average CO_2 emissions and carbon intensity from 2014 to 2024, which are the full years of the dataset we obtained. A particularly sharp drop is observed from 2023 to 2024, which may be due to biased reporting. Since most companies have not disclosed their carbon emissions for 2024, it is likely that only companies with strong commitments to sustainability or those that prioritize mitigating carbon emissions have disclosed the most recent data. As a result, we observe a very low level of carbon emissions in 2024. Similarly, the right panel shows a steep decline in carbon intensity from 2023 to 2024, which may be due to the biased disclosure of carbon emissions.

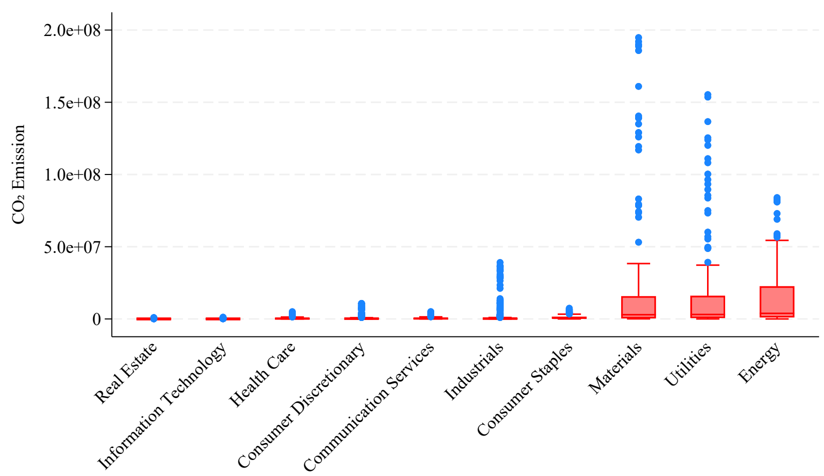

Figure 7: Box Plots: Distribution of CO_2 Emissions by Sector

Notes: Figure 7 presents box plots of carbon emissions by sector. Following the GICS by MSCI (2024), we categorize ten sectors into two groups: high-polluting and non-high-polluting, to examine whether the effect size differs in industries with a higher level of carbon emissions. The High-polluting group includes the major polluting sectors: utilities, materials, and energy sectors. We created a dummy variable for high-polluting industries, where the value is 1 if the company belongs to a high-polluting sector, and 0 otherwise.