Do Decarbonization Actions Work?

Causal and Cost-Effective Evidence from European Mining and Steel industries

Martin Kjærnes

Dissertation written under the supervision of Professor Miguel Nogueira

Dissertation submitted in partial fulfilment of requirements for the MSc in Business Analytics, at the Universidade Católica Portuguesa, 05.01.2026.

1. Abstract

Do Decarbonization Actions Work? Causal and Cost-Effective Evidence from European Mining and Steel industries

Author: Martin Kjærnes

Abstract (English)

This dissertation aims to quantify the effects of emission-reduction actions undertaken by European mining and steel companies on (i) operational CO_2 emissions and (ii) financial performance in the short term. These sectors are among the most emissions-intensive and face increasing pressure from regulators and investors to decarbonize while maintaining economic viability. Despite improvements in reported environmental indicators, it remains unclear whether higher scores translate into real emissions reductions or improved financial outcomes.

Using firm-level panel data, the study combines three complementary approaches. First, a fixed-effects difference-in-differences design estimates the causal impact of specific actions on Scope 1 and Scope 2 emissions by comparing treated firms to similar control firms and aligning outcomes around implementation timing. Second, estimated emissions impacts are combined with investment information to construct marginal abatement cost curves (MACCs), producing levelized costs per tonne of CO_2e abated based on CAPEX, assumed project lifetimes, and a discount rate. Third, a supervised machine-learning model based on ridge regression is applied to predict emissions changes and simulate recommendations for firms that did not implement actions during the sample period.

The results show no clear downward trend in aggregate industry emissions, but action-level effects indicate meaningful short-run reductions, especially for Scope 2. Cost-effectiveness varies substantially across actions and industries. The ridge model exhibits limited predictive power, highlighting challenges in generalizing short-term impacts. Overall, the findings suggest that credible progress in heavy-industry decarbonization depends on targeted interventions and causal evaluation rather than improvements in aggregate environmental scores alone.

Resumo (Português)

Esta dissertação analisa como ações corporativas de redução de emissões adotadas por empresas europeias dos setores de mineração e aço afetam (i) as emissões operacionais de CO_2 e (ii) o desempenho financeiro no curto prazo. Esses setores são altamente intensivos em emissões e enfrentam crescente pressão de reguladores e investidores para descarbonizar sem comprometer a viabilidade econômica. Apesar de melhorias reportadas em indicadores ambientais, permanece incerto se elas refletem reduções reais de emissões ou ganhos financeiros.

Com base em dados em painel ao nível da empresa, o estudo combina três abordagens. Primeiro, um modelo de diferenças-em-diferenças com efeitos fixos estima o impacto causal de ações específicas sobre as emissões de Escopo 1 e 2, comparando empresas tratadas com controles semelhantes e alinhando os resultados ao momento de implementação. Segundo, os efeitos estimados são integrados com dados de investimento para construir curvas de custo marginal de abatimento (MACC), calculando custos por tonelada de CO_2e evitada com base em CAPEX, vida útil e taxa de desconto. Terceiro, uma regressão Ridge (machine learning) é usada para prever variações nas emissões e simular recomendações para empresas que não adotaram ações no período.

Resultados não indicam uma queda consistente nas emissões agregadas, mas mostram reduções relevantes associadas a ações específicas, sobretudo no Escopo 2. O modelo preditivo apresenta poder explicativo limitado, restringindo a generalização de efeitos de curto prazo. Em conjunto, as evidências sugerem que a descarbonização na indústria pesada depende de intervenções direcionadas e avaliação causal, e não apenas de melhorias em scores ambientais.

Keywords: Predictive analytics, Decision analytics, Data-driven decision-making, Supervised learning, ESG, Business analytics, DiD, MACC, fixed-effects , cost-effective, environment

Table of contents

Equations, Figures and Tables content

2. Introduction

Over the past decade, the financial landscape has shifted, with Environmental, Social, and Governance (ESG) considerations becoming integral to investment decisions and corporate strategies. The term ESG first appeared in the United Nations Global Compact's report "Who Cares Wins" (2004), where leading financial institutions explained how incorporating sustainability could improve long-term financial markets. Although the terminology is relatively new, the foundational principles originate from the socially responsible investment and corporate social responsibility (CSR) movements that began in the 1960s and 1970s (Sparkes, 2002; Schueth, 2003).

Sustainability in a corporate context means operating in ways that preserve natural and social capital, ensuring that current economic activities do not threaten the prospects of future generations. However, the so-called green growth dilemma remains: how can companies stay competitive and profitable while reducing environmental damage? (OECD, 2020). The global push for decarbonization, driven by the Paris Agreement (2015), has accelerated this debate, but its impact and success remain unclear. Still, the core need to move toward a sustainable world persists. Empirical evidence indicates that green finance mechanisms, particularly green bonds and climate-related instruments, can effectively support emission reduction efforts (Gharleghi, Shafighi, & Nawaser, 2024). However, the overall balance between sustainability and profitability continues to be debated. Sachs et al. (2019) note that investment in fossil fuels still dominates global capital flows, while Andrade et al. (2021) find that the financial performance gap between green and traditional energy companies has narrowed, emphasizing the ongoing tension between financial goals and environmental concerns.

The classification of Scope 1, Scope 2, and Scope 3 emissions is based on the Greenhouse Gas Protocol (GHG Protocol), developed collaboratively by the World Resources Institute (WRI) and the World Business Council for Sustainable Development (WBCSD) and formally launched in 2001 following initiatives dating back to the late 1990. Within this framework, Scope 1 includes direct emission sources controlled by the companies or their operations. Scope 2 accounts for the indirect emissions from purchased electricity, steam, and cooling. Lastly, Scope 3 encompasses all other indirect emissions along the value chain, usually including emissions from transportation, use of sold products, and end-of-life treatment. Although it is not mandatory, the protocol has become the foundation of major systems and regulatory reporting standards. This makes it a natural reference point for empirical analysis of corporate decarbonization and climate-related investments.

2.1 Problem Statement

The balance between sustainability and financial performance is particularly critical for emission-intensive sectors such as the European mining and steel industries, which are the focus of the present thesis. The mining and steel industries account for over 10% of global CO_2 emissions and face increasing pressure from investors and regulators to decarbonize (IEA, 2023). Although companies in these sectors are reporting higher environmental scores, it remains unclear whether these improvements translate into measurable emission reductions or improved financial performance. Therefore, this study examines which emission-reduction actions mining and steel firms undertake, and how these actions affect realised Scope 1 and Scope 2 emissions as well as short-term financial outcomes. While sustainability efforts may be desirable, it is also reasonable to expect that such efforts must be financially viable to be maintained over time. Hence, the main focus of this thesis is to measure the actions European mining and steel companies have taken to lower their CO_2 emissions and to quantify their short-term effects, where "short term" is defined as effects observed within one to two years after implementation.

2.2 Structure of the Thesis

The remainder of the thesis is organized as follows. First, the literature review outlines the key theoretical perspectives and prior empirical findings that motivate the study and help frame the research questions. Second, the data and methods section describes the datasets, variable construction, and empirical strategy used to evaluate both environmental and financial outcomes. Third, the analysis section presents the main results and interprets them in relation to the hypotheses and the broader literature. Finally, the thesis concludes by summarizing the core findings, highlighting their practical implications, and discussing the main limitations.

3. Literature review

3.1 Core ideas in financial theory

Shareholder theory emphasizes that a company's primary goal is to maximize shareholder value. Managers act as agents for the owners, focusing exclusively on increasing financial performance as long as they operate within legal and ethical boundaries (Friedman, 1970). From this viewpoint, managers and companies should be evaluated based on the financial benefits they provide to shareholders. Non-financial goals are considered secondary and a distraction unless they enhance profitability. In this theory, ESG practices are only relevant if they contribute to shareholder value by reducing operational costs, increasing efficiency, or enhancing reputation in ways that ultimately lead to higher financial results.

On the other hand, we have stakeholder theory. This states that firms should create value for all stakeholders, not just shareholders. Stakeholders include employees, customers, suppliers, communities, regulators, and the environment. In other words, any group that can influence or be affected by the firm’s operations (Freeman, 1984). The main idea is that long-term business success depends on balancing stakeholders' interests. Sustainable growth requires trust, social legitimacy, and responsible resource use. From this perspective, ESG is not just a tool for profit but a key part of corporate purpose and long-term value creation, helping firms build resilience, maintain legitimacy, and achieve sustainable performance over time. For carbon-intensive sectors, this means that emission actions and strategies are crucial for risk management and continued access to resources and capital (Bolton & Kacperczyk, 2021; Ilhan, Sautner, & Vilkov, 2021).

Greenhouse gas emissions represent a clear negative externality from an economic perspective. For all companies, the shareholders theory exists, meaning that to keep them afloat, they need to be profitable in some way. Therefore, carbon taxes, emission trading, and performance standards align private incentives with social goals by pricing emissions or encouraging technological change. Consequently, in theory, companies should reduce emissions until the marginal abatement cost equals the relevant price. While marginal abatement cost curves (MACC) provide a statistical view of this tradeoff by ranking abatement actions from cheapest to most expensive, this method is used to analyze policy and corporate strategy to identify high-impact measures. Marginal abatement cost curves (MACCs) give a static depiction of this trade-off by ranking abatement options from cheapest to most expensive and are used in policy analysis and corporate strategy to identify “noregret" and high-cost measures (McKinsey & Company, 2009; Kesicki & Ekins, 2012). From an economic perspective, greenhouse gas emissions represent a classic negative externality: private production decisions do not internalize the full social cost of climate damage. Corrective policies, such as carbon taxes, emission trading schemes, and performance standards, are designed to align private incentives with social objectives by putting a price on carbon or mandating technological change. In theory, firms should then reduce emissions until their marginal abatement cost equals the relevant carbon price

3.2 Previous findings

The empirical literature frequently examines the relationship between corporate sustainability and financial performance. Meta-analyses of hundreds of studies generally indicate that strong ESG performance tends to lead to equal or better financial results, especially over the long term. Friede, Busch, and Bassen (2015), in a review of more than 2,000 studies, note that most find a positive link between ESG indicators and financial outcomes. More recent research has reached similar conclusions. Although the strength of these relationships varies across different methods, regions, and measurement choices (Atz, Holt, Koelbel, & Whelan, 2023), the overall evidence supports the idea that environmental performance can positively influence shareholder value.

However, recent research has shown that ratings and sustainability scores do not always correlate with actual emission reductions, especially in carbon-intensive sectors. Studies document weak and inconsistent links between firms' scores and real emissions. This indicates that much of the apparent improvement may come from better disclosure and communication rather than real changes in production processes (Berg, Kölbel, & Rigobon, 2022; Gibson, Krueger, & Schmidt, 2023). Research on greenwashing reveals that sustainability reports and ESG narratives might overstate environmental progress compared to actual emissions data. This calls for more robust, action-oriented, and emissions-based measures of performance. It reinforces the need to differentiate symbolic ESG commitments from genuine decarbonization.

In contrast, several sources in the literature on carbon risk and financial outcomes show that firms with higher carbon intensity tend to have higher expected returns and increased financial costs. Osterich and Tsiakas (2024) find that investors demand a risk premium for owning firms with high emissions under the EU Emission Trading Scheme. This indicates a negative relationship between intensity and risk-adjusted return. Furthermore, global equity and credit markets report similar findings, demonstrating that high-carbon sectors face higher financing costs and greater downside risks (Bolton & Kacperczyk, 2021; Ilhan et al., 2021). This suggests an economic rationale for decarbonization, particularly impacting industries such as mining and steel, where Scope 1 and 2 emissions primarily stem from materials.

When examining abatement costs and policymakers, the MACC technique is used to rank technologies and evaluate costs per ton of CO2. McKinsey's (2009) "Pathways to a Low-carbon Economy" study exemplifies a global and sectoral MACC analysis. Based on engineering models and scenario analyses, it aims to assign a cost to emission reductions. Similarly, Kesicki and Ekins (2012) utilize MACC in their climate research, demonstrating its usefulness for communication and its limitations. While standard MACC often overlooks uncertainty, dynamic effects, interactions, learning, and rebound measures, it generally depends on ex ante estimates rather than ex post realized performance. Less studied is firm-level ex post evidence on the costs and effectiveness of decarbonization projects. Most MACCs for industries come from "top-down" or "bottom-up" studies. A few papers use observed capital expenditure and emission data at the company level to analyze actual costs per ton and compare them to broader actions, investments, process changes, and emission controls. Additionally, although DiD methods are employed to analyze environmental regulations and policies, such as air-quality standards or carbon pricing, there is limited application of firm-level DiD to estimate the causal impact of targeted decarbonization projects on Scope 1 and 2 emissions (Callaway & Sant'Anna, 2021; Sun & Abraham, 2021).

3.1.2 Research GAP

Building on the prior literature, a gap remains in the specific context of mining and steel. This thesis contributes in two main ways: (i) it applies a firm- and time-fixed effects DiD framework to estimate the causal impact of decarbonization actions on emissions within these industries, and (ii) it combines the estimated abatement effects with project-level CAPEX to construct a data-driven MACC-style table that benchmarks cost-effectiveness across actions and sectors.

Firstly, most empirical work on composite sustainability relies on composite ESG scores or environmental scores as proxies for performance. While, as noted above, these scores often blend indicators of disclosure quality, management systems, and forward-looking commitments with actual emission outcomes, they may obscure whether improvements reflect decarbonization or just better reporting (Berg et al., 2022; Atz et al., 2023). This can be problematic in a carbon-intensive industry, where stakeholders' perceptions of absolute and relative emission reductions might matter more than policies.

While substantial research exists on the financial impacts of carbon risk and environmental performance, specific evidence for heavy industry remains limited. Most studies analyze broad market indices or diverse samples, with only a few focusing on mining and metallurgical firms. These industries face unique challenges, including technological constraints, long asset lifespans, and concentrated Scope 1 emissions from processes involving heat and chemical reactions. For these sectors, the key questions go beyond whether greener firms are more profitable; they also involve identifying which specific actions, like energy-supply projects, process-energy upgrades, material-flow improvements, or emission-control technologies, lead to significant and lasting emission reductions, and at what cost per tonne. Despite this, there is a lack of work linking actual emission reductions to explicit abatement cost curves (Oestreich & Tsiakas, 2015; Bolton & Kacperczyk, 2021). Typically, marginal abatement cost curves (MACCs) are based on engineering assessments or estimated potentials, rarely incorporating real-world firm investment data, observed emission trajectories, or the evolution of similar firms that did not invest. Conversely, most difference-in-differences (DiD) studies on environmental policy measure average emission effects but do not combine these with capital expenditure data to evaluate cost-effectiveness across different technologies, actions, and industries (Callaway & Sant'Anna, 2021; de Chaisemartin & D'Haultfœuille, 2020; Sun & Abraham, 2021).

3.3 Hypotheses development

The empirical approach in this thesis centers on two key aspects: whether specific decarbonization measures reduce operational emissions in the short run, and how the costs of these reductions vary across industries and measures. First, environmental effectiveness is evaluated using a fixed-effects difference-in-differences (FE-DiD) method, comparing firms that adopt identified decarbonization techniques with similar firms that do not adopt these techniques or adopt them later. Incorporating firm and time fixed effects, along with controls for common shocks such as energy prices and grid carbon intensity, allows the treatment effect on log Scope 1 and Scope 2 emissions to be interpreted as a causal impact of the measures.

Second, cost-effectiveness is assessed by combining these estimated annual emission reductions with project-level CAPEX and lifetime assumptions to generate marginal abatement cost curves (MACCs). The levelized cost of abatement, measured in EUR per tons of avoided CO_2e, is calculated for each industry and action type, enabling comparisons of where significant quantities of low-cost abatement are achievable and where capital costs per tonne are much higher (McKinsey & Company, 2009; Kesicki & Ekins, 2012; Bolton & Kacperczyk, 2021). Lastly, the thesis uses a supervised machine-learning model that relies on the same underlying data to predict emission outcomes and estimate abatement costs across actions and industries. This provides a benchmark for comparing the FE-DiD and MACC rankings against a flexible, data-driven model, and helps verify whether the main patterns in environmental impact and cost-effectiveness remain consistent when employing a more advanced, non-parametric approach.

4. Datasets and methodology

4.1 Emission and Financial Performance data

The data was collected from the Refinitiv platform, which is widely used in relevant studies. The platform system aggregated hundreds of indicators into pillars and sub-pillars to enable scoring, and it also includes adjustments for industry relevance and size biases. However, a major methodological challenge in all ESG research is the heterogeneity of measurements across different providers over time. As noted previously, ESG reporting has increased recently, so methodologies can vary across sources. Accordingly, by relying on a single source (in my case, Refinitiv) and being transparent about the transformation process, I aim to reduce measurement noise. Nonetheless, the methodological features must be acknowledged when interpreting the results.

The structure of my company data has a panel data format, including all the companies within my scope. The selection of the companies is based on Refinitiv sorting, and all the companies are from Europe as their continent of risk. This dataset provides the foundation for both the dependent and independent variables used in the analysis. For each firm-year observation, the dataset contains

- Scope 1 and 2 (TCO_2) and log transformed for stability
- Environmental and emission scores
- Core company variables such as assets, revenue, capital expenses and EBIT

- Industry classifications
- Firm ID
- Controls

This will serve as the basis for constructing the pre- and post-treatment baseline used for the analysis. It also provides historical values to calculate firm-level emissions by industry and trends, as shown in the descriptive section.

4.2 Emission reduction - action data

The action dataset was constructed through a structured manual collection process. For each firm in the sample, I screened publicly available sources primarily annual reports and sustainability/ESG reports, supplemented with regulatory filings, company press releases, and relevant industry databases to identify concrete, operational decarbonization measures explicitly linked to Scope 1 or Scope 2 emissions. Each action was only recorded when the source provided sufficient detail to (i) classify the measure into a predefined action category and (ii) assign an implementation year (or a clearly stated start/commissioning year). When multiple sources referred to the same project, the observation was consolidated and cross-checked to avoid duplication. The resulting dataset is therefore an event-style register of implemented actions, aligned to the firm-year panel to define treatment timing and to link actions to subsequent emission and financial outcomes. Each observation includes the following variables:

- Firm id, company name, industry, firm identifiers aligned with the Company data panel.
- Action category (energy supply, emission control, process optimization, process energy, material flow, primary metallurgy, other).
- Scope indicator
- The first year the firm implemented the action, defining the treatment timing.
- Year, the panel year aligned with outcome variables.
- Treatment status: *TREATED* for firms that implemented an action, *CONTROL* for firms with no action records.
- Capital cost of the project (Capex)
- City of implementation
- Size of project (Small, medium, big)

4.2.1 CAPEX

Regarding the marginal abandonment cost curve, it is better to consider either capital expenditure or operational expenditure. Since operational expenses were unavailable due to confidentiality, the MACC estimation was based on capital expenditure. For actions without associated capital expenditure, we initially assumed they were small, with capital costs less than $10 million USD. Therefore, by using our nearby neighbor dataset, observed or peers, to estimated all values.

$$CAPEX_j^{est} = \begin{cases} CAPEX_j & = \ Observed \\ \widehat{CAPEX}_{J^{NN}} & = \ If\ NN\ exist \\ \widehat{CAPEX}\ k, a, s^{peer} & = Last\ option \end{cases}$$

The emission reduction action data is backbone to the entire empirical approach. In the DiD analysis, the action dataset defines who is treated and when by identifying the first year each firm implements a recorded decarbonization action, which makes it possible to align Scope 1/2 emissions and financial outcomes in event time (pre vs. post) and compare treated firms to firms with no recorded actions. In the MACC construction, the same action records are then used to assign observed emission changes to specific action categories and connect them to project CAPEX (observed or estimated) and lifetime assumptions in order to compute levelized abatement costs (EUR per tCO_2e). Finally, in the machine-learning step, the action dataset provides the observed action labels and the corresponding outcome changes that the ridge model learns from, so the model can be trained on treated observations and then applied to never-treated firms to simulate expected emission effects under recommended actions

4.3 Energy data

The energy dataset is constructed based on country-by-year macro variables directly linked to their energy systems, following a panel structure from year to year. The purpose of including this is to account for exogenous shocks that may impact emission levels from firm actions. We can argue that most companies in Europe face similar risks, making their situations comparable. However, due to differences among some countries and to isolate the effect of the actions, these will be considered. Falling grid emission factors will decrease scope 2 emissions even without firm-level reductions, and higher electricity prices make energy efficiency measures more financially appealing.

- Electricity prices
- Natural gas prices
- Grid emission intensity (gCO_2/kWh)

4.4 Near neighbor data

A nearest-neighbor (NN) matching procedure was constructed to support both (i) imputing missing CAPEX values and (ii) selecting comparable control firms for the DiD design. In practice, for each firm I identified its closest "peer" firms using pre-treatment characteristics from the company panel. The matching variables were chosen to capture scale and operational similarity, including firm size proxies (e.g. revenue and total assets), baseline Scope 1 and Scope 2 emissions (levels or logs), and investment/operating characteristics such as historical CAPEX and profitability measures where available. Because these variables are measured in different units (tCO_2e, EUR/MEUR, and scores), they were standardized into z-scores before computing distances. Similarity was then defined by the sum of squared differences across the standardized characteristics, and the firm's nearest neighbor(s) were those with the smallest distance. This NN structure was used in two ways: first, to estimate missing CAPEX by borrowing information from the most similar observed peers; and second, to ensure that DiD comparisons rely on control firms with comparable baseline profiles, reducing imbalance between treated and untreated units prior to action implementation.

4.5 Data cleaning

The cleaning process was designed to prepare for the three main analysis sections, ensuring consistency and clarity across the datasets. Since the empirical framework combines independently sourced datasets, the initial step involved standardizing firm identifiers and industry classifications across all datasets. Each company was given a unique firm ID and grouped into their respective sectors. Actions were organized into logical categories to ready them for the analysis stages.

All numerical variables were reviewed for invalid values, outliers, and missingness. Some observations with non-plausible values were corrected or removed entirely. Based on the nature of the thesis, the dataset was not large from the start. Therefore, to retain most of the data and avoid removing entire rows due to missing values, mean and median imputation were implemented. This allowed the dataset to maintain its desired size and preserved most data points. It was also used to standardize the z-score of key characteristics for emission and environmental scores. Outcome variables used directly in DiD estimates were not imputed.

Finally, I combined the cleaned action-level data with firm outcomes to create a consolidated panel (firm act delta). I calculated pre-treatment baselines and post-treatment averages within event windows, removing rows with incomplete baselines, non-finite changes, or unresolved missing data to ensure both the DiD and machine learning models were trained on a consistent dataset.

4.6 Variable description

4.6.1 Geographical distribution

The figure shows the geographic distribution of the sampled firms across Europe. The dataset mainly covers Western and Northern Europe, with prominent clusters in the United Kingdom, Russia, Germany, Sweden, Finland, and France. Southern and Eastern Europe are underrepresented, although some firms are from Spain, Poland, and Austria. This uneven distribution indicates where publicly listed mining and metals firms are based and where disclosure levels are highest.

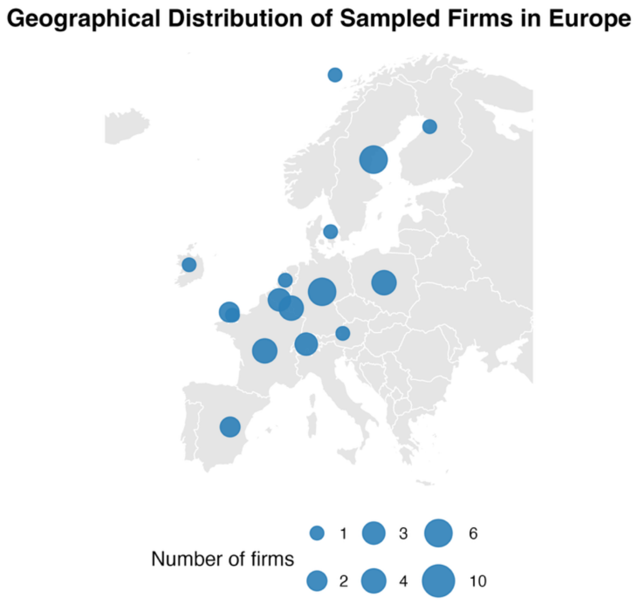


Figure 1 - Number of firms in sample.

4.6.2 Choice of dependent variable

For the analysis, the dependent variable will primarily focus on Scope 1 and 2, but it will also include revenue, environmental scores, and emission scores from Refinitiv. The decision not to include the scores in the final analysis is based on prior research on these scores. They do not always reflect the true relationships. The score is based on the company's impact and management of environmental risks and opportunities. The overall breakdown of this score is divided into three main areas: resource use, emissions, and innovation. (Refinitiv, 2022). As we can also see in the correlation spread, there is a clear positive correlation where the higher the Scope 1/2 emissions reported by the firm, the higher their environmental score tends to be. Given this and the uncertainty about how the scores are measured, the main analysis relies on Scope 1 and 2 emissions rather than the provided scores.

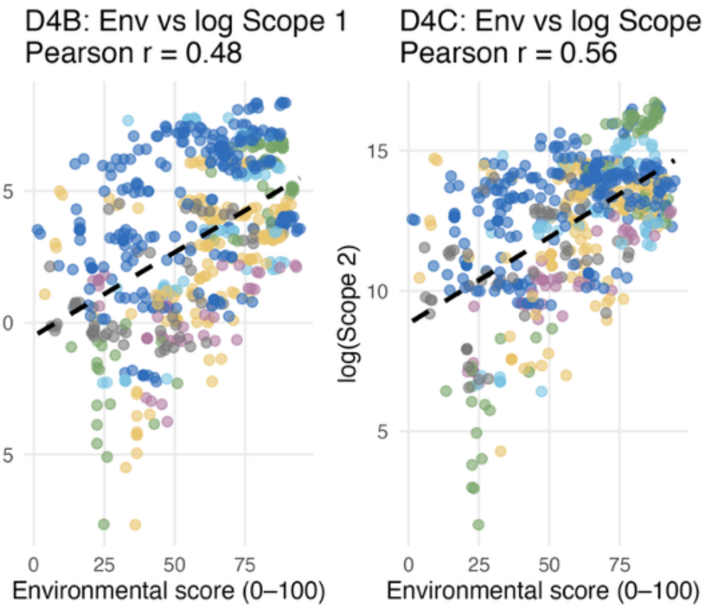


Figure 2 - Correlation between Scope and Environmental variable

As noted, this study will use Scope 1 and 2 emissions data following the GHG Protocol. The variables are reported annually at the company level and measured in metric tonns of CO2 equivalent (TCO2E), forming a panel of firms over multiple years. Some companies lack complete emission data; therefore, Refinitiv may supplement this with model-based estimates using statistical and machine learning techniques to leverage data from similar firms. For this analysis, raw emission levels are transformed using a log(1+scope) to stabilize the variables and reduce the impact of extreme values. These variables will serve as the basis for the fixed effects difference-in-difference (FE-DiD) model. Finally, the average pre-treatment emission for the firms (baseline_s1 and baseline_s2) are combined with the estimated treatment effect in logs and the FE-DiD coefficients for TCO2E, forming the foundation for abatements and marginal abatement costs used later in the MACC analysis.

4.6.3 Excluding scope 3

The analysis excludes Scope 3 emissions for three main reasons. First, the academic literature consistently highlights significant measurement uncertainty in Scope 3 reporting since these emissions depend on activities outside the firm's direct operational boundaries and largely rely on supplier- and customer-level data that are often incomplete or model-generated. Second, the action dataset used in this thesis has few, if any, interventions that explicitly target Scope 3 categories, making it methodologically challenging to link observable firm actions to changes in upstream or downstream emissions. Third, including Scope 3 could introduce systematic noise rather than useful information into both the DiD and machine learning components, since variation in reported Scope 3 levels mainly arises from differences in reporting practices rather than actual emission changes. For these reasons, the analysis focuses on Scope 1 and Scope 2, where measurement standards, action relevance, and identification strategies are much more reliable.

4.6.4 Variable distribution

As we see on the histograms, both variables exhibit a right-skewed distribution in levels. Based on this, it's justified to use the log(1+scope) transformation, which is applied throughout the analysis. After transformation, the distributions are still asymmetric but more stable compared to the previous levels. This is done to reduce influence and improve comparability across firms of different sizes. We see that the median shares the largest cluster of firms with the mode of emission levels, while a small subset of observations appears in the upper and lower tails. The scope 2 emissions have a more concentrated mass around the median, whereas scope 1 exhibits a wider spread. This indicates greater heterogeneity in firms' direct operational emissions and aligns with the intuitive notion that scope 2 tends to be more uniform across firms, since they are based in the same geographical region. So, based on the pattern, it underscores the necessity of the transformation for use in the models built later in the analysis.

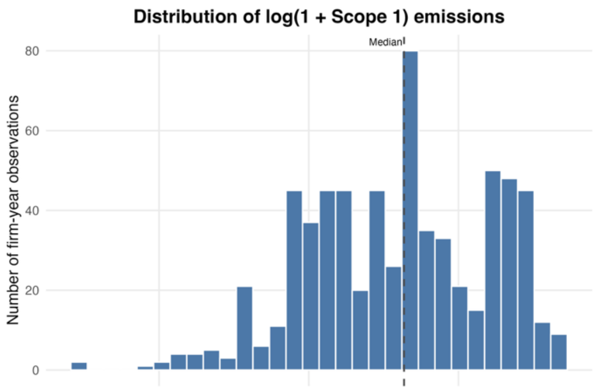


Figure 3 - Histogram of Log scope 1 emission

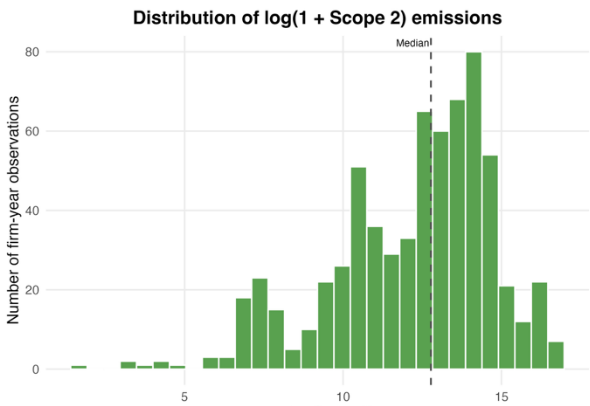


Figure 4 - Histogram of log scope2 emission.

4.6.5 Variable correlation

What we can infer from the correlation table is that firm size proxies (revenue, assets, capex) are strongly correlated with their emission scores and values. This is consistent with the industrial context of the sample. The larger the companies, the more CO2 emissions they release. We can also see that the overall environmental score and emission score are almost perfectly correlated, indicating that they are based on overlapping indicators. As we also commented before, we observe that scope variables and environmental scores are positively correlated, which supports the reason for excluding scores and focusing on realized emissions for the analysis. Due to the strong correlation with firm size proxies, which indicates overlapping information and potential multicollinearity concerns, we have partly addressed this risk in the DiD specification by including firm fixed effects, and in the machine learning components by using ridge regression. This does not mean its entirely safe, but ridge regression should penalize the magnitude of coefficients and stabilize estimates in the presence of correlated predictors.

Overall, the correlation structure consists with economic intuition, while also reinforcing the need for fixed effects and regulation for our empirical strategy.

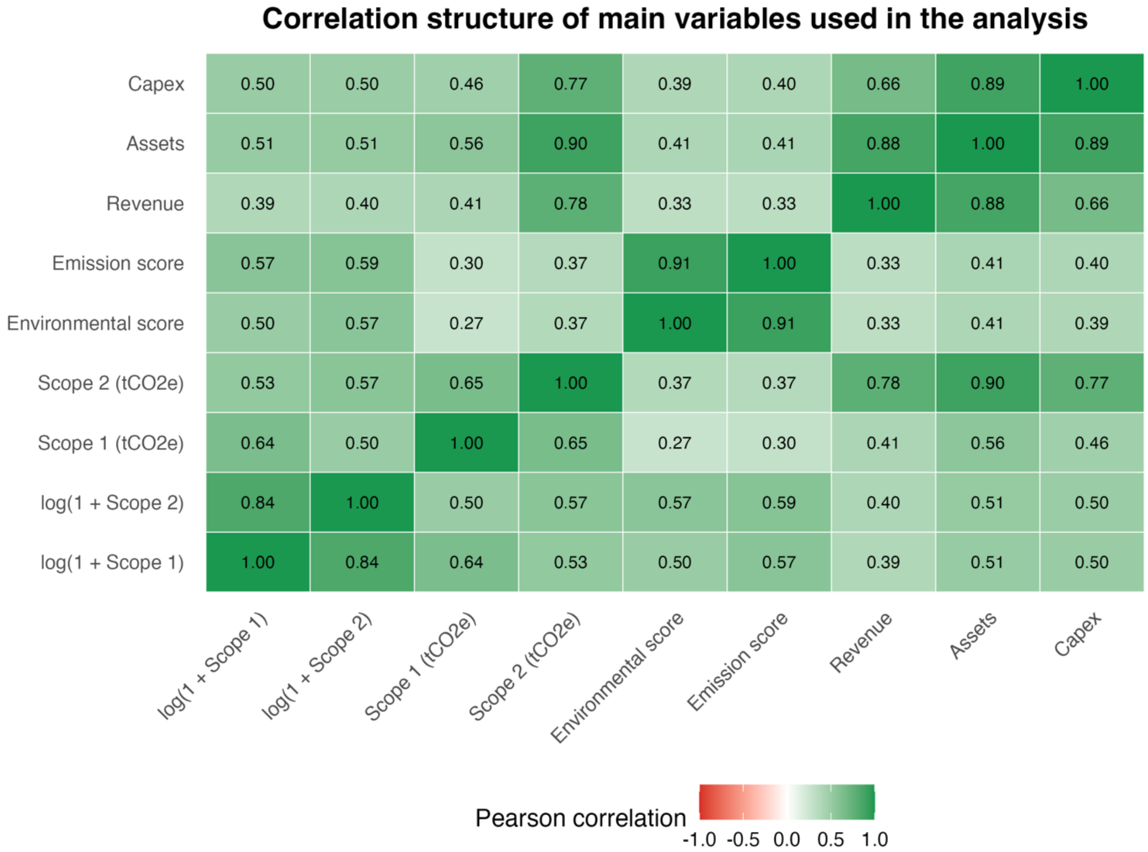


Figure 5 - Correlation of main variables from company data

4.4 Methodology

4.4.1 Differences in differences (DID)

The empirical section uses a DiD framework with firm and time fixed effects, a standard method in modern econometrics. In this setup, treatment effects are identified by comparing changes in outcome variables between treated firms and similar control firms, before and after the intervention. Including a firm fixed effect accounts for permanent differences among firms. Based on my sample, firm size varies significantly, so adding this variable helps isolate the true effect of the actions. Similarly, a time fixed effect is included to control for macroeconomic shocks. Specifically, energy prices may influence the model. Since the sample's demographics suggest that overall energy prices are similar, but some countries have their own prices, this variable is included to prevent distortions in the estimated effects. This approach is based on the “two-way fixed effects” model discussed in Wooldridge (2012) and aligns with recent literature on treatment effects heterogeneity. In climate and environmental research, similar FE-DiD models are used to estimate causal effects of regulation or technology adoption on emissions and performance (Dechezleprêtre & Sato, 2017; Fowlie, Holland & Mansur, 2012). Because this model employs a log transformation, it produces elasticities expressed as percentage changes and is later converted into TCO2E using pre-treatment emission levels, a common practice linking treatment effects to engineering quantities.

4.4.2 Marginal abetment cost curve (MACC)

Based on the DiD emission levels, then develop a decision tool by constructing marginal abatement cost curves at both the industry and action levels. MACC is a common framework in energy and climate policy analysis, summarizing the annual emission reductions and their levelized costs per ton of CO2E. The approach combines CAPEX, lifetime, and a discount rate to calculate annualized costs. This follows practices in MACC literature, such as McKinsey & Company's global abatement cost studies (Enkvist, Nauclér & Rosander, 2007), and academic work like Kesicki & Strachan (2011). While Kesicki (2012) also uses an abatement-weighted average across sectors, which I draw inspiration from. Therefore, the FE-DiD and MACC steps link causal estimation to transparent cost-effectiveness and support a natural ranking of decarbonization options.

4.4.3 ML approach

Combined with the previous methods used, then will finish this with an ML decision model. The purposes of the components are twofold: (i) to build a data-driven model that predicts the most suitable action based on its pre-intervention characteristics, and (ii) to use the predicted actions as input to simulate a model estimating the expected reduction in emissions. This structure follows the broader methodology in the modern ML econometrics literature, where the prediction is used to identify causality (Mullainathan & Spiess, 2017; Athey & Imbens, 2019).

Combining the results from previous analyses (MACC and DiD) to try to estimate the expected change in outcomes for Scope 1 and Scope 2. This approach, starting with the first prediction followed by causal mapping, aligns with the modern "causal ML" framework (Chernozhukov et al., 2018), where machine learning handles treatment assignment and causal parameters are estimated through the model. Related applications can be found in ML for forecasting environmental or energy-related outcomes (Liang et al., 2021; Knittel & Stolper, 2022).

The machine learning component is implemented as a regularized linear model using ridge regression. The ridge regression is especially suitable based on the nature of the dataset. First, the sample of treated events is small relative to the number of predictors. Additionally, several predictors are highly correlated. Given this situation, ordinary least squares would most likely produce unstable and highly variable coefficient estimates. When applying ridge regression, addressing this by adding an L2 penalty to the coefficients, shrinking them toward zero, and attempting to reduce the variance, but at the cost of introducing bias. Second, ridge remains linear in parameters and therefore easy to interpret. Predicted changes in emission can be read in the same way as the DiD model. After training on the treated, the estimated ridge model is applied out of sample to the never-treated firms. Conditioning on them adopting the MACC recommendation for their industry, the treated units serve as the training sample, while the never-treated are used for prediction. The regular, linear structure allows for interpretation for strategic purposes.

By this, the outcome variables are the realized changes in log Scope 1 and 2. The predictors are firm-level baseline characteristics, together with industry and action dummies. The model is trained only on treated firm action-based observations, which means only firms that actually have implemented actions in the sample periods. The reason for this is that the algorithm learns from observed treatments and where treatments have taken place, and to stay consistent with the causal structure in the DiD. Within the treated subsample, the ridge penalty parameter is selected by K-fold cross-validation, which creates internal train and validation splits to choose the degree of shrinkage that minimizes the prediction error

Taken together, I first try to recommend the actions based on the firm's characteristics, while then use this information linked to the implementation of action. This creates a link between the prediction, causal estimation, and cost-effectiveness evaluation.

5. Limitations

The analysis is limited to a small number of treatment events. This can be due to the way I personally collected data, but also because not all of these actions are published online. The DiD framework relies on forty firm-action observations from which causal effects are estimated. While the event study specification and parallel trends diagnostics might mitigate some concerns, the limited sample size will reduce the statistical power of the estimates and increase sensitivity to outliers or specific shocks that I have not taken into account. This will also affect the number of heterogeneous effects that can be estimated across industries and action categories.

The accuracy of measuring abatement effects depends on the quality of firm-reported emission data. Although this limitation applies to all work using Scope 1 and 2 data, these variables can differ in accuracy, reporting standards, and methodological consistency across different firms and industries. I have attempted to address this by transforming the variables to stabilize their variance and reduce the impact of level differences. Nonetheless, one must acknowledge that measurement errors may still occur, which could weaken both causal interpretation and predictive power.

The action dataset is limited by how complete the data set is, which then affects the MACC table. Not all industries are equally represented, and some actions might be disproportionately focused on a smaller group of firms. Just because some are more represented doesn't mean it is necessary or wrong. Certain parts of the sector may have contributed more to reducing emissions, but this could also be due to different reporting and publishing methods. Therefore, identifying the best actions at the industry level reflects the sample I collected, rather than a full list of abatement options. Additionally, some cost estimates for the actions are rough approximations, which will have a margin of error. While this was the best method available, the way data was collected could potentially introduce bias into the action ranking process.

The predictions for firms that have never been treated are not random guesses but scenario forecasts. The analysis considers the typical treatment patterns among those who haven't taken action. It assumes that the relationship between baseline firm characteristics and emission changes remains consistent across both treated and untreated groups. If firms that have never been treated differ systematically, the model could either overestimate or underestimate their emission changes.

Finally, all three methods used in this paper, DiD, MACC, and the ML model, rely on a common event time structure and variable construction. This improves comparability and allows us to build on each other's empirical framework, but it also increases the risk of mismeasurement in timing, baseline windows, or covariate definitions, which can propagate through the analysis. Overall, these limitations do not invalidate the findings but serve as warnings when interpreting the results. The outcome is based on a structured, data-driven approach designed to evaluate the effectiveness of abatement actions and provide firm-specific predictive insights. However, these are influenced by the scope and quality of the available data, as well as the modeling decisions made.

6. Results

The main goal of this analysis is to link different actions with both the financial and emission scores for the sector. To do this, I will examine the marginal abatement cost curve (MACC), which is based on a DID framework. Additionally, I will use a machine learning model to offer a data-driven approach for predicting the potential outcomes of these actions. In this section, we will go through the process step by step, highlighting how I develop the DiD and MACC tables, how I build the different models, and explaining the various options that could have been chosen, along with the reasoning behind our decisions. By the end of this, I will aim to answer three main questions:

Q1: How do the actions companies take lower their environmental scores?
Q2: In which industries have these actions been most effective in improving both financial and environmental performance?
Q3: What is the variation in the actions, and how can we predict the future trends?

6.1 Descriptive statistics

Before turning to the estimated effects of the emission-reduction actions, this section first presents additional descriptive statistics. The purpose is twofold: (i) to provide a clearer empirical backdrop for interpreting the subsequent results, and (ii) to place the action-level findings in a broader industry context, highlighting the main patterns and overall dynamics in emissions and performance over the sample period.

6.1.1 Actions over time

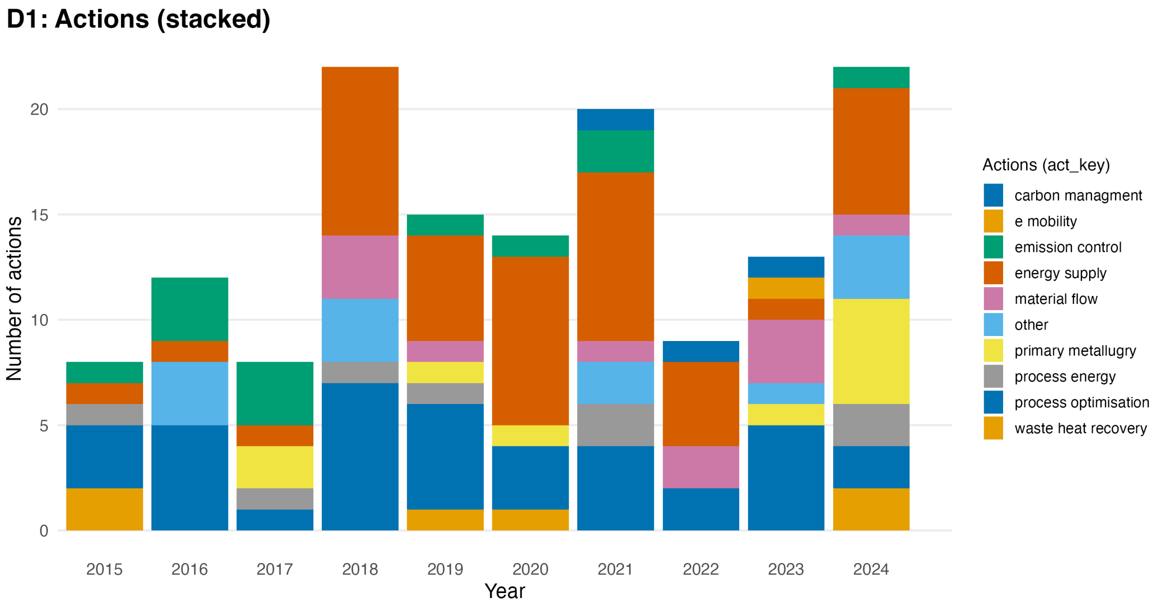


Figure 6 - Actions in period 2015 - 2024 for sample

Actions peak in 2018, 2021, and 2024, separated by quieter years like 2017 and 2022. The most consistent driver of volumes is energy supply, while carbon management remains a stable baseline across most years. Notably, primary metallurgy expands sharply in 2024, possibly indicating sector-specific efforts or policy-driven investments this year, while material flow and emission control show more episodic contributions. Since this count reflects registered actions, high bars indicate implementation rather than effectiveness. Building on this, I will further analyze the actions in the DID to determine if adoption leads to changes in both emission and financial outcomes.

6.1.2 Trends and emissions

The figure shows the median Scope 1 and Scope 2 emissions (measured in tCO_2e) for each industry over the years, serving as a baseline for our analysis. Two key points stand out. First, emissions levels vary across sectors: Diversified Mining and Iron & Steel consistently have the highest Scope 1 medians, Aluminum is in the middle, while Specialty Mining & Metals and Mining Support Services & Equipment operate at much smaller scales, highlighting the different opportunities and challenges for emission reductions in each industry. Second, the trends differ between Scope 1 and Scope 2 emissions. Scope 1 emissions declined gradually from around 2019/2020 to 2024, particularly in heavy industries such as Iron & Steel and Diversified Mining. In contrast, Scope 2 emissions are more variable, often spiking around

2020 - 2021 before stabilizing after 2022, reflecting the changing dynamics of energy markets and electricity consumption. Notably, Gold shows a renewed increase in Scope 2 emissions after 2022. The sharp drop in Mining Support Services emissions in 2023-2024 likely results from limited data or structural changes, though medians decrease, they still depend on the sample composition. I include this figure to illustrate long-term trends and differences across industries before attributing any new changes to specific actions. It helps identify where emission reductions are most possible, supports tests for differences between industries, and informs our choice to include industry fixed effects (or industry-specific trends if necessary) in our upcoming event-study and difference-in-differences analyses.

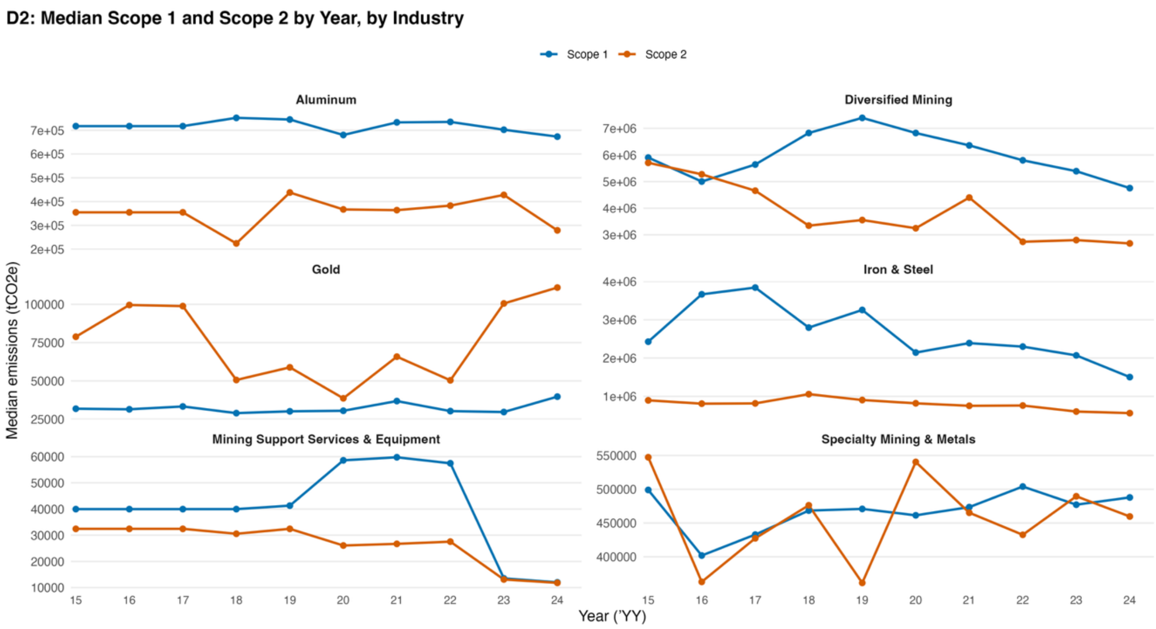


Figure 7 - Change in Scope 1 + 2 by year in sectors

6.1.3 Macros

To address our research questions, I look at country–year electricity and gas prices, as well as the grid emission factor (gCO_2/kWh). The main idea of the thesis depends on (i) the choice to take actions and (ii) measured emissions, especially scope 2. As energy prices rise, the private benefits of energy efficiency, fuel switching, and waste heat recovery become more appealing, affecting where and when firms adopt measures. At the same time, decreases in the grid emission factor reduce location-based scope 2 emissions even without firm actions. Ignoring these factors would skew the estimated effect. From the graph, we see macro shocks (like the 2021 - 2022 price spike) and differences across countries, especially in grid intensity. Including these factors helps separate system-wide influences from firm-level reductions and improves understanding of the actions firms are taking.

In the DiD design, I ensured the estimated treatment effect was not confounded by the country-level energy shock, which affects all firms simultaneously. Therefore, I included energy context controls directly into my DiD framework. For my baseline, I account for all the country-by-year shocks with fixed effects. By doing this, I ensure that the identification comes from differences within countries across firms.

$$\ln(E_{igt}) = \alpha_i + \lambda_t + \theta_g(Post_t * Treated * i) + u_{igt}$$

Equation 1 - General TWFE for the sample

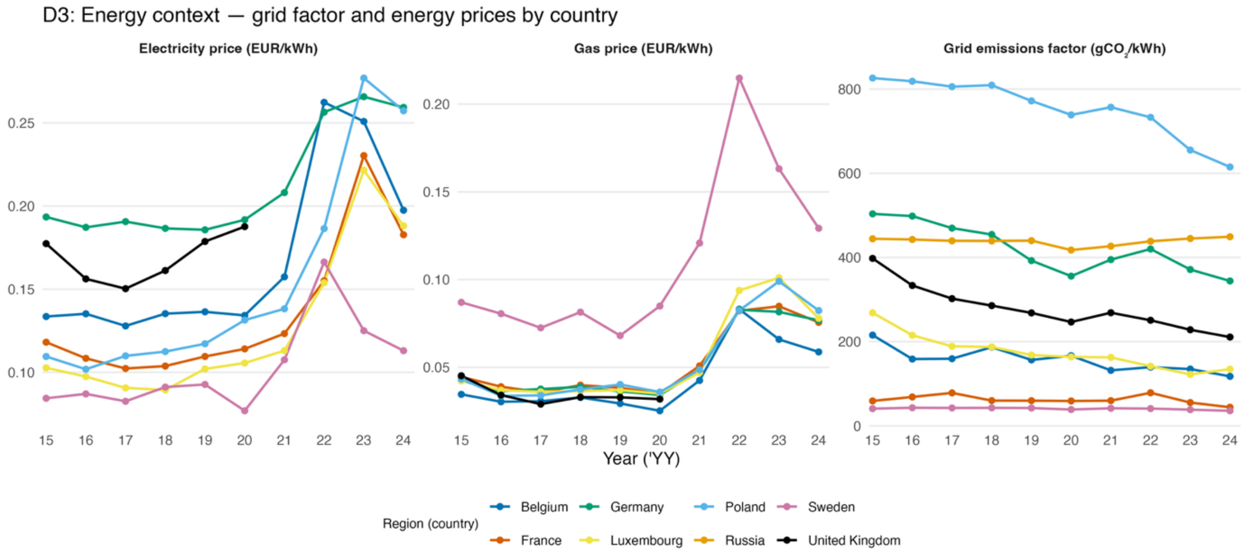


Figure 8 - Electricity, Gas and Grid emissions factors for sample

6.2 Difference in Difference

Based on the explanation I discussed, we now create a firm-level panel dataset for each action, including its implementation year. Using this, reindexing time relative to the action with K = t -T$_i$ for the firms and define a pre-window K = (-3, -2, -1) and a post-window K = (0, 1, 2). For each window, firms that have already implemented actions are considered treated, while firms that have not yet implemented actions are labeled as 'not yet implemented.' Firms that have never implemented any actions are classified as those that have not implemented any actions at all in the year being examined. This approach is based on the classic local DiD construction, which compares similar periods near the intervention period. Then examine five focus variables: (i) log-scope 1, (ii) log-scope 2, (iii) environmental score, (iv) emission score, and (v) revenue. For the aggregated data across industries, I also consider four different actions. The rationale is that these are small data batches for waste heat recovery, carbon management, e-mobility, and primary metallurgy; therefore, they are grouped under 'other,' resulting in a total of 5 actions across all 4. “Non-Gold precious metal & minerals” are also excluded.

Based on the steps outlined in the previous chapter, the DiD with FE will be explained in more formal terms. Y_{iat} be the outcome for firm i in industry a at time t, where Y is one of the focus variables explained above. For each action g and each role, $r \in \{treated, not\ treated, yet\ to\ be\ treated\}$, I compute the specific means $Y_{g,r,p}$ by averaging Y_{iat} over all firms for group g and over all periods in the pre- and post-windows defined by the event time K. The within-group change is defined by $\Delta\bar{Y}_{g,r} = \bar{Y}_{g,r}post - \bar{Y}_{g,r}pre$, while the DID gives the local average treatment effect on the treated contrast $ATT_g(Y) = \Delta\bar{Y}_g treated - \Delta\,\bar{Y}_g\,not\ treated$. For scope 1 and 2 outcome variables, as mentioned earlier, I have chosen the logarithmic expression $ATT_g(lnY)$; this will be interpreted as a proportional change of a percentage effect using $exp\{ATT_g(lnY)\} - 1$. These percentage effects back to real quantities, as we approximate the level effects in tons of CO2 by multiplying the proportional change with the treated group in the pre-window, $\Delta Y_g^{level} \approx ATTg(lnY)\bar{Y}g, treated, pre$.

To obtain interpretable summaries for industry and action levels, I computed weighted averages of the cell-specific ATT estimates. By aggregating at each industry and action, I derive a mean effect where the underlying value is weighted by the number of treated firms it includes. This means that industries and actions with more treated firms contribute more to the overall effect. While this can diminish the influence of smaller observations, the weighting reduces the risk of small values distorting the conclusion. In addition to a robustness check of the local DID calculation, I also estimated the panel FE model from.

$$Y_{it} = \alpha_i + \tau_t + \beta D_{it} + X_{it}\gamma + \epsilon_{it}$$

Equation 2 - Final TWFE within groups

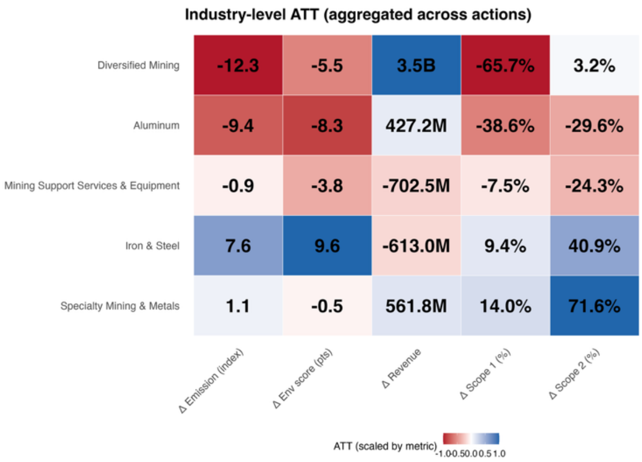


Figure 10 - Estimated DiD effects (aggregated on actions)

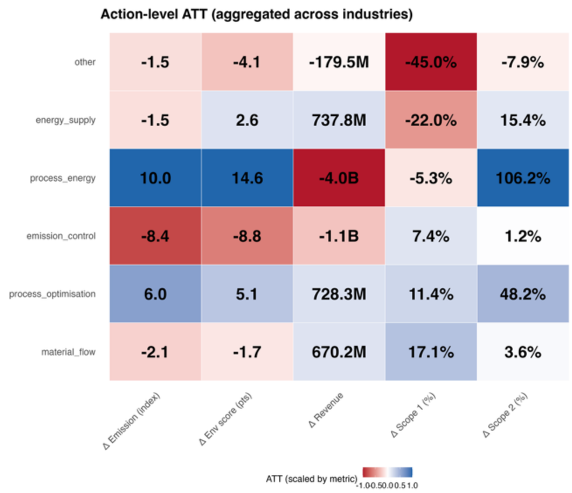


Figure 9 - Estimated DiD effect (aggregated on industry)

6.3 Marginal Abetment Cost Curve - Table

To compare actions across different industries, the DiD method was used to determine if these actions result in lower emissions. Investors and regulators also care about efficiency and the cost per TCO_2. The DiD framework provides a natural starting point because it offers a causal estimate of how key variables change when an action takes place. It will build on the framework we already discussed, further explaining how to use this to estimate the total abatement linked to capital expenditure for the aggregated actions and industry. First, the table will show how much emissions are reduced in $MTCO_2$ and then quantify the cost per CO_2 unit. This allows for ranking actions and identifying which ones are cost-effective based on their impact.

For our MACC graphs, I also included two parameters: lifespan and discount rate. These are important to incorporate into the equation to make the cost per ton of CO_2 as realistic as possible. For our discount rate, I set r = 5% to annualize capital expenditures across all projects and years. The rate is based on a generic corporate cost of capital rather than a short-term policy rate set by the ECB. Using a time-varying rate could make the cost per ton more sensitive to the timing of each action, which would also reduce comparability across industries and actions. This approach allows us to focus on cost-effectiveness rather than interest rate cycles. Another factor that's important to consider is the lifetime of the project. A high capital expenditure can be relatively inexpensive when divided over the project's lifespan. Finding the appropriate lifetime for each action was challenging; therefore, I have adopted a more general approach supported by literature, with the following lifespan for each action:

Action	Small	Medium	Big
Energy supply	15	20	25
Emission control	10	15	20
Process optimization	10	15	20
Process energy	10	12	15
Material Flow	8	10	12
Other	10	15	20

Table 1 - Estimated lifetime of actions

6.3.1 MACC implementation

Based on the FE-DiD model for log emissions, the estimated effect of treatment for g is interpreted as an approximate proportional change in emissions of the treated group, based on each firm-year observation through $\Delta E_{igt} \approx -\theta_g * E_{igt}^{Baseline}$. Where $E_{igt}^{Baseline}$ is the pre-treatment emission, where a negative $-\theta_g$ translate into a positive emission flow, and has been computed separately for both scope 1 and 2.

Each firm then belongs to an industry k, where each action is classified into an action type (a). For the AMCC level, I therefore combine industry–action cells (k, a). The total annual abatement cell is derived by summing the firm-level changes over all years, where firms belonging to industry k implemented action type a

$$Abatment_{k,a} = \Sigma i \in (k,a)\Sigma_t \Delta E_{igt}$$

Equation 3 - Cell-level total abatement (industry × action)

So, the core MACC metric is levelized based on the cost per ton of avoided emissions for industry and actions.

$$CostPerTon_{k,a} = \frac{C_{k,a}}{A_{k,a}}$$

Equation 4 - Cost-effectiveness ratio

These correspond to the “Cost S1/2 (EUR/T)” values reported in the table and plotted on the y-axis in the chart. Intuitively, CostPerTon k,a shows how much has been spent per year for each ton of CO2e avoided per year. For the industry and action levels, I computed the abatement weighted averages across cells.

$$\overline{AC}^{(S1)} = \frac{\Sigma_a CostPerTon^{(S1)}k,a * A^{(S1)}k,a}{\Sigma_a A^{(S1)}k,a}$$

Equation 5 - Mean cost per tons of CO_2e avoided (Scope 1)

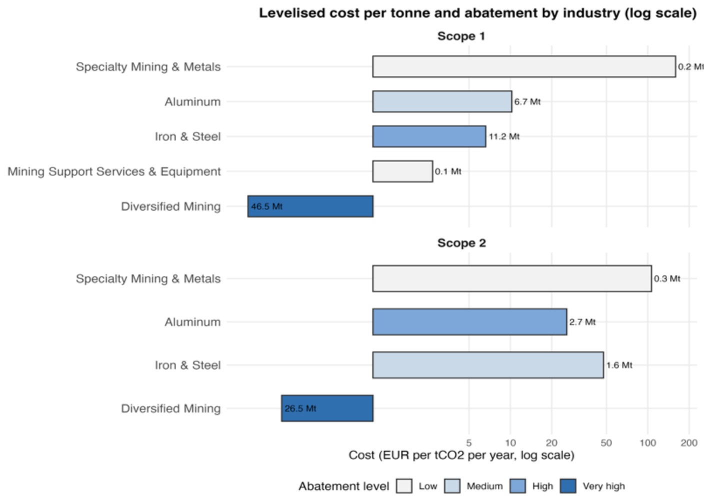


Figure 11 - Cost per tons and abatement by industry (log scale)

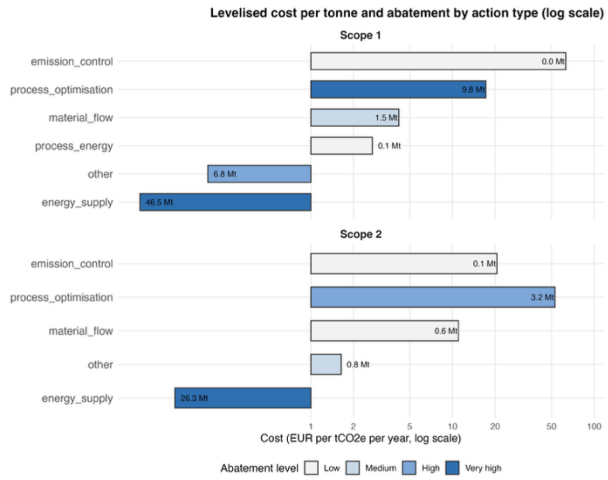


Figure 12 - Cost per tons and abetment by action (log scale)

6.4 Ridge prediction

For the final part of this analysis, I will introduce the machine learning components. The goal of this model is not to determine causal effects, as that is handled by the DiD framework. Instead, I aim to extend the observed firm-level treatment to companies that have not taken any action. This approach allows the model to serve as a prediction tool based on the same identification strategy, enabling firm-specific forecasts under different scenarios. The model is designed based on data constraints, the problem's structure, and guidance from empirical machine learning literature. There are two main reasons for including this component: first, the DiD and MACC methods estimate average causal effects at the industry and action levels. Predicting outcomes for firms that have never been treated requires a technique that captures this heterogeneity. Second, this thesis seeks to translate empirical findings into practical recommendations. For firms that have not taken any action, it aims to identify which measures might be expected to reduce scope 1 and scope 2 emissions.

The model is used here to generalize from treated firms to untreated firms, but within the methodological boundaries. Since the available dataset is small, it must balance flexibility with stability. Naturally, overly complex non-linear ML models, such as deep neural networks, gradient boosting, or random forests, would overfit severely and produce unstable estimates. Based on the selected architecture, a regularized linear model is chosen for its suitability for small sample prediction tasks and transparent economic interpretation.

The model is trained on the same event study structure as the DiD estimates for the treated firms and actions, using a pre- and post-treatment window constructed on the same principle. The realized change in Scope 1 and 2 log emissions is calculated. The percentage change is derived from the dependent variables for the prediction exercise. Since the variables are obtained from the same window as the DiD, the ML model's forecast effects are directly comparable to the causal estimates.

$$d_{lscope1} = \ln(post) - \ln(pre), \quad d_{lscioe1,pct} = e^{d_{lscope1}} - 1$$

Equation 6 - Event-window proportional change in emission

The baseline of covariates is constructed using the pre-treatment window. Variables include Scope 1, Scope 2, environmental score, emission score, and revenue. Each firm is also linked to its industry and the action type as act_key. Together, these variables form the predictor matrix. I also ensure that this dataset only contains treated firms, measuring that the model learns from the observed causal responses.

Based on the small sample size and the need for interpretability, the primary model is specified as a ridge regression, a linear model with L2 regularization. While a ridge regression solves the following optimization problem.

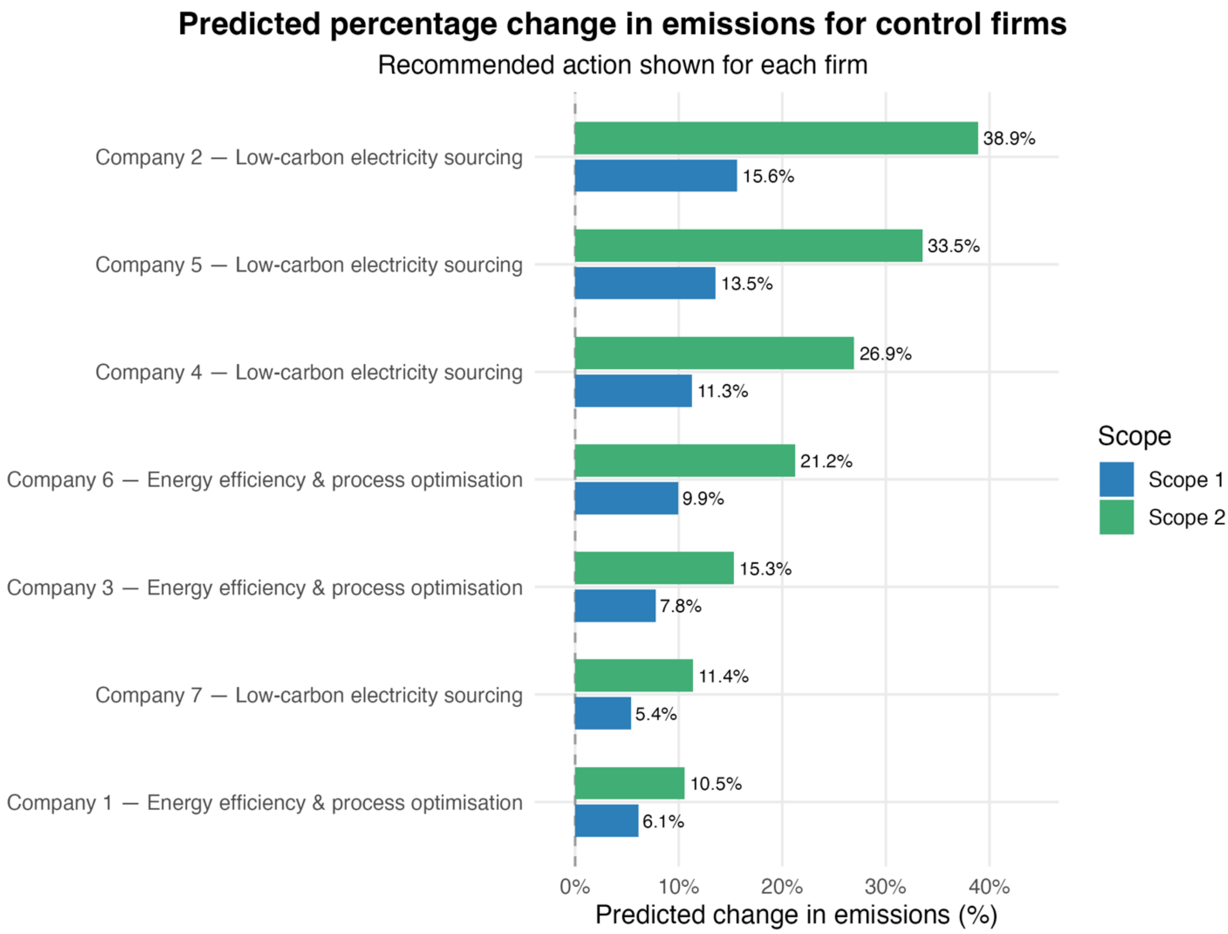
$$\hat{\beta}(\lambda) = \arg min_{\beta} \left(\sum_{i=1}^{N} (y_i - x_i{'}\beta)^2 \lambda \sum_{j=i}^{p} \beta_j^2 \right)$$

Equation 7 - Ridge regression (L2-penalized least squares) objective

Where λ is the hyperparameter controlling the degree of shrinkage. The penalization reduces the variance and aims to stabilize the coefficient estimates, in case when the number of predictors is large relative to the sample size (Hastie et al., 2009).

Two models are then estimated: (i) one predicting the expected change in Scope 1 emissions, and (ii) one predicting the change in Scope 2 emissions. For both models, industry and action types are included as factor variables. This allows for flexible differences across the sector and captures the covariance of firm-specific baseline characteristics.

Based on this, I aim to achieve three main objectives: (i) **Robustness**, regularities help to protect against overfitting when using small datasets. (ii) **Interpretability**, the estimated coefficients allow us to study how individual characteristics influence the predictive response. (iii) **Compatibility**, structure aligns with the economic logic of DID, and prediction reflects percentage changes as in the causal estimates.



7. Conclusion

7.1 Q1: How do the actions companies take lower their environmental scores?

The descriptive graphs shown in section 5.1.2 indicate no significant downward trend in emissions at the industry level over the same period. While median Scope 1 emissions remain mostly stable, Scope 2 shows a slight decline. This pattern does not suggest that the sector has failed to reduce emissions through targeted actions. Instead, it shows that industry-level medians might hide firm-level adjustments, which are offset by scale effects or different production changes. In contrast, the DiD analysis provides evidence of action effectiveness at the firm level. When evaluating the outcome, it confirms effectiveness at the firm level, especially for Scope 2. The effects align with actions related to energy supply, process optimization, and electricity efficiency measures. While Scope 1 effects are more varied and generally smaller, this probably results from measurement issues and the influence of process efficiency improvements, which produce larger average reductions than emission control measures. These have more limited short-term effects, but reductions at the firm level could be targeted for the longer term, which this paper does not cover.

7.2 Q2: In which industries have these actions been most effective in improving both financial and environmental performance?

As discussed earlier, the goal of emission-reducing actions is to ensure they are economically sustainable, which this thesis seeks to approximate by using the MACC framework. The MACC table displays the resources required for each action and how much these resources contribute to emission reductions. While this framework offers a structured way to compare actions across industries, it is not a perfect representation of the actual cost–abatement relationship. In particular, the lack of detailed operating expenditure (OPEX) data limits a full understanding of each action's total cost profile. Nevertheless, the MACC results reveal significant differences across industries and actions, with some combinations achieving much larger emission reductions at lower costs.

Beyond environmental performance, it is also important to consider how these actions interact with firms' financial outcomes. It is reasonable to believe that actions which negatively impact financial performance are unlikely to be sustained over time. Consistent with this idea, the results indicate that some sectors and actions manage to achieve both emission reductions and improvements in financial performance, suggesting that environmental and financial objectives are not necessarily in conflict. These outcomes are especially evident in actions related to energy supply and process efficiency, which tend to provide both cost savings and emission reductions.

7.3 Q3: What is the variation in the actions, and how can we predict the future trends?

The results reveal considerable variation in both the types and effectiveness of actions taken by firms across different industries. Even within the same industry, companies select different measures that vary greatly in cost, scale, and emission-reduction potential. This diversity is evident in the DiD estimates and MACC analysis, where similar actions can produce very different outcomes depending on firm characteristics, initial emission levels, and specific industry production processes. As a result, no single action consistently outperforms others across all firms; success largely depends on the specific context.

To address this variation and predict future trends, a machine learning component is integrated as a predictive tool within the causal framework. Instead of deriving new causal effects, the model learns from observed treatment outcomes and firm-level features to generate firm-specific predictions in hypothetical scenarios of action adoption. This approach allows the analysis to go beyond average industry effects and provide customized forecasts for firms that have not yet taken any actions. However, predicting future results is inherently limited by the small number of observed treatment events and the wide variability in responses.

7.4 Prediction power and significance

The empirical results indicate that both the causal and predictive components of the analysis exhibit significant statistical limitations. The difference-in-differences estimates do not have statistically significant effects across the aggregations. Although several coefficients trend in the expected direction, indicating reductions in emissions after actions are taken, the small number of action events and lower statistical power increase uncertainty for the estimated effects.

Likewise, the ridge regression models shows low predictive power when using standard accuracy metrics. This result can be explained by the small sample size and the inherently noisy relationship between firm features and emission outcomes. The goal of the machine learning model is not to produce highly precise point predictions but to offer a structured, internal recommendation for projecting observed treatment responses for untreated firms. In this way, the model acts more as a scenario-based decision tool than a forecasting device.

Overall, the limited statistical significance of the DiD estimates and the modest predictive ability of the ridge models reveal the trade-off between empirical ambition and data availability. Nonetheless, the results provide insights by integrating causal evidence with prediction. Future research could further explore these relationships to strengthen the robustness of the findings.

7.5 Overall Conclusion

Overall, the results show that emission reductions in the European mining and steel industries are driven by specific, action-level interventions rather than broad improvements in ESG or environmental scores. While descriptive industry-level trends reveal limited aggregate declines particularly for Scope 1 emissions the causal difference-in-differences analysis identifies heterogeneous but meaningful reductions following targeted measures, especially for Scope 2. This finding is consistent with prior literature documenting a weak link between sustainability scores and realized emissions (Berg et al., 2022; Gibson et al., 2023) and reinforces the need to distinguish symbolic ESG improvements from operational decarbonization.

By combining causal estimates with project-level capital expenditures, the MACC analysis further demonstrates that cost-effectiveness varies substantially across actions and industries, supporting the view that decarbonization in emission-intensive sectors is constrained by technological rigidity and capital intensity (Kesicki, 2012; Bolton & Kacperczyk, 2021). The machine-learning component complements these insights by illustrating the limits of purely predictive approaches: the low predictive power underscores that short-term emission responses are highly context-specific and difficult to generalize, a pattern consistent with recent work in applied machine learning for economic decision-making (Mullainathan & Spiess, 2017; Athey & Imbens, 2019).

Taken together, the findings directly address the research gap identified in the literature by providing firm-level, action-specific evidence on both environmental effectiveness and cost-effectiveness, rather than relying on aggregate scores or ex ante estimates. The results suggest that meaningful short-term emission reductions in mining and steel are achievable, but only through carefully selected measures whose environmental impact and financial implications are jointly considered. As such, the study contributes to the growing literature on carbon risk and decarbonization by demonstrating that effective climate strategies in heavy industry require targeted actions, causal evaluation, and transparent cost metrics, rather than broad ESG improvements alone.

8. Literature _

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Journal of Economic Perspectives,* https://doi.org/10.1146/annurev-economics-080217-053433

Athey, S., & Wager, S. (2021). Policy learning with observational data. https://doi.org/10.48550/arXiv.1702.02896

Atz, U., Holt, T., Kölbel, J. F., & Whelan, T. (2023). *ESG and financial performance: Uncovering the relationship*. NYU Stern Center for Sustainable Business.

Benuzzi, L., Brière, M., & Ramelli, S. (2024). ESG data providers: A comparative analysis of methodologies and implications. *Journal of Sustainable Finance & Investment*. https://www.researchgate.net/publication/394976667_A_comparative_analysis_of_ESG_and_non-ESG_securities_a_market_model_perspective

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

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

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics, 225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Candio, P. (2024). The effect of ESG and CSR attitude on financial performance in Europe: A quantitative re-examination. *Journal of Environmental Management, 354*, 120390. https://doi.org/10.1016/j.jenvman.2024.120390

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *American Economic Review,* https://doi.org/10.48550/arXiv.1608.00060

de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review* https://doi.org/10.1257/aer.20181169

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

Enkvist, P.-A., Nauclér, T., & Rosander, J. (2007). A cost curve for greenhouse gas reduction. *McKinsey Quarterly*,

Fowlie, M., Holland, S. P., & Mansur, E. T. (2012). What do emissions markets deliver and to whom? *American Economic Journal: Economic Policy, 4*(4), 63–93. 10.2139/ssrn.1416787

Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Pitman.

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

Friedman, M. (1970, September 13). The social responsibility of business is to increase its profits. *The New York Times Magazine*.

Galema, R., & Gerritsen, D. (2025). ESG rating changes and stock returns. *Journal of International Money and Finance, 154*, 103309. https://doi.org/10.1016/j.jimonfin.2025.103309

Gharleghi, B., Shafighi, N., & Nawaser, K. (2024). Green finance and its role in sustainability in the EU. *Cleaner Production Letters*. https://doi.org/10.1016/j.clpl.2024.100129

Gibson, R., Krueger, P., & Schmidt, P. S. (2023). ESG rating disagreement and stock returns. *Journal of Finance*.

Ibishova, A., et al. (2024). Corporate GHG emissions and financial performance: Sectoral heterogeneity in advanced vs. emerging economies. *Journal of Risk and Financial Management, 18*(10), 583. https://doi.org/10.3390/jrfm18100583

Ilhan, E., Sautner, Z., & Vilkov, G. (2021). Carbon tail risk. *The Review of Financial Studies, 34*(3), 1540–1571. https://doi.org/10.1093/rfs/hhaa071

International Energy Agency. (2023). *CO_2 emissions in 2022*. International Energy Agency. https://www.iea.org

International Energy Agency. (2023). *Iron and steel technology roadmap: Towards near-zero emissions*. IEA. https://www.iea.org/reports/iron-and-steel-technology-roadmap

Kathan, M. C., Utz, S., Dorfleitner, G., Eckberg, J., & Chmel, L. (2025). What you see is not what you get: ESG scores and greenwashing risk. *Finance Research Letters, 74*, 106710. https://doi.org/10.1016/j.frl.2024.106710

Kesicki, F. (2012). Marginal abatement cost curves for policy making—Expert-based vs. model-derived curves. *Energy, 47*(1), 86–95. https://doi.org/10.1016/j.energy.2012.08.019

Kesicki, F., & Strachan, N. (2011). Marginal abatement cost (MAC) curves: Confronting theory and practice. *Environmental Science & Policy, 14*(8), 1195–1204. https://doi.org/10.1016/j.envsci.2011.08.004

Knittel, C. R., & Stolper, S. (2022). Automating energy environmental policy analysis with machine learning. *SSRN Working Paper*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3500649

Liang, S., Song, R., Xiao, C., Liu, J., & Qu, S. (2021). Deep learning–based prediction of industrial emissions and mitigation potential. *Nature Sustainability, 4*(12), 1050–1059. https://doi.org/10.1016/j.spc.2025.03.015

McKinsey & Company. (2009). *Pathways to a low-carbon economy: Version 2 of the global greenhouse gas abatement cost curve*. McKinsey & Company.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87–106. https://doi.org/10.1257/jep.31.2.87

OECD. (2020). *Beyond growth: Towards a new economic approach*. OECD Publishing. https://www.oecd.org

Oestreich, A. M., & Tsiakas, I. (2024). Carbon emissions and firm profitability. *Journal of Sustainable Finance & Investment, 14*(4), 766–786. https://doi.org/10.1080/20430795.2024.2386476

Refinitiv. (2022). *Environmental, social and governance (ESG) scores methodology*. Refinitiv. https://www.refinitiv.com/en/sustainable-finance/esg-scores

Sachs, J. D., Woo, W. T., Yoshino, N., & Taghizadeh-Hesary, F. (2019). *Why is green finance important?* Asian Development Bank. https://www.adb.org/publications/why-green-finance-important

Schueth, S. (2003). Socially responsible investing in the United States. *Journal of Business Ethics, 43*(3), 189–194. https://doi.org/10.1023/A:1022981828869

Sparkes, R. (2002). *Socially responsible investment: A global revolution*. John Wiley & Sons.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics, 225*(2), 175–199. https://doi.org/10.1016/j.jeconom.2020.09.006

United Nations Framework Convention on Climate Change. (2015). *Paris Agreement*. https://unfccc.int/process-and-meetings/the-paris-agreement

United Nations Global Compact. (2004). *Who cares wins: Connecting financial markets to a changing world*. https://documents.worldbank.org

World Resources Institute, & World Business Council for Sustainable Development. (2004). *The greenhouse gas protocol: A corporate accounting and reporting standard* (Revised ed.). https://ghgprotocol.org

Wooldridge, J. M. (2021). *Introductory econometrics: A modern approach* (7th ed.). Cengage.