August 7, 2025

Prof. Caroline Flammer, Columbia University
Prof. Stefano Giglio, Yale University
Academic Coordinators, NBER PhD Student Workshop on Climate Finance
National Bureau of Economic Research (NBER)

Dear Professors Flammer and Giglio,

Please consider my application to the NBER PhD Student Workshop on Climate Finance on October 9, 2025. I am a PhD candidate at EM Lyon Business School and am submitting a complete draft of "Disentangling the Signals of Net-Zero Investment."

The paper asks which signals credibly predict firm decarbonization. A key finding indicates that net-zero investing should be pursued through efficiency gains rather than absolute cuts, highlighting that emission intensity necessitates capital-backed abatement instead of relying solely on divestment. Among firm-level signals, we identify effects using staggered SBTi adoption instrumented with country–year policy-driven variation in CDP reporting. Policy-induced SBTi commitments reduce emissions intensity with heterogeneity by policy exposure and balance-sheet strength; Random Forest models improve out-of-sample forecasts and narrow prediction intervals versus standard benchmarks, informing net-zero portfolio construction.

I would especially value feedback on further reducing and validating uncertainty—in emissions forecasts (calibration and coverage) and in treatment-effect estimation (timing, spillovers, instrument strength)—and on sharpening the transmission from credible decarbonization to expected cash flows and financing costs.

Information as requested:

- **Name:** Borui Qiu
- **Complete contact information:** qiu@em-lyon.com; 144 avenue Jean Jaur‘es, 69007 Lyon, France
- **University:** EM Lyon Business School
- **Expected PhD completion:** December 2026
- **Faculty adviser (reference):** Prof. Guillaume Coqueret (guillaume.coqueret@em-lyon.com)

Thank you for considering my application. I have attached my CV and the paper draft.

Sincerely,

Borui Qiu

EM Lyon Business School

144 avenue Jean Jaurès, 69007 Lyon, France

E-mail: qiu@em-lyon.com

Borui QIU

Lyon, France
+33 0603234322
QIU@EM-LYON.COM
BORUIQIU.COM

EDUCATION

Doctorant, EM Lyon QUANT Research Center, Lyon, France Nov.2023-Now
- PhD Candidate affiliated to Coactis, Université de Lyon
- Thesis Title: Financing the Green Era.

CEMS Master Double Degree - Bocconi University, Milan, Italy Aug.2020-Mar.2023
- MSc International Management(CEMS), with exchange term at Stockholm School of Economics, Sweden.
- Major track: Finance in supply chain management, courses: International Corporate finance, Macrofinance, Global operation and supply chain management, Global Strategy.

Bachelor of Engineering - Xi'an Jiaotong University (XJTU), Xi'an, China Aug.2016-Jun.2020
- Major in Chemical Engineering and Technology (Cum laude), Minor in Accounting (ACCA),
- Thesis: CCS treatment of flue gas combined with electrochemical wastewater treatment for coal-fired power plants.

INTERNSHIP EXPERIENCES

Bocconi GREEN Research Centre, Milan, Italy Research Assiatant Feb.2023-Nov.2023

Working Paper

Expanding the Zoo: The Circularity-Factor Coauthor with Maximilian Gobel and Claudio Zara

Abstract: The Circular Economy (CE) is a framework that falls in the vicinity of the environmental, social, and governance (ESG) paradigm. Yet, it relates much more to the notion of a self-sustaining production process that limits the production of non-reusable waste and the reliance on non-renewable inputs. As countries all around the world are starting to transition towards greener and more sustainable economies, investors may want to hedge themselves against the risks associated with such a transition. We therefore propose the circularity factor (CF), a hedging portfolio constructed by buying stocks with a high circularity score (CS) and shorting stocks with a low CS. Controlling for firm size, we find this to be a profitable strategy in European but not in the U.S. market.

Net-Zero Investing under decarbonizing uncertainty Coauthor with Guillaume Coqueret and Thomas Giroux

Abstract: This article sheds light on the drivers of decarbonization of global publicly-listed firms. We find that reported and estimated emissions depend on very different sets of variables. Net-zero investment on intensity improvement, different from traditional divestment-reduction association, requires investment support. We identify signals such as commitments to science-based targets (SBTi), regional climate policy, and firm-level financial resilience are critical indicators. We confirm a quasi-causal impact through staggered DiD of SBTi adoption using the IV of national-level CDP reporting. In terms of predictive accuracy, we find that machine learning models, specifically Random Forest, significantly improve forecasting accuracy for future emissions and reduce prediction un- certainty compared to traditional benchmarks. All in all, our findings deliver insights to asset managers seeking net-zero targets.

SKILLS SET

Computer skills: Matlab, Gurobi **Data analytics |** Python ML/AI
Language: Native speaker of Chinese Mandarin, fluent in English
Qualifications: ACCA Affiliate, CFA level 3 Passed, CFA Certificate in ESG Investing

I authorize the treatment of my personal data according to GDPR (EU) 2016/679.

Net-Zero Investing under decarbonizing uncertainty

Guillaume Coqueret* Thomas Giroux† Borui Qiu‡§

August 7, 2025

Abstract

This article sheds light on the drivers of decarbonization of global publicly-listed firms. We find that reported and estimated emissions depend on very different sets of variables. Net-zero investment on intensity improvement, different from traditional divestment-reduction association, requires investment support. We identify signals such as commitments to science-based targets (SBTi), regional climate policy, and firm-level financial resilience are critical indicators. We confirm a quasi-causal impact through staggered DiD of SBTi adoption using the IV of national-level CDP reporting. In terms of predictive accuracy, we find that machine learning models, specifically Random Forest, significantly improve forecasting accuracy for future emissions and reduce prediction uncertainty compared to traditional benchmarks. All in all, our findings deliver insights to asset managers seeking net-zero targets.

1 Introduction & Motivation

1.1 Context

Latest scientific consensus [1] warns that Earth is nearing 1.5°C of global warming, with a 70% chance that the 2025–2029 average will exceed this threshold as emissions continue to rise. The multiple threats posed by this alarming climate trend call for a steady decarbonization of economic systems worldwide.[2] To avoid the worst consequences of climate change and

*EMLYON Business School, 144 avenue Jean Jaures, 69007 Lyon, FRANCE. E-mail: coqueret@em-lyon.com

†CREST - Mirova, 59 Av. Pierre Mendès-France, 75013 Paris, FRANCE. E-mail: thomas.giroux@ensae-paristech.fr

‡EMLYON Business School, 144 avenue Jean Jaures, 69007 Lyon, FRANCE. E-mail: qiu@em-lyon.com

§COACTIS, Université Lyon 2, 14/16 av Berthelot, 69363 Lyon Cedex 07 FRANCE.

[1] WMO Global Annual to Decadal Climate Update

[2] We refer to Forster et al. (2023) and Lee et al. (2023) for recent updates on the estimates of the global temperature anomaly and to Malhi et al. (2020) and Richards et al. (2021) for perspectives on climate-induced damages.

achieve the ambition of the Paris Agreement to limit global warming below 1.5°C, Greenhouse gas (GHG) emissions need to drop by 45% by 2030 and ultimately reach net zero by 2050.

The financial industry may play a critical role in this regard, and investors are increasingly conscious and proactive on the matter (Krueger et al. (2020) and Bolton and Kacperczyk (2021)). For instance, in 2024, more than 300 institutions representing $60T of assets under management (AUM) had already joined the Net Zero Asset Managers initiative.

Both academics and practitioners have proposed avenues to reach so-called "*Net Zero*" targets.[3] Nevertheless, in all cases, a crucial input of such approaches is the measurement and reporting of current emissions and a credible assessment of their trajectories in the future. However, accessing current GHG emissions and trajectories accurately is challenging for investors.

Indeed, regulations requiring GHG disclosure varies around the world in their scope and accounting standards.

For example, The European Union implemented the Corporate Sustainability Reporting Directive (CSRD) in 2023, which requires companies operating in Europe to disclose their emissions beginning in 2025. It directly references the GHG Protocol's standards and is expected to impact over 50,000 companies across the European Union. Another example is California's Climate Disclosure Accountability Act from the same year. It requires companies with over $1 billion in revenues that operate in California to publicly disclose their scope 1 and 2 emissions starting in 2026, and their scope 3 emissions starting in 2027. At the federal level, the U.S. Securities and Exchange Commission (SEC) finalized a rule in March 2024 that will require companies to disclose some of their emissions if they're deemed financially material to investors. Overall, according to the World Resources Institute, just a handful of countries have mandatory GHG disclosure regulations.

This lack of reported data makes it difficult to accurately evaluate corporates GHG emissions at a large scale and therefore limits investors' ability to integrate this dimension in a consistent way into their investment decisions. Data providers have long developed estimation models to fill these gaps. However, transparency in modeling approaches and their accuracy remain limited in the industry. In continuation with the literature surrounding ESG scores discrepancies (Berg et al., 2022; Dimson et al., 2020) recent studies have documented the discrepancies among GHG data. Busch et al. (2022); Kalesnik et al. (2022); Papadopoulos (2022) compare several data providers and show that while correlations on reported direct emissions (Scope 1) are strong, around 0.97%, they tend to collapse for modeled indirect emissions (Scopes 2 and 3). This is of high importance since, the average disclosure level for the FTSE World was 58% in 2023[4]. Fortunately, Swinkels and Markwat (2023) find some improvements in recent years, especially for Scopes 1 and 2, with increasing homogeneity in reported data.

However, properly estimating current GHG emissions is only the first step in achieving net zero alignment. In addition, companies need to have ambitious transition plans to reduce their GHG emissions. To this end, as summarized by the Transition Plan Taskforce (TPT) in

[3] A non-exhaustive list is: Barahhou et al. (2022), Bolton et al. (2022), Le Guenedal et al. (2022), Le Guenedal and Roncalli (2022), Cenedese et al. (2023), Fraser and Fiedler (2023) and Roncalli (2024).

[4] "FTSE: Mind the gaps: Clarifying corporate carbon (2023)

its "Disclosure Framework"[5], a triptych is required: Ambition, Actions and Accountability. For investors to accurately assess transition risk and properly support this radical business changeover through efficient capital allocation, robust future emissions forecasts are key.

However, data providers and current research on the topic still rely on relatively simple models to predict future levels of GHG emissions. These predictions are usually simple extrapolations from recent trends at the company level (Le Guenedal et al., 2022) or portfolio level (Le Guenedal and Roncalli, 2022). Obviously, the outcome is suboptimal as these approaches do not take into account publicly available information, such as pledges, targets (e.g. Science-Based Targets) or transition plans (e.g. TPT).

In addition, according to Kalesnik et al. (2022), current forward-looking carbon scores from different data providers have very limited power to predict future changes in emissions. The authors' findings suggest that estimated emissions are at least 2.4 times less effective than self-reported emissions to predict future pathways. In addition, a recent exploratory study from Aldy et al. (2023) reveals how short-term commitments are only marginally met and emphasizes the difficulties of high-emitting companies in meeting their targets. They suggest that unrealistic and overpromising targets are the main issue toward a higher rate of success.

Therefore, building realistic, robust, and transparent systematic approaches to accurately predict future GHG emissions at the corporate level is critical to achieve world climate ambitions as set in the Paris agreement.

The present article seeks to propose a data-driven approach to determine the drivers of decarbonization in the cross-section of firms. To do so, we compile a large sample of firms around the world, where each company is characterized by an array of indicators, ranging from accounting ratios, to balance sheet composition, ESG scores, and greenhouse gas emissions.

1.2 Related literature

The present article sits at the intersection of several streams of academic research.

First, there is emerging research on firms' specific drivers of sustainable policies in general and GHG emissions reduction in particular, where current results show contradicting findings. In the more general case of environmental, social, and governance (ESG) performance, we point to the recent survey by Martiny et al. (2024). They emphasize the positive association between the proportion of ownership associated with socially responsible investors and future corporate social responsibility (CSR) scores (Hwang et al., 2022). Likewise, Kahn et al. (2023) finds a similar result for GHG emissions. Regarding debt financing, Flammer (2021) show that green bonds' issuers reduce their GHG emissions post-issuance, while Zerbib (2019) estimates a positive impact on the cost of capital, reducing the cost of the transition. On another dimension, Lee and Min (2015) and Habiba et al. (2022) point out that green R&D and innovations are associated with lower emissions.

Nevertheless, other contributions nuance the influence of ownership structure, financial flows, and financing costs. Atta-Darkua et al. (2023) argue that the real impact of sustainable investors is marginal with respect to the green transition. Moreover, Berk and Van Binsbergen (2024) show that in contradiction with theoretical findings suggesting that sustainable

[5] Summary of the TPT Recommendations

investing may produce a positive impact by making firms greener (Pástor et al., 2021), based on current data, the influence of divesting on the cost of capital is too low to meaningfully affect real investment decisions. In addition, Feldhütter and Pedersen (2023) propose a theoretical model implying that firm incentives to make green investments do not depend on their financing choice (debt versus equity). These findings are also supported by Heath et al. (2023) and Lam and Wurgler (2024). The first shows that SRI funds do not significantly change firm behavior, while the latter emphasize that only 2% of the proceeds from corporate green bonds support projects with new green features.

Besides, some studies look beyond correlation and aim to assess the causality of sustainable policies on firms' decisions. From a theoretical perspective, Acharya et al. (2023) propose a model on the impact of carbon taxes and green subsidies on the willingness of firms to invest in decarbonization technologies. The model predicts that large firms (or conglomerates) can benefit from investing in green technologies and lead the way toward decarbonization. Furthermore, Pedersen (2024) also shows that carbon prices are the most efficient tools to optimize social welfare (aggregate consumption minus dis-utility from pollution). Empirically, Adamolekun (2024) confirms that carbon pricing in the European Union reduces corporate emissions. In fact, Leffel et al. (2024) contend that state-level climate policies (e.g. financial incentives for energy efficiency) matter more for decarbonization than corporate decarbonization initiatives.

In a more descriptive way, Aldy et al. (2023) study the determinants of corporate carbon pledges and achievement. They show that companies' ability to succeed depends on the duration of the carbon pledge, the overall level of emissions, and the sales growth of each company.

Moreover, a few papers tackle the challenging exercise of ESG scores or GHG emissions forecasting based on statistical tools. Goldhammer et al. (2017) first introduce a set of company data and suggest Ordinary Least Squares (OLS) and Gamma Generalized Linear Regression (GGLR) to model GHG emissions forecasts. They achieve a higher match with reported data compared to simple extrapolation, but for a very limited set of European companies. Nguyen et al. (2021) propose an updated approach with the help of machine learning techniques to improve accuracy. They build a meta-learner that relies on the optimal set of predictors from different base models, including OLS, penalized regressions, neural networks, trees and clustering techniques, leading to a significant improve in and out-of-sample accuracy. However, they highlighted limitations in the disentanglement between direct and indirect emissions in their predictions, as well as discrepancies between industries. For example, the average R2 for most polluting industries in direct emissions (Scope 1) barely reaches 50%. Recently, Michalski and Low (2024) applied a similar approach to predict ESG scores. Pastor et al. (2024) also released a study of sectors' "*carbon burden*", computing the present value of future carbon emissions at sector level, and thereby trying to quantify the social externality arising from damages caused by corporate emissions of greenhouse gases. In doing so, they find that firms' emissions are predictable by past emissions, investment, climate score, and book-to-market.

Finally, we relate to a more mature research field studying the drivers of decarbonization at the macroeconomic level. Although most analysis tends to focus on a specific country at a time, we refer to those that take a panel approach and consider groups of countries. Several key variables stand out in this literature, such as trade openness and its variations

(Coskuner et al., 2020; Dogan and Seker, 2016b; Nguyen et al., 2021; Sharma, 2011). Other natural predictors include GDP per capita (Coskuner et al., 2020; Sharma, 2011), economic growth (Nguyen et al., 2021), energy consumption[6] (Coskuner et al., 2020; Dogan and Seker, 2016b; Sharma, 2011), environmental policies (Puertas and Marti, 2021), and financial development (Dogan and Seker, 2016b).

1.3 Summary of contributions

In the following paper, we contribute to the aforementioned literature through several ways. First, we explore the traits of firms undergoing decarbonization. One key finding is the significant gap between reported emissions and those estimated by data providers. According to our results, reported emissions largely rely on accounting data, while little consideration is given to pledges like those from the SBTI, which is surprising.

Another major insight from our results is that complex nonlinear models using panel data do not consistently outperform simpler methods. Additionally, extrapolation proves to be an unreliable benchmark due to the possibility of reversals in corporate emission strategies.

When it comes to identifying key variables for decarbonization, no clear narrative emerges. Both sustainability-linked factors, such as biodiversity footprint, and more conventional financial metrics, including historical returns and the book-to-market ratio, play a role. Moreover, analysts' sentiment and macroeconomic indicators frequently contribute to explaining carbon trends.

Our classification analysis also indicates that broader industry trends influence model performance. Specifically, the overall proportion of decarbonizing firms is closely linked to the accuracy of our models.

Lastly, we observe variations in prediction errors across different industries. Certain sectors, like utilities, are easier to predict, whereas others, such as information technology, present greater challenges. For sectors with high carbon impact—such as industrials, materials, energy, and real estate—the forecasting difficulty lies somewhere in between, making their emission trends neither straightforward nor highly unpredictable.

2 Data

2.1 Sources and construction

We collect carbon emission footprints under scopes 1, 2 and 3 for listed equities worldwide, covering a broad universe of developed and emerging markets for a period spanning from 2014 to 2023. We rely on ISS data as this is one of the main provider and it covers a very large database of both reported and estimated emissions. We also consider MSCI data as an alternative, but to ensure data consistency, we only include firms' observations from a single provider and do not mix sources. We compute our own estimate of GHG emissions intensities by scaling raw emissions by revenues sourced from Bloomberg. This approach

[6] In fact, Dogan and Seker (2016a) and Jiang and Guan (2016) show that the share of fossil fuels versus renewable in total energy consumption is a key driver.

has become standard in the industry, as well as academic research on the topic. Table 9 of the Appendix displays the name, source, and definition of each variable.

Our panel sample covers over 22,000 unique companies worldwide (North America, Europe, Japan, Asia ex Japan and Emerging markets) over the period 2014 to 2023. As shown in Figure 1, North America dominates in terms of number of companies and total market capitalization since 2015, followed by Europe and Asia. We also observe a significant bias toward small companies, reflecting the true nature of financial markets (see Table 1 for more details).

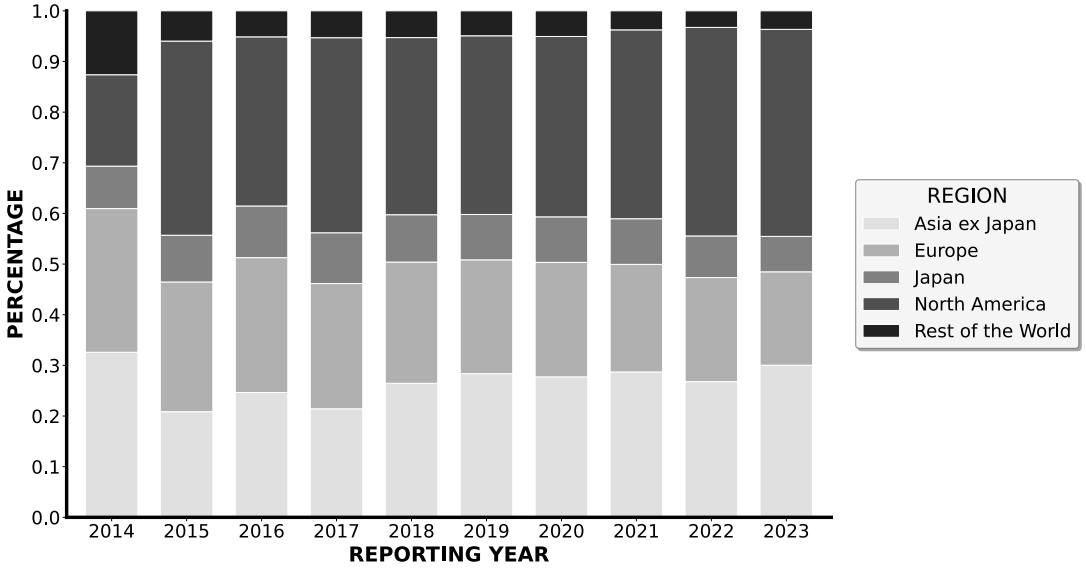


Figure 1: **Share of total market capitalisation by region and year**

Region	min	10%	50%	mean	90%	max
Asia ex Japan	0.50	58.00	736.00	4.33×10^3	7.37×10^3	1.74×10^6
Europe	0.50	48.00	990.00	1.03×10^4	1.83×10^4	7.86×10^5
Japan	3.70	55.00	376.00	3.94×10^3	6.04×10^3	9.80×10^5
North America	0.50	158.00	2.24×10^3	1.57×10^4	2.98×10^4	2.74×10^6
Rest of the World	0.60	65.00	1.06×10^3	5.97×10^3	1.08×10^4	2.92×10^5

Table 1: **Companies market capitalisation** (Million USD)

Finally, an increasing number of companies have begun reporting their direct GHG emissions over time, rising from 13% of the overall sample in 2014 to 55% in the most recent year, with a notable jump in 2022. Interestingly, the share of reported direct emissions as a proportion of the total universe footprint has followed a more modest upward trend, albeit from a much higher starting point. In 2014, the 13% of reporting companies accounted for 55% of total direct GHG emissions. By 2023, although 55% of companies reported, they represented only 83% of the total footprint. This indicates that more recent reporters tend to have relatively smaller carbon footprints than earlier ones. Nevertheless, focusing on reported

Scope 1 and 2 emissions still captures over 80% of the total universe footprint, making such analysis broadly representative.

Figure 2 summarizes this finding over time, and Table 11 provides descriptive comparison statistics for the two groups.

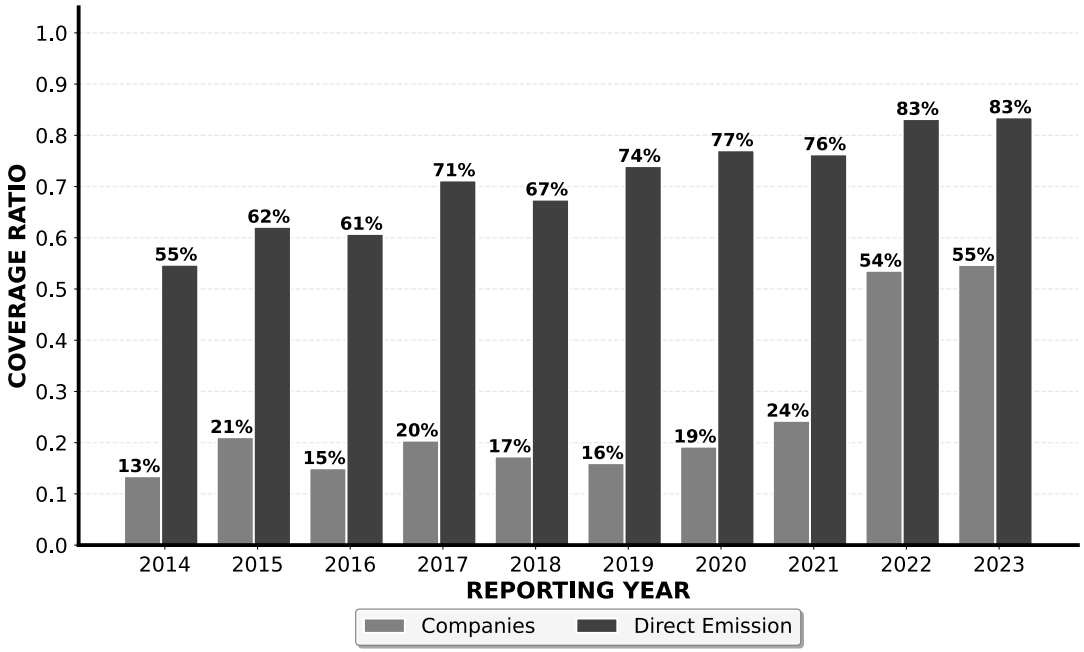


Figure 2: **Distribution of cumulative reported emissions over the period**

In addition, we identify a broad list of candidate drivers of firm decarbonization based on the literature. For clarity, we grouped them into the following categories.

Corporate Sustainability Indicators. First, we identify the ESG profile of companies as indicative of their awareness of climate-related issues and their ability to carry out change. Hence, we collect E, S, and G scores separately from ISS. Furthermore, because climate change and the biodiversity crisis are deeply intertwined, we take advantage of Iceberg-Datalab to include corporate biodiversity footprint (CBF) as a proxy of firm impact on biodiversity.

Finally, we include firms' GHG reduction targets collected by MSCI. This includes whether targets are validated by the Science Based Target Initiative (SBTi), among others. Firms' climate commitments appear to matter in the literature, as Bolton and Kacperczyk (2023) and Ramadorai and Zeni (2024) show evidence that firms that commit to reducing their carbon emissions tend to do so subsequently.

Corporate Fundamental & Market Indicators. Second, we consider the financial ability of companies to engage in decarbonization. To this end, we collect market and fundamental variables from Bloomberg, including geography and industry segmentation (GICS), market capitalization, price to book, debt to equity, free-cash-flows, long-term investments, R&D spending and others.

Corporate Ownership & Debt Structure. Third, according to several studies, the own-

ership structure influences both climate disclosure and climate performance. Following insights from Bolton and Kacperczyk (2021); Cohen et al. (2023) we include the percentage of shares held by institutional investors and the percentage of insider shares' outflow as provided by Bloomberg. Kahn et al. (2023); Rink et al. (2024) also suggest that active shareholders who engage with companies can influence their environmental and social performance. Therefore, we compute our own estimate of active sustainable shareholders. We estimate the latter by aggregating annual holdings from a comprehensive list of 780 active sustainable equity funds computed by Exane Research (2023). Details about the funds' sample can be found in the online Appendix.

Regional Indicators. Finally, we also believe that macroeconomic and regional characteristics play a significant role in firms' ability and willingness to tackle climate change and reduce their GHG emissions. We include the carbon intensity of energy mix from the IEA, and an environmental policy score from the OECD. It is worth noting the US are not covered by the OECD data and are therefore considered as laggard in advocating ambitious environmental policies during the 2014 – 2024 period.

A summary of all variables names, sources and interpretations can be found in Table 10 in the Appendix. Because corporate carbon emissions are updated on a yearly frequency, we keep this reporting frequency in the construction of our dataset. For variables with higher frequency such as market variables, we compute averages of past quarterly observations when relevant. Descriptive statistics of the universe are provided in Sub-section 2.4.

In the following study, we use GHG emissions as the dependent variable and lag all independent variables by one year in order to conduct a predictive analysis.

2.2 Outliers detection, panel construction and data imputation

In the following section, we provide further details on the intermediary steps taken to construct a consistent and clean dataset from the various sources previously mentioned.

Outliers treatment. We begin by excluding observations with non-positive GHG emissions. Next, we identify and remove outliers based on year-over-year changes in either absolute emissions or emissions intensity (normalized by sales). An observation is flagged as an outlier if it exhibits a reduction greater than 50% or an increase exceeding 200% compared to the previous year's reported value. Finally, to mitigate the influence of extreme values, all independent variables are winsorized at the 0.1^{th} and 99.9^{th} percentiles within each sector.

Panel construction. To be included in our panel dataset, firms must have at least four consecutive years of valid GHG data reported. If there are breaks in the data due to outlier treatment, we keep the most recent uninterrupted run only.

We then divide all firms into two subsets: *i) Reported* sample, which includes firms that have reported their own GHG emissions for at least one year for both Scope 1 and Scope 2; and *ii) Modeled or Estimated* sample, which includes firms for which all GHG data are estimated by the data provider.

Data imputation. We perform data imputation for independent variables with a specific approach for each group. For sustainability indicators, we apply linear interpolation, while for the financial and market variables, following Chen and McCoy (2024), we resort to cross-sectional mean imputation at the sector and size level. In practice, for all ratios and

percentage variables, we group the observations by GICS sector and fill the missing values with the cross-sectional mean of the sector each year. For absolute variables, we use the same approach, but we also take into account the market capitalization of the company compared to the sector average while grouping.

2.3 Omitted variables due to colinearity

Figure 10 in the Appendix presents the correlation matrix of all independent variables as defined above. We observe several groups of highly correlated variables and therefore drop some of them to avoid multicollinearity issues in the estimation of linear models.

First, we find that market capitalization, revenue, enterprise value, and lagged GHG emissions are highly correlated. To address this, we retain only lagged emissions, which we consider the most relevant predictor of firms' future emissions and a way to mitigate multicollinearity. ~~Second, the amount of green bonds issued is strongly correlated with the current repayment amount. We choose to keep the repayment variable because it better captures the financial drawdown and, by extension, the firm's tangible commitment to the green transition.~~ Second, all regional indicators from the OECD display strong mutual correlations. For linear models, we retain only the cross-sectional policy score, which serves as a summary indicator. We also include the GDP growth rate of the country where each firm's headquarters is located, as a macroeconomic control. Finally, energy mix variables from the IEA are also highly correlated. For linear models, we retain only coal, oil, and renewable shares; guided by the prior belief that these sources are more indicative of a firm's potential to reduce emissions. In contrast, for non-linear models such as random forests, we keep the full set of energy mix variables and allow the model to determine the most relevant predictors.

2.4 Descriptive statistics

Following the grouping described in Sub-section 2.2, we split the analysis in two groups, companies belonging to the *i) reported* sample and those belonging to the *ii) modeled or estimated* sample. We further consider four types of predictors for each Scope:

1. Raw emissions measured in tons CO2 equivalent (tCO2e) - which we refer to as **GHG**;
2. Emission intensities (**INT**), reflecting tCO2e per million of revenue
3. Relative changes in absolute emissions (Δ**GHG**);
4. Relative changes in emissions intensities (Δ**INT**).

Formally, if $\text{GHG}_{t,i}$ is the emission value of firm i and time t, then $\Delta\text{GHG}_{t,i} = \text{GHG}_{t,i}/\text{GHG}_{t-1,i} - 1$, and similarly for intensities.

We focus our baseline analysis on reported Scope 1+2 emissions (Direct Emissions) because they are less prone to disagreement and divergence, and because current indirect emissions (Scope 3) highly rely on the modeling process done by the data vendor. As an illustration, 96% of Scope 3 GHG emission in our sample are modeled. We run robustness tests on estimated emissions for scope 1+2 and both reported and estimated scope 3 emissions.

Figure 3 below and Table 11 in the Appendix provide an overview of the distribution of these variables for Scope 1+2.

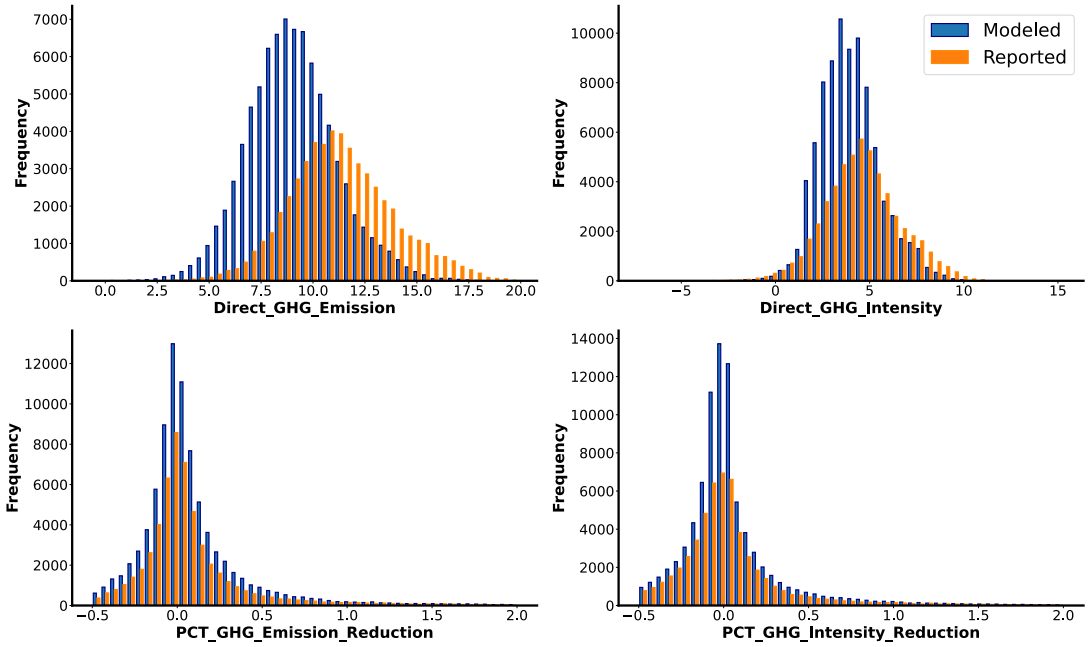


Figure 3: **Distribution of dependent variables - Scope 1+2**.

We see that reported Scope 1+2 GHG emissions tend to be larger than modeled emissions while the dynamic (yearly percentage change) has a similar distribution. This is in line with the findings of Berg et al. (2024). Several arguments are plausible to explain the differences between reported and modeled emissions. On the one hand, mandatory disclosure does not apply to all companies and may induce a size and sectorial bias as large companies and industries with significant climate-related financial materiality are targeted first by regulators. In addition, voluntary disclosure requires financial means and skills that small companies may not have or afford. On the other hand, modeled emissions may underestimate actual emissions due to model bias such as sectorial approximation and inclusion of financial variables.[7]

Turning to the exogenous variables used as potential predictors in our analysis, we present their main descriptive statistics in Table 12. We confirm the suggested bias toward larger companies, trading at higher valuations and generating higher profit margins on average for our reported sample compared to companies in the modeled sample. Reporting companies also tend to have better ESG and environmental scores, and more ambitious SBTI reduction targets than non-disclosing companies. However, we do not observe any significant differences in terms of institutional ownership nor sustainable active ownership.

[7] Due to a lack of precise model definition, we can't test this assumption.

3 The drivers of decarbonization

This section aims to identify the key determinants of decarbonization across the full sample. We begin in Sub-section 3.1 by estimating linear models, including both standard panel regressions and penalized versions. As noted in Sub-section 2.1, all explanatory variables are lagged by one year to preserve a predictive structure.

Next, in Sub-section 3.2, we employ random forest models to capture potential nonlinear relationships between firm characteristics and changes in emissions or emission intensity.

Finally, in Sub-section 3.3, we classify firms based on their realized emission trajectories and examine which characteristics differ most between those that reduce their footprint and those that continue to emit more.

3.1 Linear models - OLS and Lasso

Our analysis starts with linear models. We recall that we work with the four dependent variables outlined in Sub-section 2.4. Henceforth, we write $y_{t+1,i}$ for the time-t value of one of these variable for firm i. The model thus reads:

$$y_{t+1,i} = \alpha_i + a_t + \sum_{k=1}^{K} x_{t,i}^{(k)} b^{(k)} + e_{t+1,i}, \tag{1}$$

where the $x_{t,i}^{(k)}$ are the K predictors retained for the analysis, and listed in Table 12 in the Appendix. Note that these predictors include the lagged values of the dependent variables ($y_{t+1,i}$).

Table 2 shows the corresponding estimates with individual fixed effect. To ease interpretation, we only report t-statistics, as they provide the key message we are interested in: the sign of the relationship, and its strength. We tested two estimators for standard errors, namely Newey and West (1987) and Beck and Katz (1995), but as the latter yielded more conservative values (smaller t-statistics in absolute value), we only provide this version. Significance decisions are emphasized with a green background in the results Table.

Linear regression analysis reveals several common findings for both reported and estimated data.

A first takeaway is that past GHG emissions are consistently a strong predictor of future emissions, highlighting the inherent inertia in emission trajectories. Indeed, lagged values of absolute emissions (GHG) and intensities (INT) are associated with the largest statistics, by far. High levels of GHG emissions and intensities are associated with high levels of the same variables in the following year. In addition, we also observe that higher GHG emissions and intensities level, allow for stronger reduction potential in the following years. Indeed, high GHG emissions and intensities are associated with a negative t-statistic when explaining their future variations. This result is in line with the main findings from Pastor et al. (2024) analysis of the drivers of GHG emissions forecasts.

A second takeaway is that more than just a handful of variables are relevant for explaining the levels and changes in absolute GHG emissions and intensities. In fact, nearly every category described in Sub-section 2.1 contains at least one significant predictor. However, we uncover notable discrepancies between reported and estimated data, which raises concerns

	GHG (log)		INT (log)		ΔGHG		ΔINT	
	Reported	Modeled	Reported	Modeled	Reported	Modeled	Reported	Modeled
Direct GHG Emission	51.895	71.883			−26.161	−35.699		
Direct GHG Intensity			47.419	58.816			−25.265	−33.826
ISS Environment Score	−0.238	0.487	−0.619	−0.414	−1.473	5.968	−2.034	4.699
ISS Governance Score	0.020	1.558	1.732	4.833	1.315	0.402	0.643	0.632
ISS Social Score	0.806	0.588	0.678	−2.637	−1.593	−0.085	−2.055	−1.559
Dummy SBTi Approved	−3.275	−0.672	−1.814	−0.650	−6.303	−4.751	−6.635	−2.819
Years to Closest Target	0.022	−0.120	−0.422	−1.368	1.190	−0.029	1.718	−0.817
Biodiversity Score	0.856	0.841	−0.095	−3.590	−0.316	0.896	−3.099	−5.513
Debt to Equity	0.016	−0.661	−1.279	0.203	1.049	−0.559	−1.076	−0.036
Free Cash Flow	−0.321	−0.446	−0.732	−0.842	−0.971	−0.260	1.363	−1.026
Interest Coverage	−4.916	−2.173	−2.608	−2.369	−4.499	0.373	−2.868	−0.759
Longterm Debt	0.686	0.358	0.121	0.006	−0.320	3.793	0.647	3.485
Longterm Investment	0.081	−0.299	0.258	−0.882	−1.560	1.861	−1.107	0.840
Price to Book	−0.346	0.283	−0.690	−0.598	0.854	0.091	−1.182	−1.579
Price to Earning	0.346	0.009	0.351	0.306	−1.115	0.265	−1.769	0.302
Profit Margin	1.118	−1.501	−0.380	3.878	0.881	2.404	−0.174	−1.757
R&D Expenditure	2.119	2.313	0.437	−0.564	−0.543	−0.517	0.846	0.068
12M Stock Return	−0.056	0.972	−6.275	−1.006	7.065	0.027	−5.105	−0.788
Institutional Share	7.722	0.176	4.699	−2.674	1.060	0.156	1.137	−0.546
Internal Share Outflow	1.312	−0.981	2.856	0.155	1.425	2.248	1.367	1.597
Share Repurchase	−1.499	−0.052	−0.750	−0.135	−0.234	−0.832	2.340	−0.887
Green Fund Holding	0.261	−1.080	0.705	−1.571	−1.134	−1.093	−0.999	−1.087
Coal Energy fraction	−10.171	−8.993	−9.091	−11.302	−3.264	0.875	−5.334	−2.572
Oil Energy fraction	−6.437	−6.327	−8.027	−9.889	−2.264	−1.318	−4.888	−4.409
Renewable Energy fraction	−7.402	−1.948	−8.556	−5.714	−4.039	0.511	−7.558	−5.107
Policy Stringency	−1.870	0.312	0.211	4.122	−4.830	−0.195	−3.402	4.867

Table 2: **Panel model - individual fixed effects**. We report the t-statistics for the panel models defined in Equation (1) - all independent variables are lagged. The regressions employ two-way fixed effects (TWFE) to account for unobserved heterogeneity. The overarching column names pertain to the dependent variables. The sub-column panels pertain to the type of emissions considered as dependent variable. Standard errors are computed following Beck and Katz (1995). Colors code when statistics are larger than 2.58 (**light green**, 1% confidence level) or 3.3 (**darker green**, 0.1% confidence level) in absolute value.

about the robustness of current estimation models from data providers and underscores the need for further refinement.

Sustainability Indicators We observe significant and positive coefficients associated with environmental and governance scores for a few dependent *modeled* variables in the modeled sample, but significance vanishes for *reported* variables. The sign of the coefficient is somewhat counterintuitive, as it suggests that firms with higher ESG scores tend to have higher GHG emissions or weaker reduction trajectories in the modeled data. This finding contradicts previous literature on the role of firm governance in shaping corporate carbon policy (Elsayih et al., 2021; Haque, 2017; Oyewo, 2023), and raises questions about the precision of modeled estimates and the adequacy of ESG metrics as predictors in this context.

In addition, ambitious reduction targets reported to the SBTi have a significant effect on lowering future emissions, both in absolute terms and relative intensities. This suggests that

companies with science-based and ambitious targets are indeed reducing their emissions over time, faster than their peers without such targets, and by a greater absolute amount, even after controlling for firm fixed effects. Interestingly, this effect is not reflected in modeled estimations for absolute emissions but emerges only in the dynamic (delta), highlighting a possible limitation of current modeling approaches. In the last year in our dataset, only 16% of the firms have approved SBTi targets. This relatively low adoption rate underscores the potential for broader climate impact if such initiatives were scaled further. Moreover, the time horizon of the nearest target does not appear to have a significant effect. This raises important questions about whether the proximity of target deadlines actually influences decarbonization efforts, or whether other factors such as implementation capacity, regulatory context, or investor pressure play a more decisive role. We explore these potential drivers in more detail in the remainder of the analysis.

Finally, firms' biodiversity footprint also emerges as a significant driver, underscoring the interconnectedness between climate change and biodiversity impacts. Corporate biodiversity footprint, measured as a negative value leads to higher GHG emissions when it increases in absolute terms.

Fundamental and Market Indicators Free cash flow appears to have no significant effect, despite expectations that it might facilitate investment in transformative projects. Similarly, R&D spending does not emerge as a significant predictor. In contrast, financial leverage and resilience do appear to play a meaningful role. Interest coverage is positively associated with both absolute emission reductions and intensity-based abatement, suggesting that firms with stronger financial buffers are better positioned for the green transition. These firms are likely more capable of investing in capital-intensive low-carbon technologies and benefit from easier and cheaper access to green financing. Moreover, coefficients associated with recent financial performance—proxied by 12-month stock returns—reinforce this finding for emission intensities (i.e., GHG emissions scaled by revenues). Interestingly, however, they are positively associated with increases in absolute emissions. This asymmetry suggests that while better-performing firms may become more carbon-efficient, this does not necessarily translate into a reduction in their total emissions. Such divergence is particularly relevant in the context of the carbon premium discussion, as it underscores the distinct mechanisms driving improvements in emission intensity versus reductions in absolute emissions. It also highlights the importance of distinguishing between these two metrics when assessing real-world climate impact. Financial indicators such as the price-to-book ratio and profit margins are significant predictors in explaining estimated, but not reported—emissions. This is consistent with findings from Pastor et al. (2024), which suggest that data providers like MSCI may incorporate such financial ratios into their emissions estimation models.

Ownership and Debt Structure The influence of institutional ownership on GHG emission levels and reduction dynamics appears significant in the reported data but is absent in the estimated data. In contrast, active green ownership does not show a statistically significant effect on future emissions. The positive coefficient associated with institutional ownership suggests that asset owners could play a pivotal role in shaping the transition to a low-carbon economy. However, it also indicates that this potential remains largely untapped, as institutional investors currently tend to overweight firms with high absolute GHG emissions and emission intensities.

Regional Indicators The macroeconomic environment, particularly the energy mix, plays

a key role in shaping emission reduction dynamics. The relative use of coal, oil, renewables and gas emerges as common factors influencing both absolute and relative emissions. Despite the potential colinearity between these energy sources, the strong statistical significance of several variables suggests that the energy transition operates through multiple channels, collectively contributing to long-term net zero objectives. This finding is consistent with the observed effect of policy stringency. The OECD Environmental Policy Stringency Index significantly drives reductions in absolute GHG emissions, but has limited impact on emission intensity. This suggests that current macroeconomic instruments, such as carbon taxes or emissions trading schemes, are more effective in curbing total emissions than in improving firms' carbon efficiency through technological innovation or process optimization.

As a robustness check, we provide in Table 13 in the Appendix the results of the regression with industry fixed effect.

The main findings remain qualitatively consistent: GHG emissions show strong persistence over time, with past levels of emissions and intensities highly predictive of future values. Firms with higher initial emissions also tend to reduce emissions more significantly, confirming earlier results. A wide range of sustainability, financial, ownership, and regional variables help explain emissions patterns, although discrepancies between reported and estimated data persist. Ambitious SBTi targets are strongly associated with subsequent reductions in reported emissions. Financial resilience—captured by interest coverage and stock performance—supports emission reductions, though gains in intensity do not always lead to lower absolute emissions. Institutional ownership is linked to higher emissions in reported data, pointing to unrealized potential for investor-driven decarbonization. Structural factors such as energy mix and policy stringency remain robust drivers of emission trends.

However, we also observe notable differences. First, unlike in the company fixed effects model, ESG scores emerge as significant predictors and are generally associated with lower future emissions, as expected, across both reported and modeled data. An exception is the Governance score in the modeled sample, which surprisingly shows a positive association with emissions, raising questions about the interpretation and construction of estimated variables. Second, R&D spending is linked to higher initial emission levels but also to greater reductions in emission intensity. This suggests that while firms investing in innovation may not immediately reduce their absolute emissions, they improve efficiency, potentially optimizing internal processes rather than delivering real-world decarbonization.

In order to determine which variables matter the most, we turn to LASSO (Tibshirani (1996)) penalized regressions, a common feature selection tool in statistics. We rewrite Equation (1) as

$$y_{t+1,i} = X_{t,i}b + e_{t+1,i},$$

where the matrix $X_{t,i}$ includes the fixed effects. The LASSO solves

$$b^* = \underset{b}{\operatorname{argmin}} \, ||y - Xb||_2^2 + \lambda ||b||_1,$$

where the shrinkage parameter, λ determines the stringency of the penalization. The larger it is, the more sparse the model becomes (more estimated coefficients are set to zero). By default, the {glmnet} package in the R language spans a large number of values of λ and

generates a full matrix of estimates $\hat{B}_{k,j}$, where j is the index of λ_j. We are then interested in the proportion of times that a variable survives the penalization:

$$p_k = J^{-1} \sum_{j=1}^{J} 1_{\{\hat{B}_{k,j} \neq 0\}}. \tag{2}$$

We run the models separately on reported and estimated emissions. The corresponding values (in percents) are gathered in Table 3. We only list the top 15 variables for each category: raw footprint include GHG and INT, while changes are ΔGHG and ΔINT.

To be consistent with the initial specification in Equation (1), we include fixed effects in the model. However, they are rarely among the surviving variables and none even make it to the top 15. For the sake of completeness, we also ran models without fixed effects, and the results are qualitatively the same; the ranks are simply slightly altered (see Table 14 in the Appendix).

		Raw footprint				**Relative change**			
		GHG (log)		INT (log)		ΔGHG		ΔINT	
rank		rep.	est.	rep.	est.	rep.	est.	rep.	est.
1	Direct GHG Emission	**99**	99			**99**	99		
2	Direct GHG Intensity			**99**	99			**99**	99
3	12M Stock Reutrn	0	3	60	22	77	0	71	38
4	Policy Stringency	49	34	51	38	73	61	73	37
5	Dummy SBTi Approved	43	23	40	34	71	51	74	47
6	Renewable Energy pct	67	75	63	34	67	78	77	68
7	Coal Energy pct	71	41	67	43	66	47	67	24
8	ISS Environment Score	0	12	26	18	58	69	66	66
9	ISS Social Score	31	20	32	45	53	20	59	43
10	GDP Growth	43	48	34	39	56	74	49	51
11	Internal Share Outflow	35	47	51	8	53	28	52	42
12	Years to Closest Target	0	15	30	42	51	18	52	42
13	Price to Earning	10	0	12	8	49	0	51	18
14	Price to Book	8	0	32	1	55	18	45	28
15	Longterm Investment	14	23	27	23	49	40	45	18

Table 3: **Lasso survival rate**. We report the percentage of times that a given variable survives LASSO selection (see Equation (2)). The rank of the variable is determined by the average of the the four columns. Fixed effects are included in the model.

Although the results differ slightly from those obtained by standard panel regression, they confirm the main findings of the previous analysis, as presented in Table 2. We also continue to observe a systematic divergence between reported and modeled data across most variables, once again emphasizing the need for further refinement when working with estimated emissions.

First, past emissions emerge as the strongest predictor of future emissions, reaffirming their central role in forecasting. Second, SBTi commitments demonstrate substantial predictive power for future changes in emissions rather than current levels. This aligns with the strong negative t-statistics observed in the fixed-effects panel regressions. Yet again, the modeled data shows weaker effects, suggesting that estimation models fail to adequately

capture the impact of firm-level climate action. In addition, financial variables, such as 12-month stock returns and valuation ratios like price-to-earnings and price-to-book—consistently rank among the top predictors. Macroeconomic indicators, including policy and energy infrastructure variables, also display robust predictive power across all model specifications. In particular, the shares of coal and renewable energy in the national energy mix are selected by the LASSO procedure in 60–70% of cases for reported data, while policy stringency remains a consistent and significant predictor for both absolute and intensity-based reductions.

Interestingly, the Environmental score also appears as a significant predictor in this specification, despite being insignificant in the individual fixed-effects panel regression. This discrepancy underscores the current ambiguity in the academic literature regarding the predictive value of ESG scores, with recent studies such as Kalesnik et al. (2022); Pastor et al. (2024) reporting diverging results.

3.2 Inference from non-linear models - Random forests

To further explain emissions and intensity changes over time, we leverage on random forests. The rationale for this choice is two-fold. First, tree-based supervised learning performs very well on tabular data (Grinsztajn et al. (2022), Januschowski et al. (2022), Shwartz-Ziv and Armon (2022)). Second, these methods are readily interpretable via the feature importance metrics that are computed once the models are trained (see, e.g., Molnar (2020)). In addition, statistical tests have also been developed to evaluate the significance of features in these models (see, e.g., Mentch and Hooker (2016)), as well as confidence intervals based on resampling and deleted-d jacknife estimators (Ishwaran and Lu (2019)). We follow this approach in our analysis.

In Figure 4, we report the confidence regions for the feature importance, computed over 100 bootstraps. The most important predictor is shown at the top and all values are normalized so that the top feature has an average importance of 1.

In this section, we focus on changes in GHG emissions and intensities, as absolute levels are largely driven by their past values. Consequently, non-linear modeling is expected to provide limited value added when explaining absolute emissions, but it may better capture the dynamics of change.

Our results confirm the predictive power of many variables identified in the previous analysis. For both GHG abatement and intensity reduction, past GHG emissions remain the most important predictor. Other key variables include 12-month stock returns, components of the IEA energy mix, and a variant of the SBTi-approved target, specifically the count of climate commitments a company has made, as compiled by MSCI.

Interestingly, some variables that showed limited significance in linear panel models, such as financial ratios (price-to-book, price-to-earnings) and firm fundamentals (revenues and free cash flow), now exhibit much stronger predictive power in non-linear models. This suggests that non-linear relationships exist between financial characteristics and carbon abatement potential, which linear approaches may fail to capture. Additionally, we observe increased importance for macroeconomic indicators like GDP growth, which had minimal impact in the linear models.

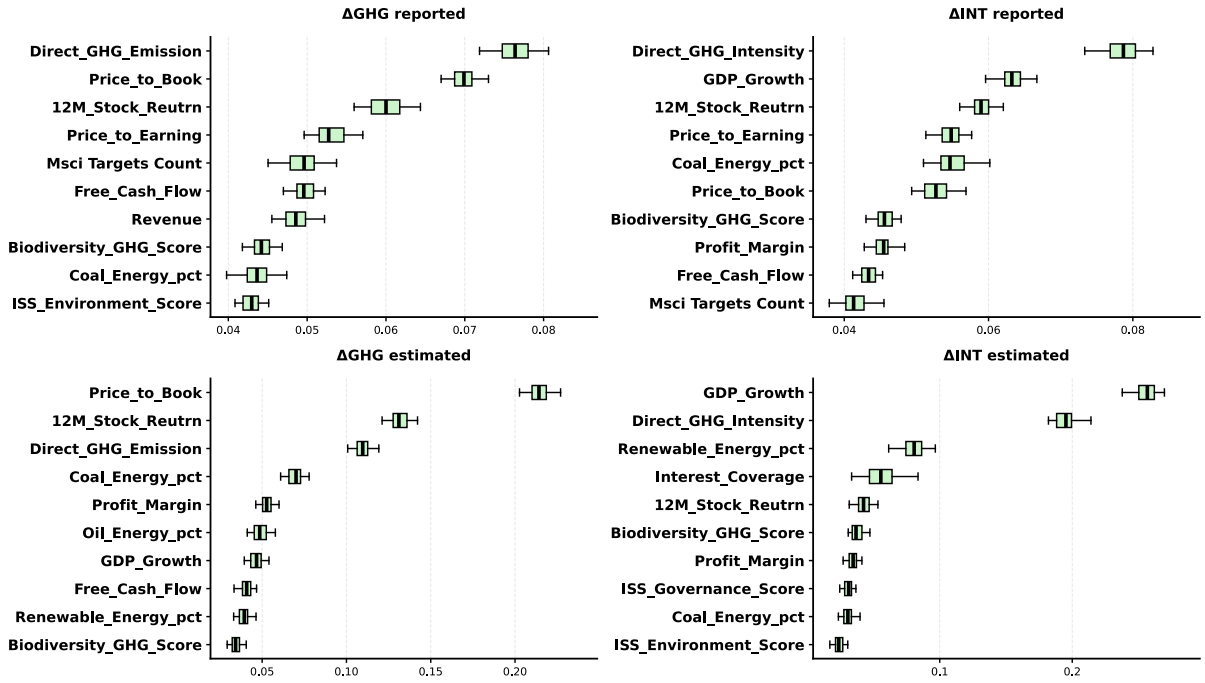


Figure 4: **Confidence regions for feature importance**. We plot the median (center), interquartile range (boxes) and 95% interval (whiskers) for the variable importance, following Ishwaran and Lu (2019). The number of bootstrap samples is 100. The forests were fitted separately on the estimated (left panels) and reported (right) variations, as well as on raw emissions (top panels) and intensities (bottom).

3.3 The characteristics of decarbonizing firms

In the previous sections, we focused our analysis on identifying the cross-sectional drivers of firms' carbon emissions and abatement. In the following subsection, we reverse the perspective and examine the characteristics of firms sorted by their GHG emissions trajectories.

First, we split the sample, on a year-by-year basis , in two groups: firms with low versus high dependent variable.[8] Then, **we aggregate all firms classified into each group to form two pooled samples**. Finally, for each characteristic, we test if the average value is significantly different between the two groups, for instance, large GHG emissions reduction versus low GHG emissions reduction.

Figure 5 displays our findings for levels of GHG emissions while Figure 11 provided in Appendix extends the analysis to emissions intensities. We plot the absolute t-statistics of the tests for which the null hypothesis is that there is no difference between the two groups. The results are consistent with previous analysis.

[8]This approach is used in Seyfi (2024) to understand the characteristics of low versus high return US stocks.

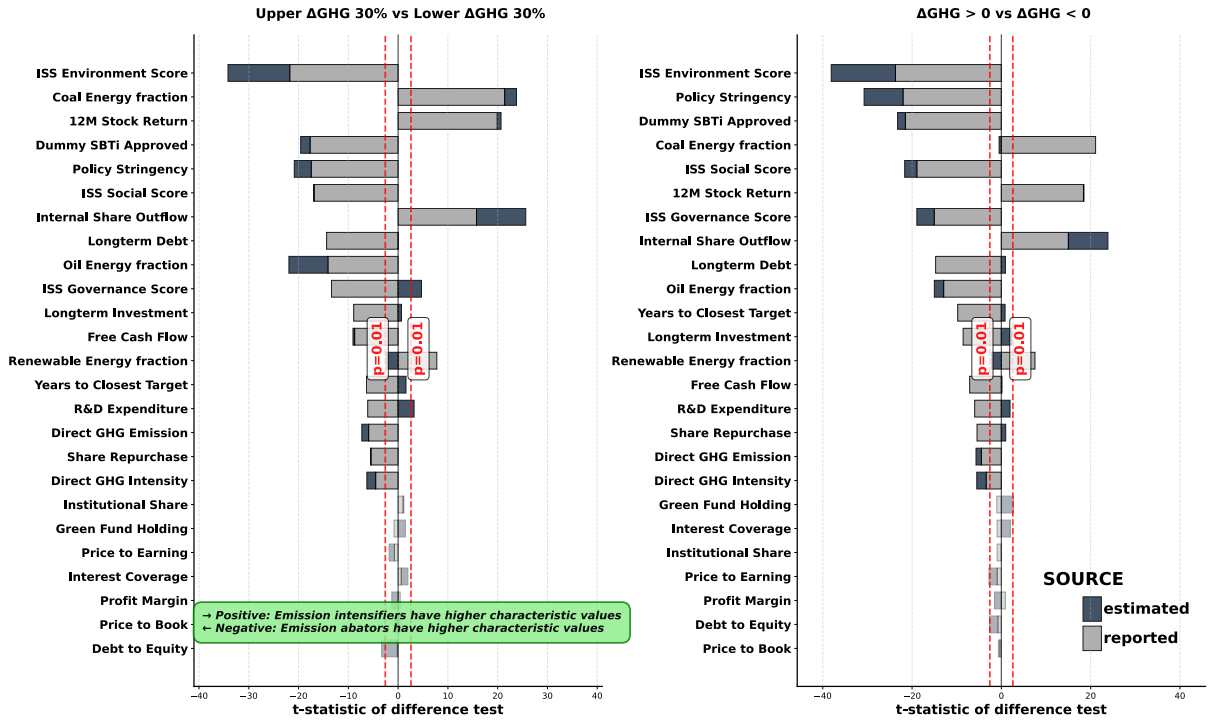


Figure 5: **Characteristics of decarbonizing firms, Pooled sample mean**. This figure displays the absolute values of the t-statistics of the tests in differences in means between characteristics, **pooled by yearly performance** , with high versus low ΔGHG.

3.4 A Causal Analysis - The Impact of SBTi on Corporate Emissions

3.4.1 Identification Strategy

To establish a causal link between Science-Based Targets (SBTi) adoption and emissions, we address the endogeneity of voluntary climate commitments using an instrumental variable (IV) approach within a difference-in-differences (DiD) framework. Our strategy leverages the staggered rollout of the unified CDP-ICLEI regional reporting system, which began in 2018. This reform created an authoritative disclosure channel, enhancing peer pressure on firms. The staggered adoption across different jurisdictions provides quasi-random variation in the timing of this external pressure, which is plausibly uncorrelated with firms' pre-existing emission trends.

The staggered nature of the policy rollout makes traditional DiD estimators biased. We therefore implement the robust estimator from Callaway and Sant'Anna (2021), which is specifically designed for such settings. This method avoids the "bad control" problem by constructing valid counterfactuals for each treatment cohort, using only not-yet-treated or never-treated firms. We define treatment cohorts (g) based on the year (2018, 2019, 2021, or 2023) the CDP-ICLEI system was implemented in a firm's headquarter country. We implement a two-stage procedure to estimate the Local Average Treatment Effect (LATE) for each cohort (g) and time period (t):

First Stage (Cohort-specific):

$$\text{SBTi}_{i,g,t} = \pi_{g,t} \cdot \text{CDP}_{i,g,t} + X'_{i,g,t}\Gamma_{g,t} + \epsilon_{i,g,t} \tag{3}$$

Reduced Form (Cohort-specific):

$$\Delta\text{Emissions}_{i,g,t} = \beta_{g,t} \cdot \text{CDP}_{i,g,t} + X'_{i,g,t}\Theta_{g,t} + u_{i,g,t} \tag{4}$$

where $\text{CDP}_{i,g,t}$ equals 1 for treated firms in cohort g at time $t \geq g$, and 0 for control firms. The cohort-specific Local Average Treatment Effect is $LATE_{g,t} = \beta_{g,t}/\pi_{g,t}$.

3.4.2 Main Results

Our primary IV-DiD results, presented in Table 4, demonstrate a significant causal impact of SBTi adoption on corporate emissions intensity. The first stage (Panel A) confirms the instrument's relevance: exposure to the CDP program increases the probability of SBTi adoption by 3.7 percentage points. The reduced-form estimate (Panel B) shows this program leads to a 0.047 percentage point annual reduction in GHG intensity. Combining these, the Local Average Treatment Effect (LATE, Panel C) reveals our main finding: for firms induced to participate by the program ("compliers"), SBTi adoption causes a 1.264% annual reduction in emissions intensity than its peer. The lack of a significant effect on absolute emissions suggests this decarbonization is driven primarily by enhanced carbon efficiency.

To ensure the credibility of our causal estimates, we perform two key validation tests summarized in Table 5. First, we test the exogeneity of our instrument by examining a placebo test with 18 pre-treatment firm characteristics. The results support our assumption, with 17 of the 18 tests yielding insignificant coefficients. The sole exception was for insider share outflow (Internal Share Outflow), a result likely attributable to high imputation for

Panel & Variable	Estimate	Std. Error	95% Conf. Interval
Panel A: First Stage			
CDP $\rightarrow$ SBTi Adoption (π)	0.037***	0.004	[0.029, 0.045]
Panel B: Reduced Form			
CDP $\rightarrow$ Direct GHG Emissions (%)	0.001	0.008	[-0.015, 0.016]
CDP $\rightarrow$ GHG Intensity (%)	-0.047***	0.008	[-0.063, -0.031]
Panel C: LATE (SBTi $\rightarrow$ Emissions)			
SBTi $\rightarrow$ Direct GHG Emissions (%)	0.013	0.218	[-0.414, 0.441]
SBTi $\rightarrow$ GHG Intensity (%)	-1.264***	0.254	[-1.761, -0.767]

Table 4: **Main IV-DID Results using Aggregated Callaway-Sant'Anna Estimator.** This table reports aggregated treatment effects. The LATE is calculated as the ratio of the reduced-form estimate to the first-stage estimate. Standard errors are clustered at the firm level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Significance stars for panels A and B are based on the 95% confidence interval not covering zero.

non-disclosure information. Second, we test the exclusion restriction using a triple-difference model. The results show that the instrument's effect does not systematically differ between eventual SBTi adopters and non-adopters for either direct emissions (p = 0.586) or GHG intensity (p = 0.062). This supports our assumption that the CDP program influences emissions through the intended SBTi adoption channel. Finally, to visually inspect the parallel trends assumption for our first stage, we plot an event study in Figure 6. The plot confirms the absence of significant pre-treatment effects, further strengthening our identification strategy.

To explore potential heterogeneity across adoption cohorts, we use the Callaway & Sant'Anna estimator aggregated at the cohort level ('type = "group"'). Table 15, in the appendix, presents the results from this cohort-by-cohort analysis. The findings reveal significant heterogeneity but also reinforce our main conclusion: the program's effectiveness is primarily driven by the initial, largest cohort of adopters. Crucially, for all later cohorts, the LATE estimates on both direct GHG emissions and GHG intensity are not statistically significant. The only exception is a marginally significant effect on direct emissions for the small 2021 cohort. This pattern suggests that the causal link between the disclosure program and tangible emission reductions is concentrated in the initial wave of adoption, and the generalized effect is still not revealed for the short, limited observation panel.

The Science-Based Targets Initiative, especially when implemented under quasi-regulatory pressure, effectively reduces uncertainty and indicates authentic alignment with net-zero goals. The notable decrease in emissions intensity among "complier" firms indicates that these commitments result in concrete actions, exceeding limited aspirational pledges or potential greenwashing. Such a framework offers investors a means to identify "validated pledges," which serve as more solid indicators of future efficiency improvements compared to solely voluntary targets, particularly in the context of firm or product-level net-zero commitments aimed at enhancing carbon efficiency. This strategy promotes a sustainable decarbonization pathway at the firm level by integrating efficiency gains into core processes that ultimately influence the carbon footprint of products and services.

Panel A: Pre-treatment Outcomes Test			
Variable Category	**Variables Tested**	**Passed Tests**	**Pass Rate**
Lagged Emissions	2	2	100%
Financial Performance	6	6	100%
Capital Structure	4	4	100%
Market Valuation	2	2	100%
Ownership & Payouts	4	3	75%
Overall	**18**	**17**	**94%**

Panel B: Triple Difference Test			
Outcome Variable	**Triple Interaction**	**P-value**	**Result**
Direct GHG Emissions (%)	0.0017	0.586	Pass
GHG Intensity (%)	0.0058	0.062	Pass

Table 5: **Validity Tests for IV-DID Identification.** Panel A reports results from regressing 18 pre-treatment firm characteristics on the instrumental variable; a test "passes" if the coefficient is statistically insignificant ($p > 0.05$). Panel B reports triple-difference results examining whether the instrument affects eventual SBTi adopters and non-adopters differentially.

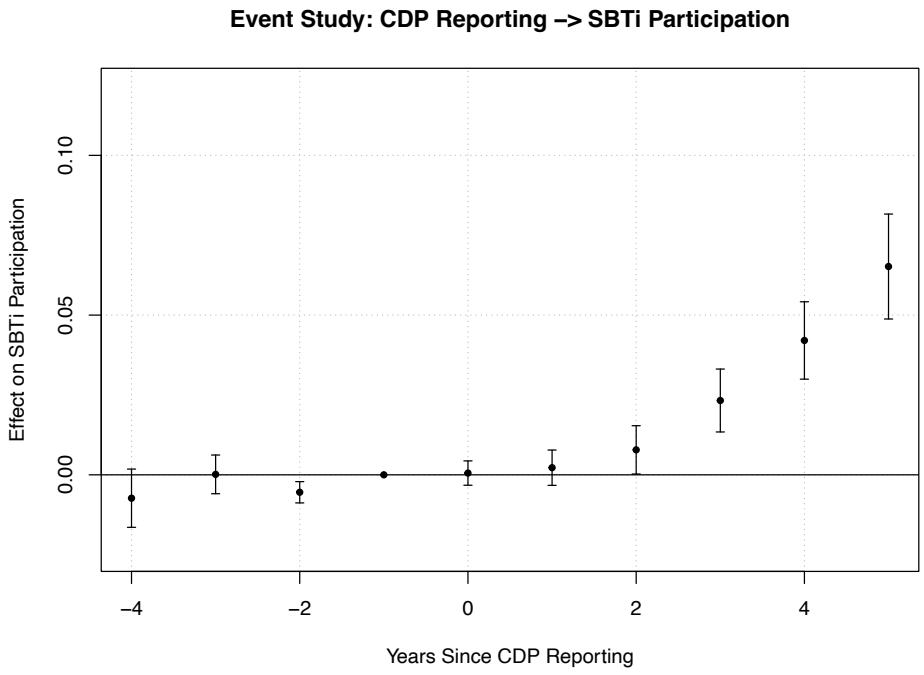


Figure 6: Event study plot of the effect of the CDP regional reporting program on SBTi adoption. The plot displays the dynamic coefficients with 95% confidence intervals. The absence of statistically significant effects in the pre-treatment periods ($t < 0$) supports the parallel trends assumption for the first stage of our IV design.

4 Forecasting

4.1 Out-of-sample analysis

Being able to understand what influences emission dynamics in the cross-section of firms is crucial for investors seeking portfolios decarbonization in accordance with Net-Zero pledges (see, e.g., Bolton et al. (2022) and Le Guenedal et al. (2022)). However, corporate strategies with respect to emissions evolve constantly, both at the micro and the macro level, hence forecasting carbon footprints is hard and besting simple benchmarks is even more challenging.

In this section, we analyze the out-of-sample accuracy of our emission forecasts. Importantly, this section only focuses on **reported** emissions. Due to data restriction, the training set consists of all observations, except from the last year (2023), while the test set encompasses all points from this last year. In terms of algorithms, we stick to random forests, as in the previous section. Formally, the model is

$$y_{t+1,n} = f(\boldsymbol{x}_{t,n}) + e_{t+1,n}, \tag{5}$$

where f is the non-linear model and $e_{t+1,n}$ the residuals. The predictors $\boldsymbol{x}_t$ are the variables used so far in the paper. Note that we use the notation y as is customary in the literature; in our study, it will stand for (log) emissions, (log) intensities, and variations thereof. Importantly, we focus solely on *reported* data. We work with logarithmic values to avoid the squared errors to be heavily biased towards the large emitters only. We also highlight the lag between the dependent variable (at time $t+1$) and the predictors (in t). The metric we will report is the RMSE (root mean squared error), likely the most common choice in regression tasks:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(y_{T+1,n} - \hat{f}(\boldsymbol{x}_{T,n})\right)^2}, \tag{6}$$

where $\hat{f}$ is the trained model, T is the testing date and N the corresponding number of firms.

In order to assess the quality of our forecast, we define two benchmark values. The first benchmark is the RMSE when the prediction is the (log) emission from the previous year, i.e., assuming nothimng has changed from 2021 to 2022. The second benchmark is the error we make when extrapolating the trend of emissions and intensities using a linear model. For this benchmark, we restrict the testing set to the firms for which all points between 2016 and 2022 are available. Indeed, the extrapolation mostly makes sense if a minimum of points are available. To motivate this choice, we show in Figure 7 a few examples of log emission trends for large corporations. These trends show that extrapolating, which seems natural, seems to be a fairly promising approach.

As is customary in machine learning practice, we run the analysis on several values of the important parameters of the model, which we summarize in Table 6 below.

Our main results are gathered in Table 7. The first rows (Panel A) indicate the benchmark errors when predicting the four dependent variables (in columns): GHG, GHG PCT (relative variation in emissions), INT and INT PCT (relative variation in intensities). We report the errors both on the full sample of 2022 and also on the subset of firms for which all points are available ("sub" columns in the table).

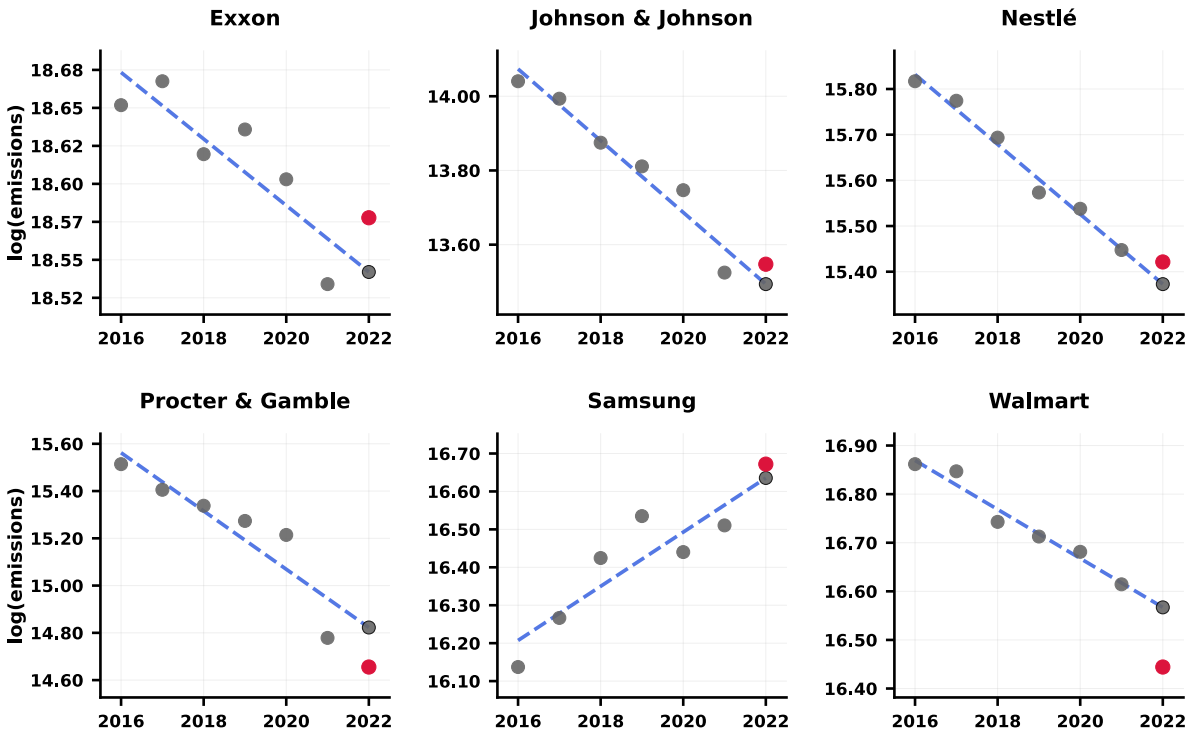


Figure 7: **Extrapolation for large firms**. We show the dynamics of reported log emissions for the six largest firms in the sample (in terms of revenue) for which all seven years of data are available. The first six years are shown in grey and then last point for 2023 is the extrapolation. The realized value for 2023 is shown in red.

short name	description	tested values
ntree	number of trees	150, 500, 1500
mtry	% of columns used to train each tree	0.6, 0.9
sampsize	% of original sample used for each tree	0.6, 0.9
nodesize	max. size of leaves as % of training sample	0.0005, 0.001, 0.005
maxnodes	max. number of leaves per tree	1000, 2000, 5000

Table 6: **Hyper-parameters**. We list the hyper-parameters we used in our first baseline attempt, along with the tested values.

An interesting takeaway from these preliminary results is that extrapolation is not the best option. This can be surprising, given the seemingly promising examples shown in Figure 7, but in fact, while the good performance of extrapolation holds for large firms, it is much more erratic for smaller firms.Borui: I didn't update this, THE ONLY past result In Figure 12 in the Appendix, we provide a few examples for which firms experience sharp shocks in their emission trajectory in 2022, leading extrapolation to fail badly. This may explain why the simpler prediction that assumes constant emissions can work better, as it is less exposed to contrarian moves.

dependent variable	**GHG**		**GHG PCT**		**INT**		**INT PCT**	
sample	all	sub	all	sub	all	sub	all	sub
PANEL A: benchmarks								
constant	0.891	**0.205**	0.304	0.231	0.916	0.246	0.332	0.271
extrapolation	-	0.274	-	-	-	0.303	-	-
PANEL B: random forests (statistics across HP combinations)								
min	**0.886**	0.211	**0.297**	**0.229**	**0.900**	0.250	**0.327**	**0.267**
median	0.889	0.223	0.297	0.229	0.905	0.257	0.328	0.268
max	0.927	0.308	0.298	0.231	0.939	0.311	0.328	0.269
PANEL C: learning from errors (statistics across HP combinations)								
min	0.892	0.207	-	-	0.905	**0.243**	-	-
median	0.893	0.210	-	-	0.906	0.247	-	-
max	0.894	0.214	-	-	0.906	0.249	-	-

Table 7: **Out-of-sample forecasting accuracy**. We report the root mean squared errors (RMSE) of the predictive models for the year 2023. In Panel A, we report two benchmarks: the constant value (from 2022) and the extrapolated one, from 2016 to 2022. For the extrapolation, the estimations are run on the set of firms for which all eight data points of reported emissions are available (we refer to this set as "*sub*"). In Panel B, we report the statistics (minimum, median and maximum) of the RMSE when spanning the hyperparameter space with 108 combinations. Panel C shows results from learning from errors approach. Training is performed on 2016-2022 data with out-of-sample predictions on 2023 data.

The next batch of rows (Panel B in Table 7) provides the minimum, median and maximum of the RMSE across the 108 hyperparameter combinations of Table 6. Plainly, they show that the sophisticated algorithms do not perform better than the heuristic benchmarks, except perhaps for the relative change in intensities (INT PCT).

To try to improve on these disappointing results, we tested another approach. Given the relative good performance of the constant benchmark, we sought to predict the drivers of the raw variations. In doing so, we hope to reduce the error from the simplest benchmark. In this case, the dependent variable in the models becomes

$$\Delta_{i,t} = \begin{cases} \log(\text{GHG}_{i,t}) - \log(\text{GHG}_{i,t-1}) & \text{for emissions} \\ \log(\text{INT}_{i,t}) - \log(\text{INT}_{i,t-1}) & \text{for intensities.} \end{cases} \tag{7}$$

Upon calibration of the model, the estimates for log emissions and log intensities are then

$$\log(\text{GHG}_{i,t-1}) + \hat{\Delta}_{i,t}, \quad \log(\text{INT}_{i,t-1}) + \hat{\Delta}_{i,t},$$

where $\hat{\Delta}_{i,t}$ is the value predicted from the model (either for emissions or for intensities). In particular, this is suited to raw numbers, but not to the GHG and INT percent changes. The corresponding errors are gathered in Panel C of Table 7. They are all smaller than in Panel B, suggesting that this approach is better for forecasting purposes, compared to the brute RF models. Nevertheless, they mostly do not beat the benchmarks, except for one exception.

4.2 Classification

The results in the previous section are disappointing both because the sophistication in models does not enhance performance, but also because performance itself is not easy to interpret - this is a common drawback of the RMSE.

To bypass these limitations, we propose to forecast decarbonization as a binary (dummy) variable. A firm is termed "*decarbonizing*" at year t if its emissions in t are lower than in year $t-1$. We will use the same designation for intensity reduction. The corresponding value for the dependent variable is one and the only other possible value is zero if the firm is projected to increase its emissions in 2022. We consider three benchmarks:

1. **constant status**: a firm is forecasted to decarbonize in 2022 if it had reduced its footprint in 2021 (compared to 2020).
2. **regression forecast**: we use the linear trend from 2016 to 2021 to forecast a value in 2022. If it is lower than that in 2021, then we predict a decarbonization.
3. **regression slope**: given the above trend, we predict lower emissions if the slope is negative.

The forecasting accuracy of these benchmarks is reported in Panel A in Table **??**. The accuracy is simply the proportion of correct predictions. When the samples are perfectly balanced, a natural yardstick is 50%, which corresponds to a coin toss.

We do obtain metrics above 50%, especially in Panel C when predicting GHG emissions. The models have a much harder time forecasting intensities, for reasons that remain unclear. Perhaps one potential explanation comes from the plot next to Table **??**. It shows that the aggregate proportion of firms that decarbonize is much more stable for raw emissions compared to intensities. In particular, the proportion of firms reducing their emissions is always above 50%, which boosts out-of-sample performance.

4.3 Sector discrepancies

Finally, we dive into the dependence of our results to sectors. In Figure 8, we depict the RMSE, but this time averaged for each sector separately. This reveals pronounced discrepancies, with utilities being the industry that appears the easiest to forecast, while information technology is the hardest.

Some sectors are harder to abate and also more material when it comes to GHG emissions. This is the case of Industrials, Real Estate and Energy for instance. The former are among the hardest sectors to forecast, while for the other two, our models fare better.

5 Conclusion

This paper seeks to leverage data mining to understand the characteristics of decarbonizing firms. First, there are large discrepancies between reported and estimated emissions. The former, calculated by data providers are much based, according to our results, on accounting data, and they counter-intuitively give little weight to pledges such as those recorded by the SBTI.

dependent variable	$1_{\Delta GHG<0}$		$1_{\Delta INT<0}$	
sample	all	sub	all	sub
PANEL A: Benchmarks				
constant	0.544	0.553	0.534	0.545
linear prediction	-	0.556	-	0.546
slope sign	-	0.543	-	0.529
PANEL B: Random Forest predictions (regression)				
min	0.525	0.519	0.494	0.482
median	0.535	0.537	0.511	0.509
max	0.540	0.550	0.527	0.534
PANEL C: Random Forest predictions (classification)				
min	0.582	0.633	0.533	0.556
median	**0.588**	**0.639**	**0.551**	**0.573**
max	0.592	0.645	0.562	0.585

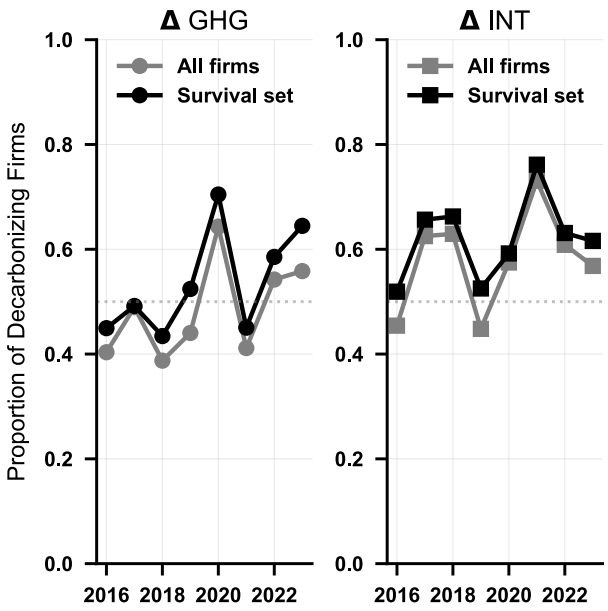


Table 8: **Classification of decarbonization**. In the left table, we report the accuracy of the predictive models for the year 2023. In Panel A, we report three benchmarks: i) the decarbonization status of 2022 is assumed to remain, ii) the linear prediction is used to determine if emissions or intensities will decrease and iii) the slope of the linear regression is used (a negative slope meaning decarbonization). For the latter two, the estimations are run on the set of firms for which all eight data points of reported emissions are available (we refer to this set as "*sub*"). In Panel B, we report the statistics (minimum, median and maximum) of the accuracy when spanning the hyperparameter space with 108 hyperparameter combinations. In this case, the RF models are used to generate predictions for 2023 and decarbonization is inferred from the realized 2022 emissions and intensities. In Panel C, it is the binary variable of decarbonization that is directly forecasted. In the right plot, we depict the evolution of the proportion of decarbonizing firms, both with respect to GHG and INT, and also discriminating between reported (black) and estimated (grey) figures.

A second important takeaway is that sophisticated nonlinear models based on panel data do not outperform simple methods and that extrapolation is not the best benchmark due to potential reversals is corporate emission policies.

With regards to the important variables for decarbonization, it is hard to tell a compelling story. Sustainability-linked features, such as biodiversity footprint matter, but so do other more mundane financial indicators, such as past returns and the book-to-market ratio. Analysts' sentiment and macroeconomic indicators are, too, often relevant in explaining carbon trajectories.

Our analysis on classification further shows that aggregate trends have an impact on model accuracy. Global proportions of decarbonizing firms strongly relates with the performance of the models we used.

Finally, our study reveals the differences of errors across industries. Some industries (e.g., utilities) are easier to predict than others (information technology). With regards to sectors with high carbon materiality, e.g., industrials, materials, energy and real estates, the situation

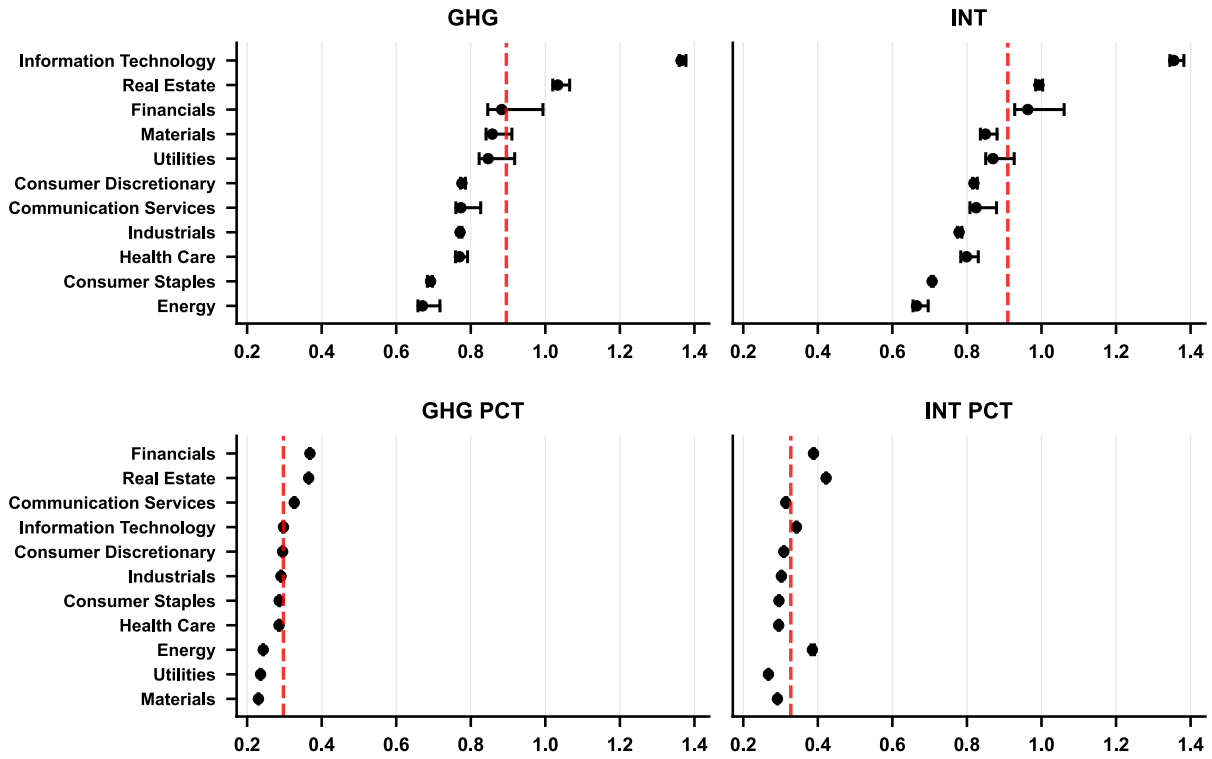


Figure 8: **Sector variations**. This plot shows the RMSE of the predictive models from Table 7 (Panel B) aggregated at the sector level. The error bars mark the variations across the hyper-parameters. The dashed red lines shows the RMSE over all sectors.

is not clear cut. They belong to the set of industries for which emissions are neither very hard nor simple to forecast.

6 Appendix

6.1 Variable definitions

Parameter	Source	Description
I. GHG Emissions		
ISS_GHG_S1	**ISS ESG**	Scope 1 GHG emissions, expressed in tons equivalent CO2 (tCO2e)
ISS_GHG_S2	**ISS ESG**	Scope 2 GHG emissions, expressed in tons equivalent CO2 (tCO2e)
ISS_GHG_S3	**ISS ESG**	Scope 3 (upstream & downstream) GHG emissions, expressed in tons equivalent CO2 (tCO2e)
MSCI_GHG_S1	**MSCI**	Scope 1 GHG emissions, expressed in tons equivalent CO2 (tCO2e)
MSCI_GHG_S2	**MSCI**	Scope 2 GHG emissions, expressed in tons equivalent CO2 (tCO2e)
MSCI_GHG_S3	**MSCI**	Scope 3 (upstream & downstream) GHG emissions, expressed in tons equivalent CO2 (tCO2e)

Table 9: **Description of dependant variables**

Parameter	Source	Description
I. Sustainability Indicators, previous period		
ISS_Environment_Score	**ISS ESG**	ISS Environmental Score
ISS_Social_Score	**ISS ESG**	ISS Social Score
ISS_Governance_Score	**ISS ESG**	ISS Governance Score
Biodiversity_GHG_Score	**Iceberg DataLab**	Corporate Biodiversity Footprint (CBF), expressed in $km2.MSA$
Dummy_SBTi_Approved	**MSCI**	Corporate voluntary emissions targets indicator
Years_to_Closest_Target	**MSCI**	Years to the closest emissions target
II. Fundamental & Market Indicators		
12M_Stock_Return	**Bloomberg**	Average annualised day-to-day return net of dividend
Revenue	**Bloomberg**	Sales revenue turnover, yearly fundamental
Price_to_Earning	**Bloomberg**	Price to earning ratio, calendar year average
Price_to_Book	**Bloomberg**	Price to book ratio, calendar year average
Debt_to_Equity	**Bloomberg**	Total debt to total equity ratio, yearly fundamental

Continued on next page

(Continued)

Parameter	Source	Description
Longterm_Debt	**Bloomberg**	Long term debt on the balance sheet, yearly fundamental
Profit_Margin	**Bloomberg**	Profit margin ratio, yearly fundamental
Interest_Coverage	**Bloomberg**	Interest coverage ratio, yearly fundamental
R&D_Expenditure	**Bloomberg**	Research and development expenditure, income statement yearly fundamental
Free_Cash_Flow	**Bloomberg**	Free cash flow, yearly fundamental
Longterm_Investment	**Bloomberg**	Long term investment, yearly fundamental
III. Ownership & Debt Structure		
Internal_Share_Outflow	**Bloomberg**	P85ercentage of insider shares outflow at the end the reporting period
Institutional_Share	**Bloomberg**	Percentage of float shares held by institutions, calendar year average
Green_Fund_Holding	**Exane, Factset**	Percentage of shares held by *"Active"* sustainable ESG funds as identified by Exane (2023).
Share_Repurchase	**Bloomberg**	Total value of shares repurchased, yearly fundamental
IV. Regional Indicators		
Policy_Stringency	**OECD**	OECD Cross sectional Environmental policy Stringency, No US information, considered 0
Coal_Energy_pct	**IEA**	Coal Energy Country Level Energy Consumption Ratio from IEA yearly
Oil_Energy_pct	**IEA**	Oil Energy Country Level Energy Consumption Ratio from IEA yearly
Renewable_Energy_pct	**IEA**	Renewables Energy Country Level Energy Consumption Ratio from IEA yearly

Table 10: **Description of explanatory variables**

6.2 Descriptive statistics

	type	N	min	25%	median	mean	75%	max	std. dev.
GHG	estimated	83833	-0.63	7.63	8.97	9.05	10.36	19.14	2.16
GHG	reported	53708	-0.28	9.64	11.14	11.31	12.84	19.82	2.50
INT	estimated	83833	-5.07	2.89	3.87	4.00	4.89	15.29	1.59
INT	reported	53708	-7.26	3.28	4.51	4.58	5.79	12.03	2.02
ΔGHG	estimated	83833	-0.50	-0.08	0.02	0.08	0.14	2.00	0.31
ΔGHG	reported	53708	-0.50	-0.08	0.01	0.07	0.13	2.00	0.32
ΔINT	estimated	83833	-0.50	-0.10	-0.01	0.04	0.09	2.00	0.30
ΔINT	reported	53708	-0.50	-0.13	-0.02	0.03	0.10	2.00	0.32

Table 11: **Descriptive statistics of dependent variables for Scope 1+2** We report the baseline indicators for our four dependent variables. Emissions are initially in tons equivalent CO_2, and we then take the log. Variations in emissions and intensities ($\Delta GHG_{t,i}$ and $\Delta INT_{t,i}$) are trimmed below 50% and above 200%. Emissions are reported after the log. Intensities are computed as raw emissions divided by Revenue in USD.

.

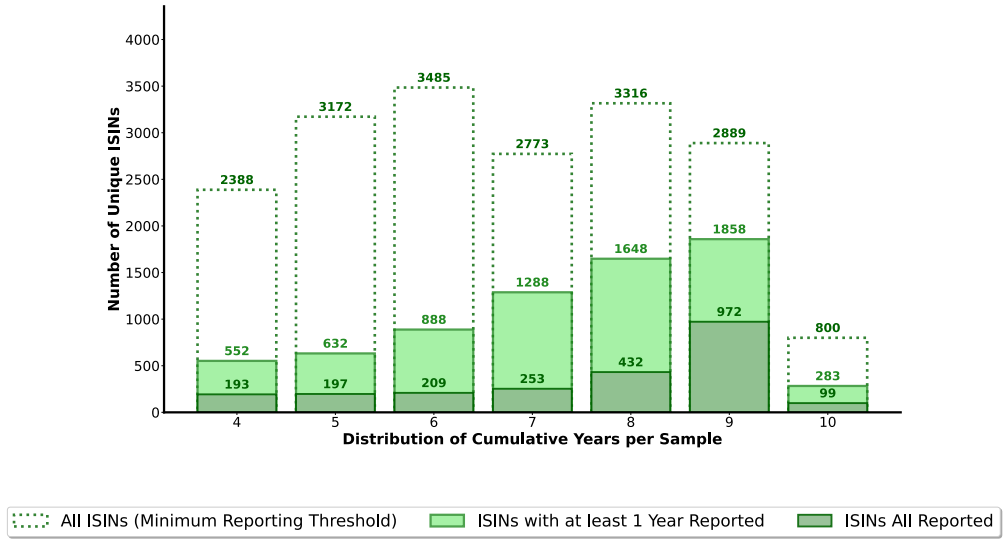


Figure 9: **Distribution of cumulative reported emissions over the period**

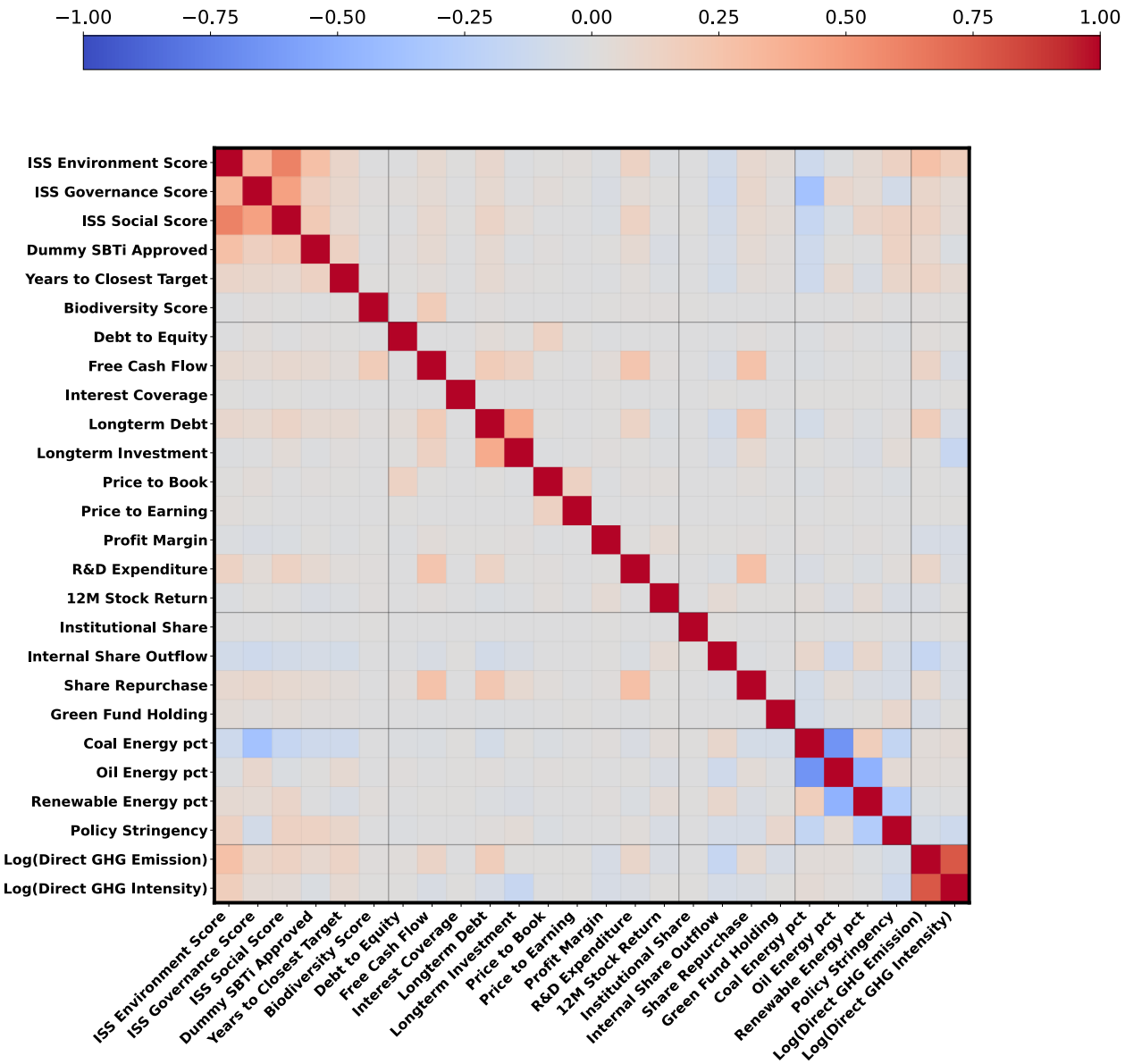


Figure 10: **Correlation matrix of all initial variables**.

Table 12: **Descriptive Statistics of Key Indicators.** We report summary descriptive statistics of independent variables. We also highlight the if significant sample mean t-test results of the modeled group and the reported group. light green for 5% level, dark green for 1% level.

Panel A: Reported

Parameter	count	mean	std	min	1%	50%	99%	max
I. Sustainability Indicators								
ISS Environment Score	53708	1.74***	0.38	1.00	1.10	1.70	2.87	3.56
ISS Governance Score	53708	2.27***	0.48	1.00	1.15	2.25	3.33	3.82
ISS Social Score	53708	1.81***	0.33	1.02	1.16	1.76	2.82	3.32
Dummy SBTi Approved	53708	0.04***	0.21	0.00	0.00	0.00	1.00	1.00
Years to Closest Target	53708	1.57***	3.80	0.00	0.00	0.00	18.00	48.00
Biodiversity Score	53708	−441.43	2.44×10^5	-5.32×10^6	-2.62×10^4	−755.92	−0.04	3.95×10^7
II. Fundamental & Market Indicators								
Debt to Equity	53708	145.31 ***	1.47×10^3	0.00	0.00	55.39	1.24×10^3	2.69×10^5
Free Cash Flow	53708	170.38 ***	1.18×10^3	-1.90×10^4	−940.24	37.50	3.14×10^3	1.27×10^5
Interest Coverage	53708	1.19×10^3	1.54×10^5	-4.70×10^5	−99.00	11.18	3.87×10^3	3.55×10^7
Longterm Debt	53708	$3.60 \times 10^{3***}$	1.47×10^4	−0.07	0.00	403.05	5.79×10^4	3.62×10^5
Longterm Investment	53708	$4.83 \times 10^{3***}$	3.94×10^4	0.00	0.00	21.17	1.13×10^5	1.24×10^6
Price to Book	53708	5.48	64.19	0.04	0.26	1.68	39.04	5.84×10^3
Price to Earning	53708	260.76 ***	2.37×10^4	0.10	2.70	17.62	433.86	3.42×10^6
Profit Margin	53708	9.29***	61.42	-5.39×10^3	−59.51	6.80	101.24	1.99×10^3
R&D Expenditure	53708	44.71***	364.38	0.00	0.00	0.00	841.80	1.99×10^4
12M Stock Return	53708	11.70**	47.58	−91.93	−61.36	4.73	182.19	1.84×10^3
III. Ownership & Debt Structure								
Institutional Share	53708	7.88×10^5	1.72×10^8	0.00	0.00	51.58	141.57	3.97×10^{10}
Internal Share Outflow	53708	4.85***	11.15	0.00	0.00	0.47	57.97	110.59
Share Repurchase	53708	39.14***	422.34	−859.45	0.00	0.00	831.50	3.14×10^4
Green Fund Holding	53708	0.14***	0.86	0.00	0.00	0.00	2.08	62.91
IV. Regional Indicators								
Coal Energy pct	53708	0.08***	0.08	0.00	5.01×10^{-3}	0.06	0.39	0.42
Oil Energy pct	53708	0.44***	0.08	0.16	0.23	0.45	0.62	0.75
Renewable Energy pct	53708	0.08***	0.07	0.00	0.02	0.06	0.29	0.69
Policy Stringency	53708	2.26***	2.21	0.00	0.00	2.13	7.30	7.42

Continued on next page

(Continued)

Parameter	count	mean	std	min	1%	50%	99%	max
Panel B: Modeled								
I. Sustainability Indicators								
ISS Environment Score	83833	1.60***	0.26	1.00	1.07	1.63	2.38	3.48
ISS Governance Score	83833	2.15***	0.28	1.00	1.25	2.16	2.94	3.57
ISS Social Score	83833	1.71***	0.23	1.00	1.10	1.69	2.33	3.22
Dummy SBTi Approved	83833	1.30×10^{-4}***	0.01	0.00	0.00	0.00	0.00	1.00
Years to Closest Target	83833	0.10***	1.30	0.00	0.00	0.00	1.00	39.00
Biodiversity Score	83833	-1.72×10^{3}	1.05×10^{4}	-1.44×10^{5}	-1.17×10^{4}	-850.70	-3.24	4.90×10^{5}
II. Fundamental & Market Indicators								
Debt to Equity	83833	117.55 ***	1.02×10^{3}	0.00	0.00	36.48	1.02×10^{3}	9.18×10^{4}
Free Cash Flow	83833	16.75***	266.34	-1.92×10^{4}	-354.19	3.88	438.99	3.58×10^{4}
Interest Coverage	83833	788.32	3.67×10^{4}	-2.26×10^{6}	-518.60	7.60	9.64×10^{3}	6.49×10^{6}
Longterm Debt	83833	436.40 ***	1.67×10^{4}	0.00	0.00	12.26	3.90×10^{3}	2.34×10^{6}
Longterm Investment	83833	713.64 ***	2.99×10^{4}	0.00	0.00	3.12	7.29×10^{3}	2.96×10^{6}
Price to Book	83833	5.33	85.65	0.02	0.20	1.47	31.46	5.28×10^{3}
Price to Earning	83833	2.16×10^{3}***	7.69×10^{4}	0.02	1.91	19.65	1.13×10^{3}	1.02×10^{7}
Profit Margin	83833	-19.43 ***	2.00×10^{3}	-4.02×10^{5}	-196.72	4.52	87.65	1.16×10^{4}
R&D Expenditure	83833	6.92***	52.85	-2.67	0.00	0.00	98.10	5.92×10^{3}
12M Stock Return	83833	21.50**	961.59	-98.33	-68.21	1.94	237.22	1.05×10^{5}
III. Ownership & Debt Structure								
Institutional Share	83833	5.67×10^{4}	1.10×10^{7}	0.00	0.00	16.22	132.33	2.60×10^{9}
Internal Share Outflow	83833	10.78***	16.13	0.00	0.00	2.95	68.66	127.02
Share Repurchase	83833	2.05***	47.31	-32.64	0.00	0.00	34.23	7.23×10^{3}
Green Fund Holding	83833	0.03***	0.53	0.00	0.00	0.00	0.62	63.14
IV. Regional Indicators								
Coal Energy pct	83833	0.11***	0.11	0.00	1.53×10^{-3}	0.07	0.41	0.42
Oil Energy pct	83833	0.42***	0.10	0.16	0.23	0.43	0.65	0.75
Renewable Energy pct	83833	0.07***	0.06	0.00	0.02	0.06	0.29	0.73
Policy Stringency	83833	1.93***	1.81	0.00	0.00	2.04	6.73	7.42

6.3 OLS, with time, region and industry group fixed-effects

	GHG (log)		INT (log)		ΔGHG		ΔINT	
	Reported	Modeled	Reported	Modeled	Reported	Modeled	Reported	Modeled
Direct_GHG_Emission	463.722	710.684			−21.273	−21.211		
Direct_GHG_Intensity			314.076	424.625			−16.812	−23.272
ISS_Environment_Score	1.571	−4.477	0.314	−2.723	−3.576	0.322	−2.872	2.978
ISS_Governance_Score	−2.690	5.906	−1.525	5.220	−2.696	2.623	−0.409	−0.205
ISS_Social_Score	1.653	−2.055	1.063	−1.224	−1.519	−2.825	−2.205	−1.425
Dummy_SBTi_Approved	3.718	−0.888	2.206	−1.012	−9.156	−1.102	−9.089	−2.225
Years_to_Closest_Target	1.928	1.469	1.113	0.332	1.907	2.485	0.812	−0.615
Biodiversity_Score	0.622	2.825	−0.025	−3.130	−1.062	1.775	−1.161	−6.040
Debt_to_Equity	0.957	−0.962	0.749	1.286	0.342	−0.741	−1.848	1.276
Free_Cash_Flow	1.610	0.686	−1.042	0.752	−0.867	−0.700	−0.399	−0.420
Interest_Coverage	−2.093	0.887	1.905	−1.422	−5.536	0.794	−2.624	−1.694
Longterm_Debt	5.844	−1.573	2.525	−1.749	−0.508	0.793	−2.739	−1.172
Longterm_Investment	2.074	1.753	0.777	0.399	−1.168	0.527	0.747	−0.376
Price_to_Book	−1.477	0.045	−2.155	−0.179	−0.296	0.116	−4.824	−0.761
Price_to_Earning	−1.641	−0.465	−0.680	2.349	−0.596	−0.844	−1.931	0.800
Profit_Margin	−2.038	−6.159	0.502	3.744	−0.628	0.240	−0.861	−0.851
R&D_Expenditure	4.709	1.944	−2.318	−0.597	0.310	1.565	−4.761	−3.084
12M_Stock_Return	1.613	1.395	−7.369	−0.136	11.076	0.887	−6.214	−1.451
Institutional_Share	−0.626	−0.185	8.865	0.175	−2.321	−0.721	1.953	−0.431
Internal_Share_Outflow	−0.689	−0.572	−1.769	−2.856	3.227	4.922	−0.020	−1.088
Share_Repurchase	−1.579	0.475	−1.475	−1.952	−1.264	−1.124	0.558	−1.343
Green_Fund_Holding	−0.974	0.112	1.084	0.932	−2.402	1.542	−0.384	0.156
Coal_Energy_pct	−5.681	−6.988	−9.357	−15.797	−1.128	4.900	−4.978	−3.350
Oil_Energy_pct	−5.845	−8.230	−7.956	−12.752	−1.432	−0.843	−4.966	−5.072
Renewable_Energy_pct	−2.043	−0.502	−5.119	−6.434	−1.392	2.145	−7.138	−6.768
Policy_Stringency	2.844	2.236	4.244	6.537	−3.994	0.870	−3.293	5.212

Table 13: **Panel Regression Results - Time + Region + Industry Fixed Effects**. We report the *t*-statistics for models with time, region, and industry fixed effects - all independent variables are lagged. The regressions control for time effects (year), region effects (country), and industry effects (GICS sector). Standard errors are robust to heteroskedasticity. Colors indicate significance levels: light green for 1% level, dark green for 0.1% level.

6.4 LASSO surviving rates without fixed effects

		Raw footprint				Relative change			
		GHG (log)		INT (log)		ΔGHG		ΔINT	
rank		rep.	est.	rep.	est.	rep.	est.	rep.	est.
1	Direct_GHG_Emission	**99**	**98**			**98**	**99**		
2	Policy_Stringency	36	15	42	21	**96**	85	**97**	62
3	Dummy_SBTi_Approved	9	0	22	6	**94**	54	**98**	57
4	Coal_Energy_pct	39	34	42	0	**97**	47	**92**	76
5	12M_Stock_Reutrn	9	0	35	0	**96**	35	**91**	43
6	Direct_GHG_Intensity			**99**	**98**			**92**	**96**
7	Longterm_Debt	0	15	24	25	83	45	89	40
8	ISS_Governance_Score	10	4	28	0	67	38	**96**	72
9	GDP_Growth	0	0	18	28	52	**92**	**98**	**99**
10	Internal_Share_Outflow	4	0	0	0	88	85	59	25
11	Biodiversity_GHG_Score	19	0	30	30	64	70	83	82
12	Longterm_Investment	7	0	25	0	79	27	65	53
13	Oil_Energy_pct	9	30	24	22	77	**95**	64	38
14	Years_to_Closest_Target	0	0	22	4	56	65	82	61
15	ISS_Environment_Score	13	2	25	33	**94**	80	39	81

Table 14: **Lasso survival rate without TWFE**. We report the percentage of times that a given variable survives LASSO selection. The rank of the variable is determined by lag variables first, then by the average of ΔGHG reported and ΔINT reported columns. No fixed effects are included.

6.5 Characteristics of decarbonizing firms intensities

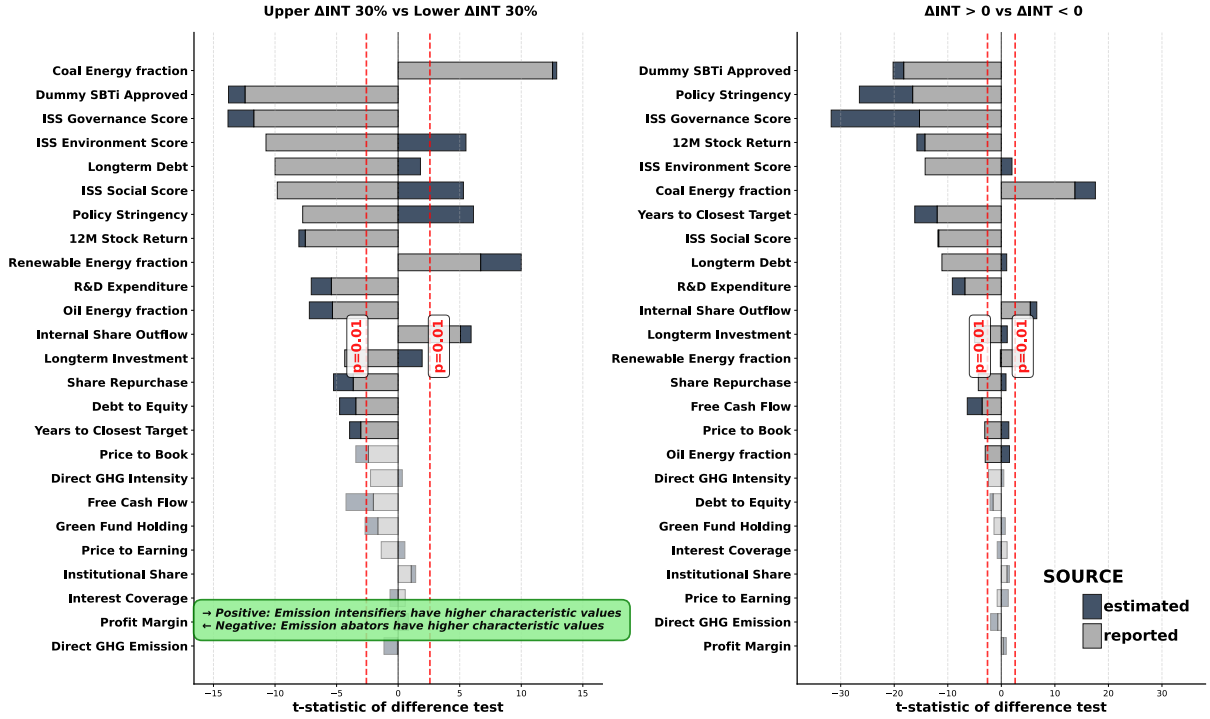


Figure 11: **Characteristics of (intensity) decarbonizing firms, Pooled**. We plot the absolute values of the t-statistics of the tests in differences in means between characteristics, **pooled by yearly performance** , with high versus low ΔINT.

	First Stage	LATE Estimates	
Cohort	**π (CDP $\rightarrow$ SBTi)**	**Direct GHG (%)**	**Intensity (%)**
2018	0.042***	-0.039	-1.309***
	(0.004)	(0.191)	(0.222)
2019	-0.015**	-1.524	-2.737
	(0.006)	(1.860)	(2.057)
2021	0.460**	-0.272*	0.047
	(0.185)	(0.140)	(0.082)
2023	-0.048***	-3.357	0.196
	(0.004)	(3.654)	(2.482)

Table 15: **Heterogeneity Analysis by Adoption Cohort using C&S Estimator.** This table reports cohort-specific IV-DID estimates using the Callaway & Sant'Anna (2021) method aggregated by group. Standard errors, clustered at the firm level, are in parentheses. The LATE is calculated as the ratio of the group-specific reduced-form estimate to the group-specific first-stage estimate. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

6.6 Large errors from extrapolation

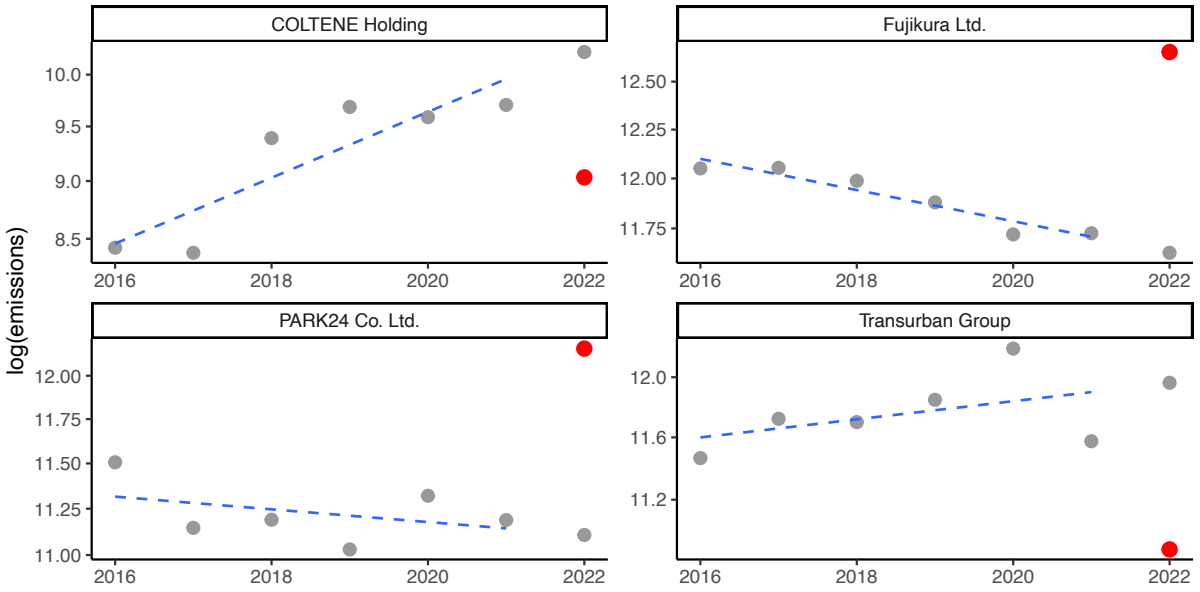


Figure 12: **Large errors from extrapolation**. We show the dynamics of log emissions for a sample of firms which have experience sharp breaks in 2022. The first six years are shown in grey and the last point is the extrapolation. The realized value for 2022 is shown in red.

References

Acharya, V. V., R. Engle, and O. Wang (2023). Large firms, common ownership, and incentives to decarbonize and innovate. *Working Paper*.

Adamolekun, G. (2024). Carbon price and firm greenhouse gas emissions. *Journal of Environmental Management 349*, 119496.

Aldy, J. E., P. Bolton, Z. Halem, and M. T. Kacperczyk (2023). Behind schedule: The corporate effort to fulfill climate obligations. *Journal of Applied Corporate Finance, Forthcoming*.

Atta-Darkua, V., S. Glossner, P. Krueger, and P. Matos (2023). Decarbonizing institutional investor portfolios: Helping to green the planet or just greening your portfolio? *SSRN Working Paper 4212568*.

Barahhou, I., M. Ben Slimane, T. Roncalli, and N. Oulid Azouz (2022). Net zero investment portfolios-part 1. The comprehensive integrated approach. *SSRN Working Paper 4283998*.

Beck, N. and J. N. Katz (1995). What to do (and not to do) with time-series cross-section data. *American Political Science Review 89*(3), 634–647.

Berg, F., J. F. Koelbel, and R. Rigobon (2022). Aggregate confusion: The divergence of esg ratings. *Review of Finance 26*(6), 1315–1344.

Berg, F., J. Oliver Huidobro, and R. Rigobon (2024). On the importance of assurance in carbon accounting. *SSRN Working Paper 4734240*.

Berk, J. and J. H. Van Binsbergen (2024). The impact of impact investing. *SSRN Working Paper 3909166*.

Bolton, P. and M. Kacperczyk (2021). Do investors care about carbon risk? *Journal of Financial Economics 142*(2), 517–549.

Bolton, P. and M. Kacperczyk (2023). Firm commitments. Technical report, National Bureau of Economic Research.

Bolton, P., M. Kacperczyk, and F. Samama (2022). Net-zero carbon portfolio alignment. *Financial Analysts Journal 78*(2), 19–33.

Busch, T., M. Johnson, and T. Pioch (2022). Corporate carbon performance data: Quo vadis? *Journal of Industrial Ecology 26*(1), 350–363.

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics 225*(2), 200–230. Themed Issue: Treatment Effect 1.

Cenedese, G., S. Han, and M. T. Kacperczyk (2023). Carbon-transition risk and net-zero portfolios. *SSRN Working Paper 4565220*.

Chen, A. Y. and J. McCoy (2024). Missing values and the dimensionality of expected returns. *Journal of Financial Economics Forthcoming*.

Cohen, S., I. Kadach, and G. Ormazabal (2023). Institutional investors, climate disclosure, and carbon emissions. *Journal of Accounting and Economics 76*(2-3), 101640.

Coskuner, C., M. K. Paskeh, G. Olasehinde-Williams, and S. S. Akadiri (2020). Economic and social determinants of carbon emissions: evidence from organization of petroleum exporting countries. *Journal of Public Affairs 20*(3), e2092.

Dimson, E., P. Marsh, and M. Staunton (2020). Divergent esg ratings.

Dogan, E. and F. Seker (2016a). Determinants of co2 emissions in the European Union: the role of renewable and non-renewable energy. *Renewable energy 94*, 429–439.

Dogan, E. and F. Seker (2016b). An investigation on the determinants of carbon emissions for OECD countries: empirical evidence from panel models robust to heterogeneity and cross-sectional dependence. *Environmental Science and Pollution Research 23*, 14646–14655.

Elsayih, J., R. Datt, and Q. Tang (2021). Corporate governance and carbon emissions performance: empirical evidence from Australia. *Australasian Journal of Environmental Management 28*(4), 433–459.

Feldhütter, P. and L. H. Pedersen (2023). Is capital structure irrelevant with ESG investors? *SSRN Working Paper*.

Flammer, C. (2021). Corporate green bonds. *Journal of financial economics 142*(2), 499–516.

Forster, P. M., C. J. Smith, T. Walsh, W. F. Lamb, R. Lamboll, M. Hauser, A. Ribes, D. Rosen, N. Gillett, M. D. Palmer, et al. (2023). Indicators of global climate change 2022: annual update of large-scale indicators of the state of the climate system and human influence. *Earth System Science Data 15*(6), 2295–2327.

Fraser, A. and T. Fiedler (2023). Net-zero targets for investment portfolios: An analysis of financed emissions metrics. *Energy Economics 126*, 106917.

Goldhammer, B., C. Busse, and T. Busch (2017). Estimating corporate carbon footprints with externally available data. *Journal of Industrial Ecology 21*(5), 1165–1179.

Grinsztajn, L., E. Oyallon, and G. Varoquaux (2022). Why do tree-based models still outperform deep learning on typical tabular data? *Advances in Neural Information Processing Systems 35*, 507–520.

Habiba, U., C. Xinbang, and A. Anwar (2022). Do green technology innovations, financial development, and renewable energy use help to curb carbon emissions? *Renewable Energy 193*, 1082–1093.

Haque, F. (2017). The effects of board characteristics and sustainable compensation policy on carbon performance of UK firms. *British Accounting Review 49*(3), 347–364.

Heath, D., D. Macciocchi, R. Michaely, and M. C. Ringgenberg (2023). Does socially responsible investing change firm behavior? *Review of Finance 27*(6), 2057–2083.

Hwang, C. Y., S. Titman, and Y. Wang (2022). Investor tastes, corporate behavior, and stock returns: An analysis of corporate social responsibility. *Management Science 68*(10), 7131–7152.

Ishwaran, H. and M. Lu (2019). Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. *Statistics in mMdicine 38*(4), 558–582.

Januschowski, T., Y. Wang, K. Torkkola, T. Erkkilä, H. Hasson, and J. Gasthaus (2022). Forecasting with trees. *International Journal of Forecasting 38*(4), 1473–1481.

Jiang, X. and D. Guan (2016). Determinants of global CO2 emissions growth. *Applied energy 184*, 1132–1141.

Kahn, M. E., J. G. Matsusaka, and C. Shu (2023). Divestment and engagement: The effect of green investors on corporate carbon emissions. *SSRN Working Paper 4609861*.

Kalesnik, V., M. Wilkens, and J. Zink (2022). Do corporate carbon emissions data enable investors to mitigate climate change? *Journal of Portfolio Management 48*(10), 119–147.

Krueger, P., Z. Sautner, and L. T. Starks (2020). The importance of climate risks for institutional investors. *Review of Financial Studies 33*(3), 1067–1111.

Lam, P. and J. Wurgler (2024). Green bonds: New label, same projects. Technical report, National Bureau of Economic Research.

Le Guenedal, T., F. Lombard, T. Roncalli, and T. Sekine (2022). Net zero carbon metrics. *SSRN Working Paper 4033686*.

Le Guenedal, T. and T. Roncalli (2022). Portfolio construction with climate risk measures. *SSRN Working Paper 3999971*.

Lee, H., K. Calvin, D. Dasgupta, G. Krinner, A. Mukherji, P. Thorne, C. Trisos, J. Romero, P. Aldunce, K. Barret, et al. (2023). Ipcc, 2023: Climate change 2023: Synthesis report, summary for policymakers. contribution of working groups I, II and III to the sixth assessment report of the intergovernmental panel on climate change.

Lee, K.-H. and B. Min (2015). Green R&D for eco-innovation and its impact on carbon emissions and firm performance. *Journal of Cleaner Production 108*, 534–542.

Leffel, B., T. P. Lyon, and J. P. Newell (2024). Filling the climate governance gap: Do corporate decarbonization initiatives matter as much as state and local government policy? *Energy Research & Social Science 109*, 103376.

Malhi, Y., J. Franklin, N. Seddon, M. Solan, M. G. Turner, C. B. Field, and N. Knowlton (2020). Climate change and ecosystems: threats, opportunities and solutions.

Martiny, A., F. Testa, J. Taglialatela, and F. Iraldo (2024). Determinants of Environmental, Social and Governance (ESG) performance: A systematic literature review. *Journal of Cleaner Production Forthcoming*, 142213.

Mentch, L. and G. Hooker (2016). Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. *Journal of Machine Learning Research 17*(26), 1–41.

Michalski, L. and R. K. Y. Low (2024). Determinants of corporate credit ratings: Does esg matter? *International Review of Financial Analysis Forthcoming*, 103228.

Molnar, C. (2020). *Interpretable machine learning*. Lulu.

Newey, W. and K. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Nguyen, D. K., T. L. D. Huynh, and M. A. Nasir (2021). Carbon emissions determinants and forecasting: Evidence from g6 countries. *Journal of Environmental Management 285*, 111988.

Nguyen, Q., I. Diaz-Rainey, and D. Kuruppuarachchi (2021). Predicting corporate carbon footprints for climate finance risk analyses: a machine learning approach. *Energy Economics 95*, 105129.

Oyewo, B. (2023). Corporate governance and carbon emissions performance: International evidence on curvilinear relationships. *Journal of Environmental Management 334*, 117474.

Papadopoulos, G. (2022). Discrepancies in corporate ghg emissions data and their impact on firm performance assessment.

Pástor, L., R. F. Stambaugh, and L. A. Taylor (2021). Sustainable investing in equilibrium. *Journal of financial economics 142*(2), 550–571.

Pastor, L., R. F. Stambaugh, and L. A. Taylor (2024). Carbon burden. *SSRN Working Paper 4998860*.

Pedersen, L. H. (2024). Carbon pricing versus green finance. *SSRN Working Paper 4382360*.

Puertas, R. and L. Marti (2021). Eco-innovation and determinants of GHG emissions in OECD countries. *Journal of Cleaner Production 319*, 128739.

Ramadorai, T. and F. Zeni (2024). Climate regulation and emissions abatement: Theory and evidence from firms' disclosures. *Management Science*.

Richards, C., R. Lupton, and J. M. Allwood (2021). Re-framing the threat of global warming: an empirical causal loop diagram of climate change, food insecurity and societal collapse. *Climatic Change 164*(3), 49.

Rink, S., M. Dumrose, and Y. Matheis (2024). Consistency or transformation? finance in climate agreements. *SSRN Working Paper 4835984*.

Roncalli, T. (2024). Portfolio alignment and net zero investing. *SSRN Working Paper 4891271*.

Seyfi, S. (2024). Essence of the cross section. *SSRN Working Paper 4466972*.

Sharma, S. S. (2011). Determinants of carbon dioxide emissions: empirical evidence from 69 countries. *Applied Energy 88*(1), 376–382.

Shwartz-Ziv, R. and A. Armon (2022). Tabular data: Deep learning is not all you need. *Information Fusion 81*, 84–90.

Swinkels, L. and T. Markwat (2023). Corporate carbon emissions data for equity and bond portfolios. *Managerial Finance 50*(1), 118–139.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology 58*(1), 267–288.

Zerbib, O. D. (2019). The effect of pro-environmental preferences on bond prices: Evidence from green bonds. *Journal of banking & finance 98*, 39–60.