An anatomy of decarbonizing firms

Guillaume Coqueret[*] Thomas Giroux[†] Borui Qiu[‡§]

May 15, 2025

Abstract

This article sheds light on the drivers of decarbonization in the cross-section of global publicly-listed firms. We find that reported and estimated emissions depend on very different sets of variables. With respect to reported footprint, biodiversity loss, R&D expenditures and institutional ownership matter the most. In terms of predictive accuracy, we find that sophisticated models based on panel data do not best simple benchmarks and that linear extrapolation is not the best alternative. Moreover, we report marked difference in forecasting errors across sectors, with utilities being the easiest to predict and information technology the hardest. All in all, our findings deliver insights to asset managers seeking net-zero targets.

1 Introduction & Motivation

1.1 Context

Latest scientific consensus shows that the Earth is already about 1.1°C warmer than it was in the late 1800s, and emissions continue to rise. Hence, the multiple threats posed by climate change call for a steady decarbonization of economic systems worldwide.[1] To avoid the worst consequences of climate change and 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 an critial 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.

[*]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]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.

Both academics and practitioners have proposed avenues to reach so-called "*Net Zero*" targets.[2] 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 limit 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[3]. Fortunately, Swinkels and Markwat (2023) find some improvements in recent years, especially for Scopes 1 and 2, with increasing homogeneity in reported data.

Yet, properly estimating current GHG emissions is only the first step to achieve net zero alignment. In addition, companies need to engage in an ambitious transition plans to reduce their GHG emissions. To this end, as summarized by the Transition Plan Taskforce (TPT) in its "Disclosure Framework"[4], 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 a current research on the topic still rely on relatively simple models to predict forward levels of GHG emissions. These predictions are usually simple extrapolation from recent trends at the company (Le Guenedal et al., 2022) or portfolio (Le Guenedal and Roncalli, 2022) level. Obviously, the outcome is suboptimal as these ap-

[2] 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).

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

[4] Summary of the TPT Recommendations

proaches 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 in predicting 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. Moreover, a recent exploratory study from Aldy et al. (2023) reveal how short terms commitments are only marginally met and emphasize on the difficulties of high emitting companies to meet their targets. They suggest that over-promising and unrealistic targets are the main issue toward a greater 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 the world's climate ambitions as set in the Paris agreement.

The present article seeks to propose a data-based approach to determine the drivers of decarbonization in the cross-section of firms. To do so, we compile a large sample of firms worldwide, where each company is characterized by a large 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 confluence of several streams of academic research. First, there is an 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 linked to socially responsible investors and future corporate social responsibility (CSR) scores (Hwang et al., 2022). Likewise, Kahn et al. (2023) find a similar outcome 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 regard to the green transition. Moreover, Berk and Van Binsbergen (2024) show that in contradiction with theoretical findings suggesting that sustainable investing may produce 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. Also, 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 further supported by Heath et al. (2023) and Lam and Wurgler (2024). While the first shows that SRI funds do not significantly change firm behavior, the latter emphasize that only 2% of corporate green bond proceeds 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) lay out a model on the impact of carbon taxes and green subsidies on firm's willingness to invest in decarbonization technologies. The model predicts that large firms (or conglomerates) can benefit from investing in green technologies and lead the way towards decarbonization. Furthermore, Pedersen (2024) also shows that carbon prices are the most efficient tools to optimize social welfare (aggregate consumption minus disutility 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 length of the carbon pledge, the overall level of emissions, and each company's sales growth.

Moreover, a few papers tackle the challenging exercise of ESG scores or GHG emissions forecasting based on statistical tools. Goldhammer et al. (2017) where the first to introduce a set of companies' data and they suggested Ordinary Least Squares (OLS) and Gamma Generalized Linear Regression (GGLR) to model GHG emissions forecasts. They achieve higher match with reported data compared to simple extrapolation, but for a very limited set of European companies. Nguyen et al. (2021) updated the 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. Their significantly improve in and out-of-sample accuracy. Nevertheless, they highlighted limitations in disentangling between direct and indirect emissions in their predictions, as well as discrepancies between industries. For instance, the average R2 for most polluting industries in direct emissions (Scope 1) barley 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 levels, and thereby trying to quantify the social externality arising from damages caused by corporate emissions of greenhouse gases. 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 macro-economic level. While most analysis tend to focus a specific country at a time, we rather 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[5] (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 sig-

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

nificant 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 2015 to 2022. We rely on ISS data as they are one of the mainstream provider and they gathered a very large database of both reported and estimated emissions. We compute our own estimate of GHG emissions intensities by scaling raw emissions with Enterprise Value Including Cash (EVIC) that we obtain from Bloomberg. This approach has become standard in the industries as well as academic research on the topic. Table 7 of the appendix displays the name, source and definition of each variable.

Our panel sample covers 6456 unique companies over the period across four regions (North America, Europe, Japan and Asia ex Japan). As shown in Figure 1, North America dominates in terms of number of companies and total market capitalisation, followed by Europe and Asia. We also observe a significant bias toward small size companies, reflecting the true nature of financial markets (See Table 1 for more details).

Finally, companies reporting on GHG emissions increase over time, ranging from 40% to 70% in the latest year. Surprisingly, while these companies are larger in terms of market capitalisation than the one not disclosing, we also observe a similar proportions in terms of total amount of GHG emissions, as the reported data account for roughly 40% of total GHG emissions in our dataset for the first year to more than 70% in 2022. Figure 2 summarize this finding over time.

We further identify a broad list of candidate drivers of firms' decarbonization based on the literature. For clarity, we grouped them in the following categories.

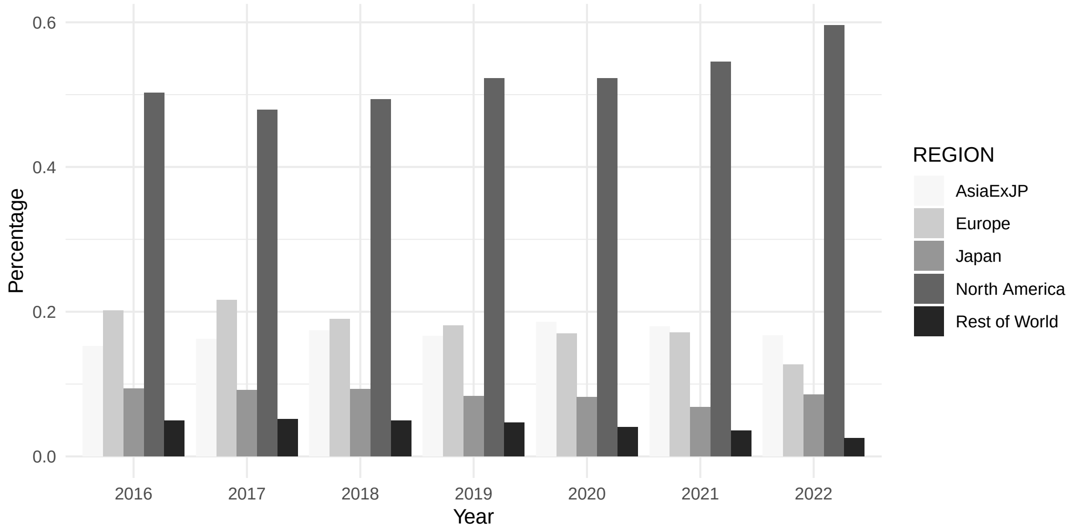

Figure 1: **Share of total market capitalisation by region and year**

Region	min	10%	50%	mean	90%	max
North America	20.00	562.00	3.39×10^3	1.69×10^4	3.34×10^4	2.35×10^6
Europe	32.00	354.00	2.19×10^3	9.46×10^3	2.39×10^4	3.74×10^5
Japan	78.00	576.00	2.55×10^3	6.83×10^3	1.67×10^4	2.72×10^5
Asia Ex Japan	19.00	504.00	2.85×10^3	8.69×10^3	1.79×10^4	6.79×10^5
Rest of World	48.00	520.00	2.58×10^3	7.00×10^3	1.40×10^4	1.60×10^5

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

Sustainability Indicators. First, we identify the firms' ESG profile as indicative of their awareness toward climate-related issues and ability to conduct change. Hence, we gather E,S and G scores separately from ISS and Refinitiv. In addition, because climate change and biodiversity crisis are deeply intertwined, we leverage on Iceberg-Datalab to include corporates biodiversity footprint (CBF) as a proxy of the firms' impact on biodiversity. Finally, we include firms' GHG reduction targets reported by the Science Based Target initiative (SBTi). The usefulness of those targets is supported by the evidence of Bolton and Kacperczyk (2023) and Ramadorai and Zeni (2024), who find that firms that commit to reducing their carbon emissions indeed tend to do so subsequently.

Fundamental & Market Indicators. Second, we consider firms' financial ability 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. We also include momentum from sell-side analysts in the form of the proportion of positive price target and earnings per share revisions over the last three months.

Ownership & Debt Structure. Third, according to several studies, the ownership structure influences both climate disclosure ande performance. Following insights from Bolton and Kacperczyk (2021); Cohen et al. (2023) we include the percentage of float shares held

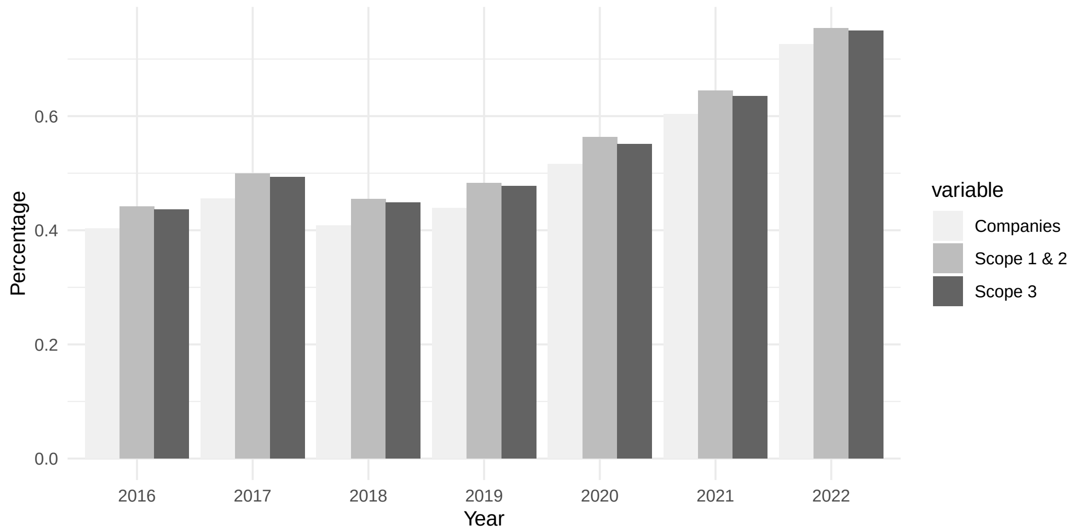

Figure 2: **Share of companies and emissions reported by year**

by institutional investors as provided by Bloomberg. Kahn et al. (2023); Rink et al. (2024) also suggest that active shareholders engaging with companies can influence their environmental and social performance. Therefore, we add the percentage of insider shares' outflow and 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.

In addition, we also include the structure of debt owners by computing a green debt ratio that is the sum of green, social and sustainability-linked bonds (GSSB) over the total debt issued.

Regional Indicators*.* Finally, we also believe that macro-economic and regions characteristics play a significant role in firms' ability and willingness to tackle climate change and reduce their GHG emissions. We include the GDP growth from the World Bank, 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 2013 – 2023 period.

A summary of all variables names, sources and interpretations can be found in Table 8 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 lag all independent variables by one year, with the aim to conduct a predictive analysis and be able to forecast next year GHG emissions based on year t observations.

2.2 Outliers detection and data imputation

Outliers treatment. We remove all non-positive values of the GHG emissions data, as well as companies with non-consecutive tickers over the years to avoid any reporting inconsistencies. We brute force the detection of outliers by setting a threshold at 50% reduction as well as the 200% boosting in the intensity of emissions, compared to the previous year, based on the assumption that such changes are unrealistic for a continuous running business. For all independent variables, we apply a 0.1% percentile winsorization by sector to avoid extreme value distortion.

Data imputation. We conduct 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 capitalisation of the company compared to the sector average while doing the grouping.

2.3 Omitted variables due to colinearity

Figure 8 in the Appendix presents the correlation matrix of all independent variables as defined in the above. We observe several groups with high correlation, and therefore drop some of them to avoid the colinearity issues in linear models estimation.

ESG Scores. We will use the three E, S and G pillars provided by ISS. Similar scores from Eikon (Refinitiv) were discarded, as they were redundant.

Size factor. Market capitalisation, revenue and enterprise value, as well as the lag emission are highly correlated with each other. By keeping the lag emission only, we not only avoid the colinearity issues but we also keep the most relevant information to explain firms' future emissions.

Green Bond. The amount of green bond issued is highly correlated with the current repayment amount for green bonds. Thus we keep the latter, and assume that the drawdowns is more indicative of the firm's financial commitment to the green transition.

Regional Policy. All three regional indicators from OECD are highly correlated with each other. We keep the cross sectional policy score for linear models. However, for deep learning models; such as random forest, we keep all the energy mix to let the model decide which one is the most relevant. Moreover, we include the GDP growth of the country where the firm's headquarters are located.

Country Energy Stack. Energy mix data from IEA appear to have strong correlation. Hence, drop several energy sources and keep coal, electricity, gas and oil mix with the prior belief that those energy mix could be more indicative of the firm's ability to reduce their emission. Again, for deep learning models, we will keep all the energy mix variables.

2.4 Descriptive statistics

We split the summary of the data in two and discriminate between the dependent variables and the predictors.

For each scope, we identify four types of dependent variables: *i)* raw emissions measured in tons CO2 equivalent (tCO2e) - which we refer to as **GHG** - *ii)* emission intensities (**INT**), reflecting tCO2e per million of enterprise value (EVIC), *iii)* relative changes in absolute emissions (Δ**GHG**) and *iv)* 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 because they are less prone to disagreement and divergence, and hence are more reliable. We run robustness tests on estimated emissions for scope 1+2 and both reported and estimated scope 3 emissions. Figure 3 below and Table 9 in the appendix provide an overview of the distribution of these variables for our baseline sample.

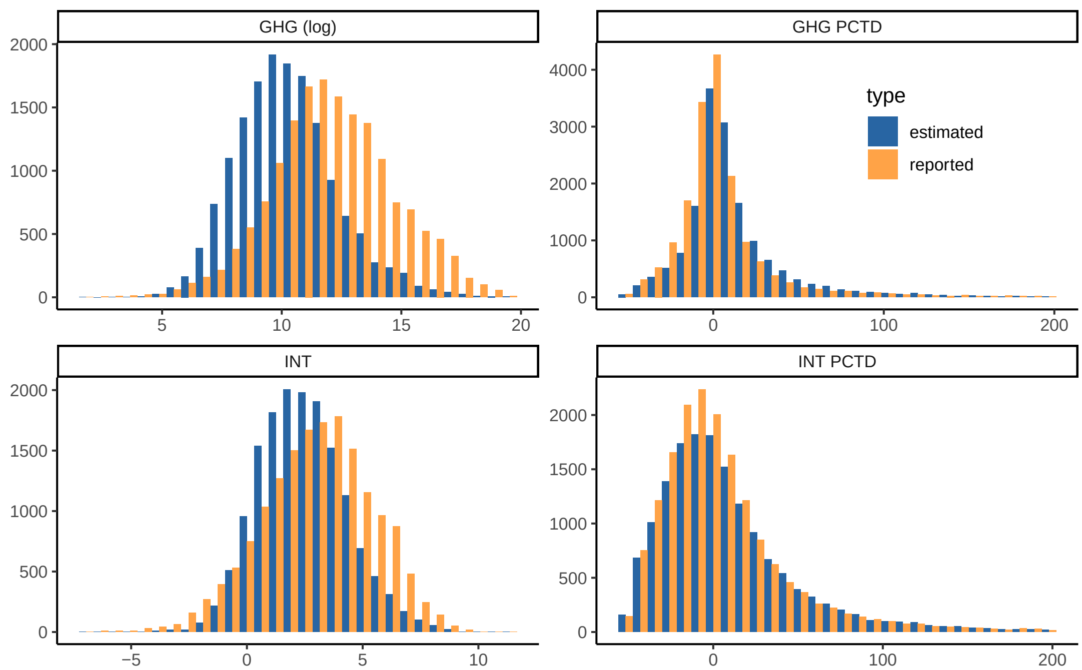

Figure 3: **Distribution of dependent variables**.

Reported Scope 1+2 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.[6]

Now, looking at the exogenous variables that we use as potential predictors in our forth-

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

coming analysis, we display the main descriptive statistics in Table 10. 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.

The correlation matrix of all variables presented in Table 8 in the Appendix, reveals an intriguing fact - there is a strong correlation between absolute levels of GHG emissions and the intensities, once scaled by enterprise value including cash (EVIC). A key reason lies in the computation of EVIC. Indeed, EVIC calculation includes not only the market capitalization of a company but also short-term and long-term debt, as well as any cash or cash equivalents on the company's balance sheet. As such, EVIC is less prone to volatility than market capitalization solely and therefore, GHG intensities yearly percentage change is mostly driven by changes in the numerator. Aside from this, once controlled for outliers and colinearity as described in Section 2.2, our set of exogenous drivers have a low number of strongly correlated variables.

3 The drivers of decarbonization

This section seeks to decipher the determinants of decarbonization over the full sample. First, in Section 3.1, we run linear models: both simple panels and penalized ones. As explained in Section 2.1, we lag all explanatory variable by one year. Next, in Section 3.2, we resort to random forests in order to capture potential nonlinearities in the features that drive emission and intensity reductions. Finally, in Section 3.3, we group firms according to their realized variations in emissions and figure out which characteristics differ the most between those that reduce footprint versus those that tend to pollute 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 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 10 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 industry group 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.

	GHG (log)		ΔGHG		INT (log)		ΔINT	
	estimated	reported	estimated	reported	estimated	reported	estimated	reported
ISS_E	2.338	0.281	2.421	−0.348	−2.110	−0.696	−2.077	0.975
ISS_S	−2.536	0.017	−2.361	−0.046	3.063	0.634	2.654	−0.281
ISS_G	0.483	−2.019	0.481	−1.338	−2.933	−5.076	−2.832	4.369
CBF	−0.315	−1.829	−0.615	−1.725	−0.503	3.767	−0.833	−3.889
CBF_INT	5.036	3.651	4.644	3.543	−6.746	−6.016	−7.083	6.245
BLG_RETURN	7.249	0.174	6.908	0.031	5.174	1.026	5.732	−1.371
BLG_PE	−0.154	0.381	−0.073	0.306	−0.203	1.597	−0.074	−1.502
BLG_PS	0.880	0.532	0.474	0.654	2.484	−0.138	2.777	−0.567
BLG_PB	5.142	1.664	4.842	0.976	2.557	0.660	4.202	−1.819
BLG_DE	0.650	3.233	1.038	3.923	0.585	1.368	−0.518	−0.175
BLG_LT_DEBT	1.440	−2.282	1.420	−1.319	−0.950	−1.531	−1.549	1.457
BLG_PROFIT	−0.105	−0.856	−0.568	−1.029	1.087	0.146	1.432	0.650
BLG_IC_RATIO	−0.800	0.778	−1.085	0.922	0.276	0.606	−0.321	−0.066
BLG_RD_EXP	3.462	−3.095	3.516	−2.781	0.640	−3.574	0.708	3.832
BLG_FCF	0.020	−0.192	−0.276	−0.005	−0.918	−2.113	−1.027	2.489
BLG_LT_INV	−1.431	−0.240	−1.205	−0.463	−0.619	−0.412	−0.347	0.319
FE_REVUP_PCT	5.075	2.899	3.932	2.545	−1.562	−0.024	−2.338	0.598
BLG_SHARE_OUT	2.153	4.715	1.632	4.788	0.297	3.828	0.698	−4.545
BLG_INSTI_SHARE	−0.341	−5.015	−0.423	−5.019	0.011	−2.075	−0.443	2.356
FS_GREEN_SHARE	−0.593	−0.834	−0.899	−1.373	0.188	0.303	0.068	−0.389
BLG_SHARE_REPUR	0.399	2.488	0.601	2.166	0.548	0.416	0.232	−0.533
FS_GB_PAID	0.539	0.021	0.309	−0.221	0.682	0.000	0.732	0.193
WB_GDP	2.536	6.712	2.066	6.942	1.993	4.577	2.074	−5.255
IEA_COAL	−1.928	7.883	−2.174	7.395	4.991	8.591	3.883	−7.772
IEA_OIL	−3.387	4.084	−3.423	3.052	3.453	2.325	2.445	−0.513
SBTI	1.695	−7.442	1.324	−7.789	1.878	−4.576	1.327	5.406
GHG_S12_LOG	654.430	815.410	−14.227	−9.534				
GHG_S12_PCTD	2.430	0.356	2.549	0.970				
INT_S12_LOG					422.807	547.208	−12.396	10.854
INT_S12_PCTD					−3.139	−5.143	−1.940	3.946

Table 2: **Panel model - industry group 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.

The 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. Similarly, an increase in GHG emissions is difficult to overcome in the following year. This was expected because the evolution of emissions is highly persistent over time. 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 to explain the levels and changes of absolute emissions and intensities. In fact, almost each category described in Section 2.1 has at least one significant predictor. However, we uncover notable discrepancies between reported and estimated data, suggesting potential limitations of current estimation models and emphasize the need for further refinement.

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. We also observe a significant and negative effect of higher governance score on future GHG emissions, supporting the idea that greater governance improves chances of reducing GHG emissions.[7]

Additionally, certain firm-level financial characteristics, such as sales, debt-to-equity ratio, demonstrate a consistent influence on GHG emissions. Positive financial performance, proxied by average monthly return, have a positive impact on reducing future GHG emissions, both in intensities but also in absolute terms. Higher debt ratio appears as a barrier to engage in GHG emissions reductions. However, the structure of the debts seems to matter. Indeed, long-term debt seems to allow to engage in emissions reduction in absolute terms.

The role of the macroeconomic environment, particularly GDP growth and energy mix (specifically coal and oil usage) is another common factor affecting emission reduction dynamics.

Despite these shared insights, the analysis also uncovers notable discrepancies between reported and estimated data. While environmental and social scores show limited predictive power for reported GHG emissions, they exhibit strong significance for estimated emissions. The result on reported data are in line with Kalesnik et al. (2022) but the latter potentially indicate modeling biases. More surprisingly, the relationship between environmental scores and absolute emissions is positive for estimated data, contrary to expectations.

Moreover, the impact of ambitious reduction targets reported to SBTI is evident only in reported data. The more ambitious the reduction target, the lower GHG emissions are and become, both in absolute values and intensities. This result suggest that while companies with more ambitious targets are often ones with lower emissions than their peers from the same sectors, they are nonetheless achieving higher reductions than their peers too. Moreover, current estimation models appear to not adequately capture the effect of such commitments on future emissions.

Similarly, few traditional valuation ratios have an impact on future GHG emissions. The price-to-book ratio appears significant in explaining estimated emissions, but not reported ones. This is in line with findings from Pastor et al. (2024) studying the drivers of GHG forecasts from MSCI and potentially underscores the fact that data providers use the B/M

[7] We point to Haque (2017), Elsayih et al. (2021) and Oyewo (2023) for prior work on the impact of firm governance on corporate carbon policy.

ratio as component of their estimation models. With regards to reported emissions, the debt-to-equity ratio stands out.

Furthermore, companies with higher R&D spending have significantly lower GHG emissions, both in level, intensities and reduction dynamic over time. The same holds for institutional ownership.

We also see analysts' sentiment (FE_REVUP_PCT) being a steady predictor of future GHG emissions for estimated data, while it is not significant for reported data. Again, this may point to a component of data vendors' internal models. Finally, the influence of institutional ownership on GHG emission levels and reduction dynamics is significant for reported data but absent in estimated data.

As a robustness check, we provide in Table 11 in the Appendix the results of the regression with individual fixed-effects instead of industry group fixed-effects. The main findings are qualitatively similar.

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 re-write 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

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

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 from Equation (1), we include fixed effects in the model. However, they are rarely among the surviving variables and none make it even near 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 12 in the appendix).

Our results confirm the main findings from the previous section presented in Table 2. Past emissions appears to be the most predictive variables of future emissions. Biodiversity footprint is also strongly related, emphasizing the deep interconnections between climate and biodiversity impacts of economic activities. In addition, financial health and characteristics such as past returns, analyst sentiment, sales, debt to equity and R&D show up again in the top of the list. Finally, macro-economic variables such as GDP growth and country energy mix, as well as shareholder structure, especially institutional ownership are confirmed

		Raw footprint					Relative change			
		GHG (log)		INT (log)			ΔGHG		ΔINT	
rank		est.	rep.	est.	rep.		est.	rep.	est.	rep.
1	GHG_S12_LAG	99	98	56	47	BLG_PB	99	76	99	78
2	INT_S12_LAG	22	0	99	99	BLG_SALES	87	96	74	93
3	ISS_E	9	0	32	28	BLG_SHARE_OUT	81	89	67	87
4	BLG_RETURN	0	0	28	33	FE_REVUP_PCT	95	57	79	93
5	CBF_INT	0	0	26	32	WB_GDP	55	60	99	99
6	BLG_PB	10	0	28	16	CBF_INT	75	75	83	75
7	BLG_SALES	16	0	32	4	BLG_RD_EXP	85	75	80	67
8	WB_GDP	0	0	31	16	BLG_RETURN	67	65	99	74
9	INT_S12_PCTD_LAG	7	0	0	28	BLG_LT_DEBT	71	88	73	72
10	FE_REVUP_PCT	0	0	6	29	INT_S12_PCTD_LAG	68	76	77	82
11	BLG_CAP	0	0	8	25	GHG_S12_PCTD_LAG	80	75	74	72
12	BLG_RD_EXP	0	0	12	20	SBTI	55	99	76	71
13	BLG_SHARE_OUT	0	0	18	11	BLG_DE	85	60	84	59
14	BLG_SHARE_REPUR	7	0	19	0	OECD_CROSS	37	92	81	78
15	BLG_FCF	0	0	24	0	ISS_E	85	55	79	67
16	BLG_INSTI_SHARE	0	0	24	0	BLG_LT_INV	71	75	66	64
17	IEA_COAL	0	0	0	23	BLG_FCF	75	49	69	70
18	SBTI	0	0	14	7	INT_S12_LAG	71	75	53	63
19	BLG_EVIC	0	0	17	3	BLG_INSTI_SHARE	47	89	54	71
20	IEA_ELEC	0	0	14	4	BLG_PS	75	48	84	46

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.

as significant drivers of firms future GHG emissions. However, it is noteworthy to see that environmental score comes up as a significant driver too, while despite having a quite large t-statistic, it was not the case when looking at the panel regression estimates. This underlines the current ambiguous academic findings about E,S and G scores usefulness in predicting future GHG emissions, as different studies such as Kalesnik et al. (2022); Pastor et al. (2024) present 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.

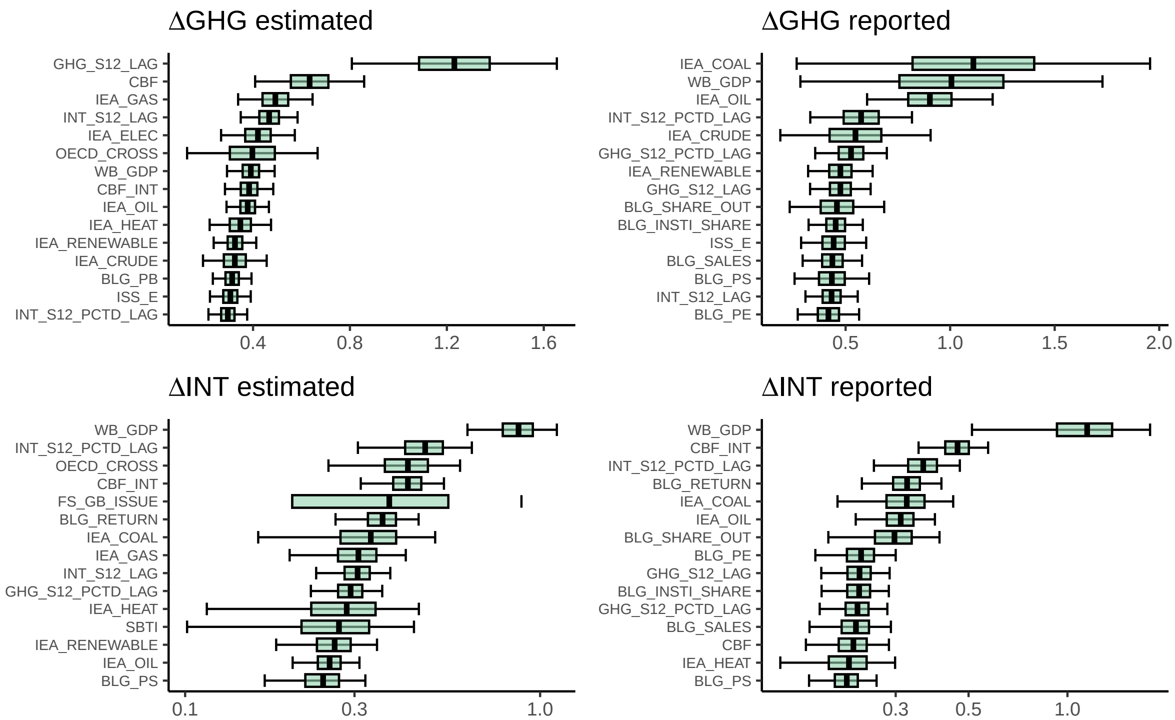

Figure 4: **Confidence regions for feature importance**. We plot the median (center), interquartile range (boxes) and 95% interval (whiskers) for the variable importance, following Ishwaran and Lu (2019). The number of bootstrap samples is 100. The forests were fitted separately on the estimated (left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

We focus on changes of GHG emissions and intensities as absolute values are strongly driven by past values. Therefore non-linear modeling will make less of a difference to explain absolute values compared to their dynamic.

Our results confirm the predictive power of past values of the dependent variable. In the top left panel, GHG past emissions has a feature importance median almost twice as large as the next feature which is the biodiversity footprint of the company. As in Table 2, this emphasizes how intertwined climate and biodiversity are. Biodiversity intensity is also the second predictor for changes in reported intensities (bottom right panel). We find relatively smaller values for reported data and intensities, but past emissions stay among the top predictors.

Macroeconomic variables appear more significant compared to non-penalized linear models for GHG emissions. Interestingly, for estimated GHG emissions, some new variables appear, for instance related to the OECD and to the IEA. The latter were already associated with significant t-statistics in Table 2.

In all panels, local economic growth (WB_GDP) is key. This can be perhaps explained by the fact that growth can have an impact on the denominator of the intensity and it also echoes the high (absolute) statistics in Table 2.

The nonlinear models confirm the results reported above. The main difference with Table 2 is the absence of financial ratios like B/M or D/E. The return was significant for estimated data in the table, and in Figure 4, it only stands out for intensities - both reported and estimated.

3.3 The characteristics of decarbonizing firms

This last sub-section pertains to the characteristics of sorted firms. We first split the sample, on a year-by-year basis, in two groups: firms with low versus high dependent variable.[8] Then, for each group, we compute the average of characteristics. Finally, for each characteristic, we test if the average value is significantly different, from one group to the other (high GHG or INT versus low GHG or INT). Figure 5 displays our findings for levels of GHG emissions while Figure 9 provided in Appendix extends the analysis to GHG intensities. It plots the absolute t-statistics of the tests for which the null is that there is no difference between the two groups.

We observe common drivers among top vs bottom performers across both GHG absolute emissions and their dynamic for reported and estimated values. First, some firms' financial characteristics such as returns, sales, and price-to-book ratios have significant influence. Analysts' sentiment (FE_REVUP_PCT) also contributes to differentiate between top and bottom performers. In addition, past GHG emissions' dynamic is a critical predictor of firms with future high or low levels of emissions.

Nonetheless, we still find many discrepancies between reported and estimated data. Indeed, estimated data seems to rely on financial momentum, leveraging on past returns and analysts sentiment, as well as the price to book to account for companies valuation. It is not what we observe for reported data however. In this sample, the structure of long term debt and investment matters more than valuation and financial momentum to explain top versus low carbon performers. Moreover, the ownership structure of the company and its commitments to reduce GHG emissions (SBTI) are only slightly explaining top versus low performers in estimated data while their are of first importance when considering reported data only.

Finally, when comparing Table 5 to Table 9, we find that the impact of institutional ownership is more relevant in explaining top versus bottom performers in carbon intensities than it is for absolute emissions. This underlines the importance of carbon intensities as a metric for investment decisions, whereas absolute level of emissions are currently not being well integrated in portfolio alignment frameworks. Therefore, institutional investors may overweight their investments in companies with low emissions intensities while missing the big picture of absolute emissions. The results also logically emphasize the increased impact of firms' market capitalization and financial performance in explaining top versus bottom performers in GHG intensities versus absolute emissions.

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

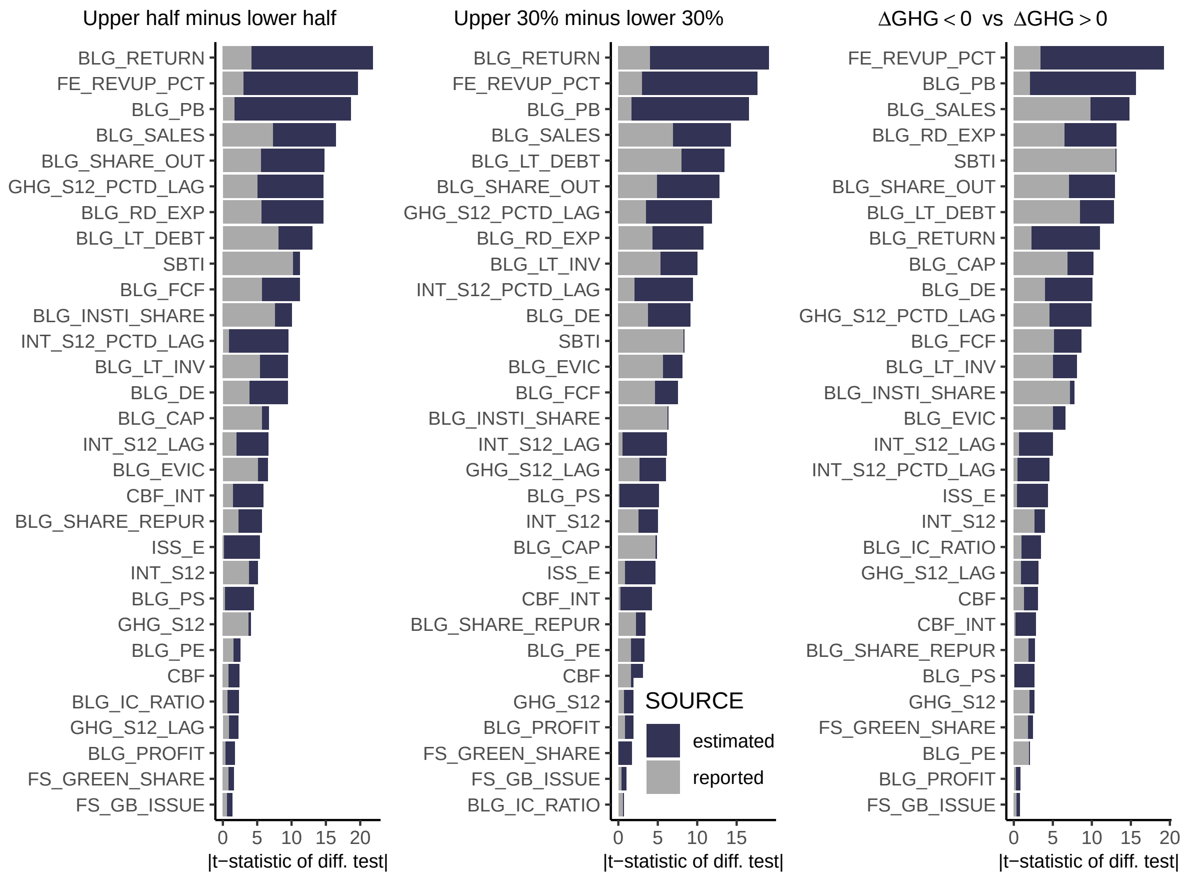

Figure 5: **Characteristics of decarbonizing firms**. We plot the absolute values of the t-statistics of the tests in differences in means between characteristics with high versus low ΔGHG.

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 (2022), 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

$$y_{t+1,n} = f(\boldsymbol{x}_{t,n}) + e_{t+1,n}, \tag{3}$$

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:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(y_{T+1,n} - \hat{f}(\boldsymbol{x}_{T,n})\right)^2}, \tag{4}$$

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 6 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 4 below.

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

Our main results are gathered in Table 5. 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).

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 6, but in fact, while the good performance of extrapolation holds for large firms, it is

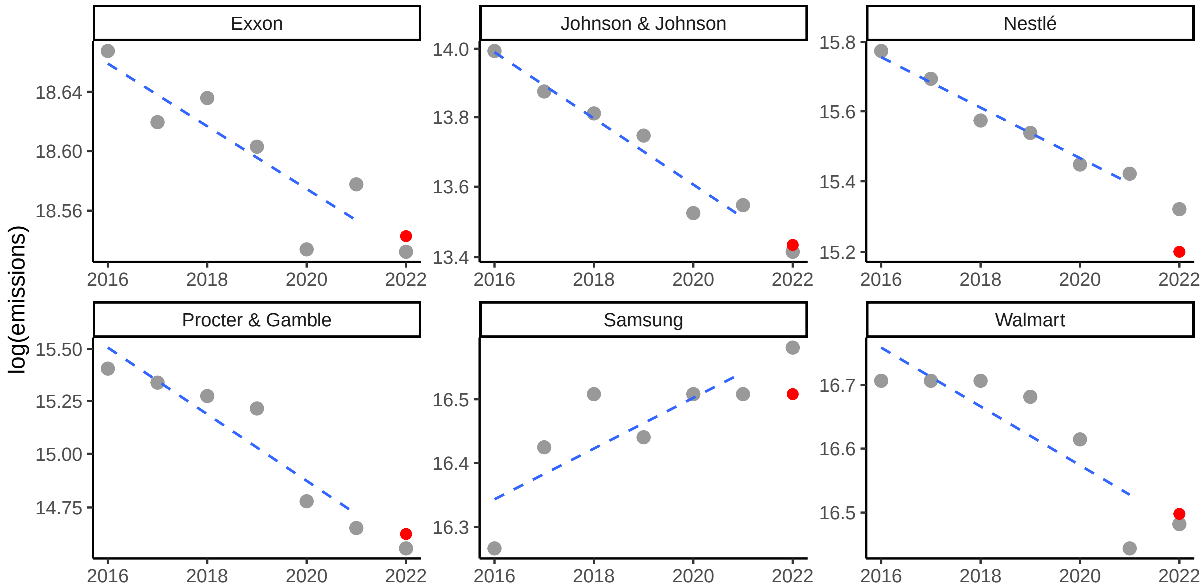

Figure 6: **Extrapolation for large firms**. We show the dynamics of reported log emissions for the six largest firms in the sample (in terms of revenue) for which all seven years of data are available. The first six years are shown in grey and the last point is the extrapolation. The realized value for 2022 is shown in red.

much more erratic for smaller firms. In Figure 10 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.

The next batch of rows (Panel B in Table 5) provides the minimum, median and maximum of the RMSE across the 108 hyperparameter combinations of Table 4. 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

$$\Delta_{i,t} = \begin{cases} \log(\text{GHG}_{i,t}) - \log(\text{GHG}_{i,t-1}) & \text{for emissions} \\ \log(\text{INT}_{i,t}) - \log(\text{INT}_{i,t-1}) & \text{for intensities.} \end{cases} \tag{5}$$

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 5. They are all smaller than in Panel

dependent variable	**GHG**		**GHG PCT**		**INT**		**INT PCT**	
sample	all	sub	all	sub	all	sub	all	sub
PANEL A: benchmarks								
constant	**0.242**	**0.219**	**0.277**	**0.235**	0.337	**0.293**	0.455	0.373
extrapolation	-	0.299	-	0.264	-	0.393	-	0.411
PANEL B: random forests (statistics across HP combinations)								
min	0.249	0.224	0.284	0.245	0.348	0.302	0.416	0.352
median	0.251	0.226	0.292	0.252	0.352	0.305	**0.419**	**0.354**
max	0.271	0.229	0.296	0.257	0.373	0.310	0.424	0.356
PANEL C: learning from errors (statistics across HP combinations)								
min	0.243	0.221	-	-	0.336	0.297	-	-
median	0.245	0.223	-	-	0.339	0.299	-	-
max	0.248	0.225	-	-	0.342	0.302	-	-

Table 5: **Out-of-sample performance**. We report the root mean squared errors (RMSE) of the predictive models for the year 2022. In Panel A, we report two benchmarks: the constant value (from 2021) and the extrapolated one, from 2016 to 2021. For the extrapolation, the estimations are run on the set of firms for which all seven 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 described in Table 4. The minimum value of each column is highlighted in bold.

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 6. 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.

dependent variable	$1_{\Delta GHG<0}$		$1_{\Delta INT<0}$	
sample	all	sub	all	sub
PANEL A: Benchmarks				
constant	0.504	0.513	0.471	0.474
linear prediction	-	0.506	-	0.445
slope sign	-	0.522	-	0.391
PANEL B: Random Forest predictions (regression)				
min	0.501	0.487	0.445	0.444
median	0.510	0.506	0.478	0.479
max	0.524	0.527	0.496	0.503
PANEL C: Random Forest predictions (classification)				
min	0.565	0.583	0.426	0.436
median	0.584	0.600	0.446	0.453
max	0.593	0.621	0.474	0.487

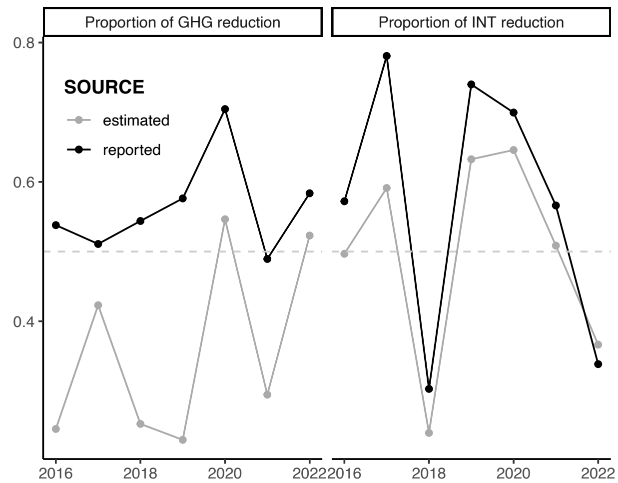

Table 6: **Classification of decarbonization**. In the left table, we report the accuracy of the predictive models for the year 2022. In Panel A, we report three benchmarks: i) the decarbonization status of 2021 is assumed to remain, ii) the linear prediction is used to determine if emissions or intensities will decreases 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 seven 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 described in Table 4. In this case, the RF models are used to generate predictions for 2022 and decarbonization is inferred from the realized 2021 emissions and intensities. In Panel C, it is the binary variable of decarbonization that is directly forecasted. In the right plot, de 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.

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 6. 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 7, 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

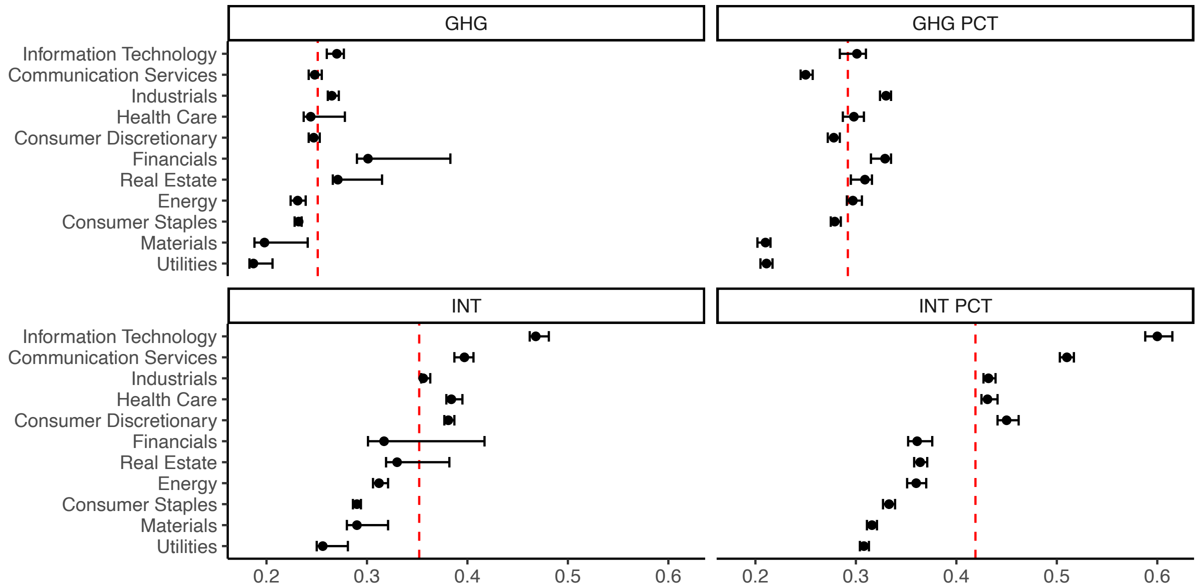

Figure 7: **Sector variations**. This plot shows the RMSE of the predictive models from Table 5 (Panel B) aggregated at the sector level. The error bars mark the variations across the hyperparameters. The dashed red lines shows the RMSE over all sectors.

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.

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 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)
II. Emissions characteristics		
ISS_Source	**ISS ESG**	GHG Emissions source (reported vs. estimated)

Table 7: **Description of dependant variables**

Parameter	**Source**	Description
I. Sustainability Indicators, previous period		
ISS_E	**ISS ESG**	ISS Environmental Score
ISS_S	**ISS ESG**	ISS Social Score
ISS_G	**ISS ESG**	ISS Governance Score
EK_S	**Refinitiv**	Refinitiv Social Pillar Score
EK_G	**Refinitiv**	Refinitiv Governance Pillar Score
EK_E	**Refinitiv**	Refinitiv Environmental Pillar Score
CBF	**Iceberg DataLab**	Corporate Biodiversity Footprint (CBF), expressed in $km2.MSA$
CBF_INT	**Iceberg DataLab**	Corporate Biodiversity Footprint (CBF), scaled by EVIC
SBTI	**ISS ESG**	Corporate voluntary emissions targets indicator
II. Fundamental & Market Indicators		
COUNTRY_ISO	**Bloomberg**	Country of main quotation

Continued on next page

(Continued)

Parameter	Source	Description
GICS_SECTOR	**Bloomberg**	GICS Sector
GICS_IG	**Bloomberg**	GICS Industry Group
BLG_RETURN	**Bloomberg**	Average annualised day to day return net of dividend
BLG_EVIC	**Bloomberg**	Monthly end enterprise value including cash (EVIC), year average
BLG_SALES	**Bloomberg**	Sales revenue turnover, yearly fundamental
BLG_PE	**Bloomberg**	Price to earning ratio, calendar year average
BLG_PS	**Bloomberg**	Price to sales, calendar year average
BLG_PB	**Bloomberg**	Price to book ratio, calendar year average
BLG_CAP	**Bloomberg**	Monthly end market capitalization, calendar year average
BLG_DE	**Bloomberg**	Total debt to total equity ratio, yearly fundamental
BLG_LT_DEBT	**Bloomberg**	Long term debt on the balance sheet, yearly fundamental
BLG_PROFIT	**Bloomberg**	Profit margin ratio, yearly fundamental
BLG_IC_RATIO	**Bloomberg**	Interest coverage ratio, yearly fundamental
BLG_RD_EXP	**Bloomberg**	Research and development expenditure, income statement yearly fundamental
BLG_FCF	**Bloomberg**	Free cash flow, yearly fundamental
BLG_LT_INV	**Bloomberg**	Long term investment, yearly fundamental
FE_REVUP_PCT	**FactSet**	Percetnage of analyst upward revisions in the last 3 months, clendar year average.
III. Ownership & Debt Structure		
BLG_SHARE_OUT	**Bloomberg**	Percentage of insider shares outflow at the end the reporting period
BLG_INSTI_SHARE	**Bloomberg**	Percentage of float shares held by institutions, calendar year average
FS_GREEN_SHARE	**Exane, Factset**	Percentage of shares held by "*Active*" sustainable ESG funds as identified by Exane (2023).
BLG_SHARE_REPUR	**Bloomberg**	Total value of shares repurchased, yearly fundamental
FS_GB_ISSUE	**FactSet**	Accumulated Green Bond issuance (% of total long-term debt)
FS_GB_PAID	**FactSet**	Accumulated Green Bond repaid (% of total long-term debt)
IV. Regional Indicators		
OECD_CROSS	**OECD**	OECD Cross sectional Environmental policy Stringency, No US information, considered 0
OECD_INT	**OECD**	OECD Cross International Environmental policy Stringency, No US information, considered 0
OECD_SEC	**OECD**	OECD Cross Sectoral Environmental policy Stringency, No US information, considered 0

Continued on next page

(Continued)

Parameter	**Source**	Description
WB_GDP	**World Bank**	GDP growth percentage. Taiwan from different source
IEA_COAL	**IEA**	Coal Energy Country Level Energy Consumption Ratio from IEA yearly
IEA_CRUDE	**IEA**	Crude Energy Country Level Energy Consumption Ratio from IEA yearly
IEA_ELEC	**IEA**	Electricity Energy Country Level Energy Consumption Ratio from IEA yearly
IEA_HEAT	**IEA**	Heat Energy Country Level Energy Consumption Ratio from IEA yearly
IEA_GAS	**IEA**	Natural Gas Energy Country Level Energy Consumption Ratio from IEA yearly
IEA_OIL	**IEA**	Oil Energy Country Level Energy Consumption Ratio from IEA yearly
IEA_RENEWABLE	**IEA**	Renewables Energy Country Level Energy Consumption Ratio from IEA yearly

Table 8: **Description of explanatory variables**

6.2 Descriptive statistics

	type	N	min	25%	median	mean	75%	max	std. dev.
GHG	estimated	13583	3.84	8.93	10.18	10.34	11.53	19.53	2.08
GHG	reported	14583	3.26	10.63	12.15	12.26	13.82	19.75	2.47
ΔGHG	estimated	13583	-49.60	-3.45	5.17	11.27	18.45	194.10	30.71
ΔGHG	reported	14583	-49.46	-10.63	-1.90	2.74	7.64	194.10	29.05
INT	estimated	13583	-4.03	1.18	2.44	2.52	3.71	9.55	1.87
INT	reported	14583	-6.05	1.56	3.16	3.10	4.69	10.85	2.32
ΔINT	estimated	13583	-49.87	-18.98	-1.23	6.23	20.91	195.24	38.01
ΔINT	reported	14583	-49.87	-19.63	-3.45	3.86	17.09	195.24	36.56

Table 9: **Descriptive statistics of dependant 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 ($\Delta \text{GHG}_{t,i}$ and $\Delta \text{INT}_{t,i}$) are trimmed below 50% and above 200%. Emissions are reported after the log. Intensities are computed as raw emissions divided by Enterprise Value including Cash (EVIC) in USD.

.

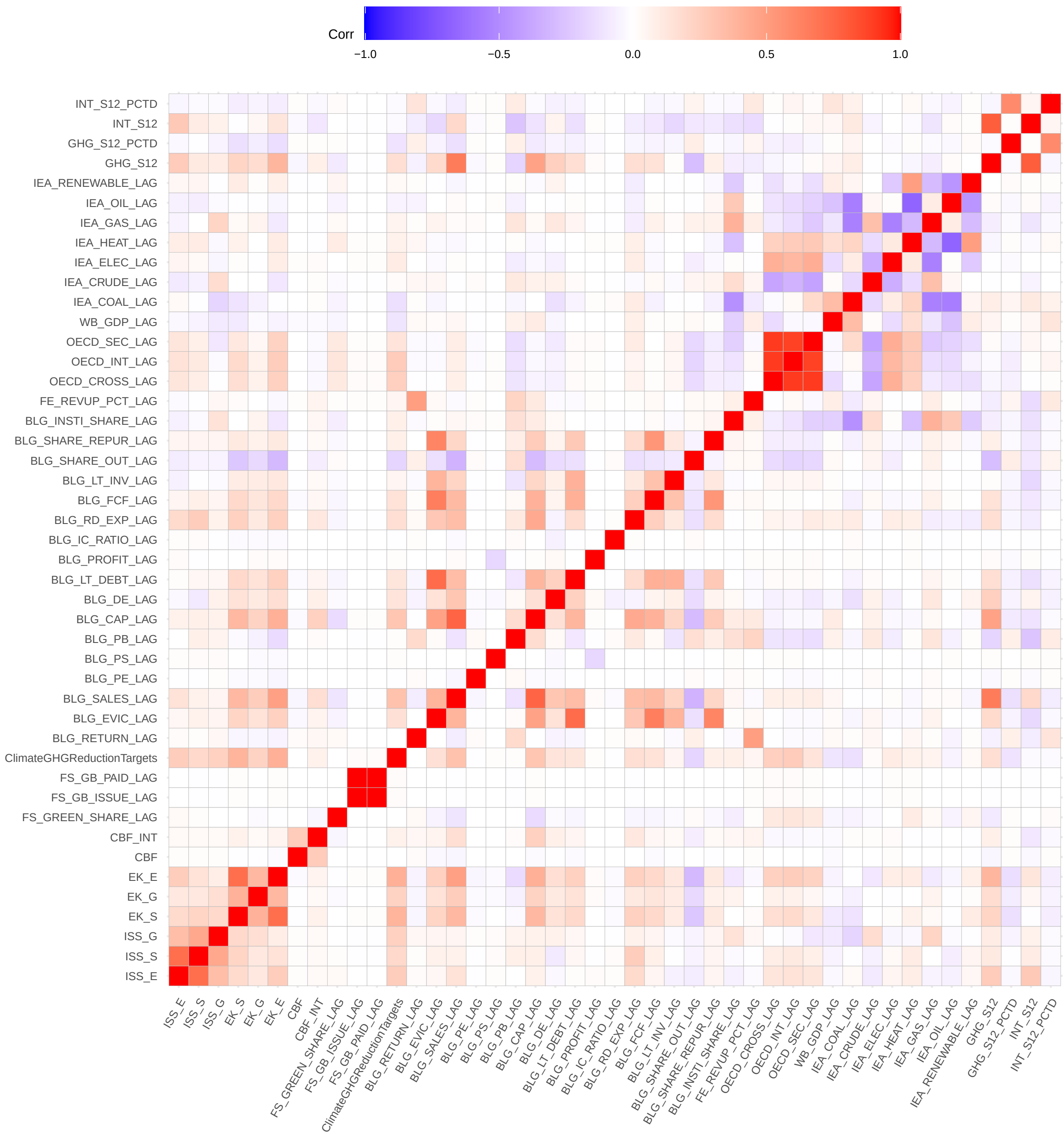

Figure 8: **Correlation matrix of all initial variables**.

Panel A: Modeled

Parameter	count	mean	std	min	1%	50%	99%	max
I. Sustainability Indicators								
ISS_E	13583	1.64	0.35	1.00	1.03	1.68	2.72	3.31
ISS_S	13583	1.76	0.31	1.06	1.11	1.72	2.82	2.92
ISS_G	13583	2.19	0.32	1.02	1.26	2.26	2.96	3.39
EK_S	13583	40.59	15.74	0.36	3.23	45.44	78.50	96.85
EK_G	13583	48.09	16.58	0.42	6.24	50.58	83.61	97.31
EK_E	13583	27.13	19.02	0.00	0.00	30.79	71.08	92.29
CBF	13583	-3.34×10^{3}	1.06×10^{4}	-8.94×10^{5}	-4.94×10^{4}	-819.40	-0.16	1.21
CBF_INT	13583	-2.75	10.66	-556.10	-35.65	-0.44	-1.00×10^{-5}	6.00×10^{-5}
SBTI	13583	0.03	0.28	0.00	0.00	0.00	2.00	3.00
II. Fundamental & Market Indicators								
BLG_RETURN	13583	11.52	38.55	-79.48	-57.40	6.30	128.14	426.00
BLG_EVIC	13583	6.39×10^{3}	2.46×10^{4}	74.41	153.41	2.20×10^{3}	6.12×10^{4}	8.14×10^{5}
BLG_SALES	13583	5.48	1.39	0.83	2.41	5.42	9.00	11.25
BLG_PE	13583	62.89	563.41	0.42	3.72	19.94	603.11	2.68×10^{4}
BLG_PS	13583	4.18	27.69	0.03	0.13	1.88	26.13	1.61×10^{3}
BLG_PB	13583	1.31	0.67	0.10	0.34	1.17	3.63	6.04
BLG_CAP	13583	7.40	1.22	3.20	4.76	7.35	10.46	12.79
BLG_DE	13583	3.54	1.64	0.00	0.00	3.92	6.82	9.66
BLG_LT_DEBT	13583	1.23×10^{3}	5.17×10^{3}	0.00	0.00	188.60	1.99×10^{4}	1.47×10^{5}
BLG_PROFIT	13583	-25.15	1.34×10^{3}	-8.19×10^{4}	-166.55	7.84	90.63	583.82
BLG_IC_RATIO	13583	568.83	5.90×10^{3}	-1.55×10^{4}	-381.15	23.25	6.73×10^{3}	3.34×10^{5}
BLG_RD_EXP	13583	1.15	1.37	0.00	0.00	0.57	4.99	8.11
BLG_FCF	13583	42.48	223.34	-3.21×10^{3}	-282.35	14.31	725.39	4.90×10^{3}
BLG_LT_INV	13583	1.68×10^{3}	1.78×10^{4}	0.00	0.00	6.64	3.06×10^{4}	6.93×10^{5}
FE_REVUP_PCT	13583	57.87	28.32	0.00	0.00	60.00	100.00	100.00
III. Ownership & Debt Structure								
BLG_SHARE_OUT	13583	1.30	1.14	0.00	0.00	1.09	4.06	4.44
BLG_INSTI_SHARE	13583	4.01	0.77	0.00	1.58	4.20	4.98	5.70
FS_GREEN_SHARE	13583	0.22	1.00	0.00	0.00	0.00	3.66	24.68
BLG_SHARE_REPUR	13583	14.23	54.61	-7.00×10^{-5}	0.00	2.41	203.98	2.58×10^{3}
FS_GB_ISSUE	13583	9.00×10^{-4}	0.02	0.00	0.00	0.00	0.00	0.93

Continued on next page

(Continued)

Parameter	count	mean	std	min	1%	50%	99%	max
FS_GB_PAID	13583	8.80×10^{-4}	0.02	0.00	0.00	0.00	0.00	0.93
IV. Regional Indicators								
OECD_CROSS	13583	1.61	1.95	0.00	0.00	0.00	6.63	7.31
OECD_INT	13583	1.63	1.88	0.00	0.00	0.00	6.79	8.29
OECD_SEC	13583	2.36	2.47	0.00	0.00	0.00	6.00	6.30
WB_GDP	13583	2.68	3.49	-11.17	-5.57	2.29	13.39	24.48
IEA_COAL	13583	0.06	0.09	0.00	5.89×10^{-3}	0.02	0.39	0.39
IEA_CRUDE	13583	2.32×10^{-3}	3.01×10^{-3}	0.00	0.00	1.39×10^{-3}	0.01	0.04
IEA_ELEC	13583	0.23	0.04	0.05	0.13	0.22	0.33	0.48
IEA_HEAT	13583	0.02	0.03	0.00	0.00	3.91×10^{-3}	0.13	0.20
IEA_GAS	13583	0.18	0.07	0.00	0.03	0.20	0.31	0.35
IEA_OIL	13583	0.45	0.08	0.24	0.24	0.48	0.66	0.71
IEA_RENEWABLE	13583	0.07	0.05	0.00	0.00	0.06	0.25	0.64

Panel B: Reported

Parameter	count	mean	std	min	1%	50%	99%	max
I. Sustainability Indicators								
ISS_E	14583	1.81	0.37	1.00	1.03	1.76	2.80	3.51
ISS_S	14583	1.85	0.32	1.02	1.11	1.78	2.82	3.32
ISS_G	14583	2.30	0.41	1.00	1.27	2.29	3.32	3.66
EK_S	14583	61.40	18.84	0.89	13.89	60.56	95.21	98.20
EK_G	14583	59.30	19.24	0.10	12.51	58.56	94.19	99.45
EK_E	14583	57.91	21.05	0.00	7.15	58.08	95.80	99.13
CBF	14583	-3.91×10^{3}	4.50×10^{4}	-5.32×10^{6}	-4.51×10^{4}	-882.48	-0.10	377.17
CBF_INT	14583	-1.00	4.37	-272.35	-14.35	-0.09	-1.00×10^{-5}	6.69×10^{-3}
SBTI	14583	0.86	1.16	0.00	0.00	0.00	3.00	3.00
II. Fundamental & Market Indicators								
BLG_RETURN	14583	10.89	34.45	-79.48	-54.03	7.71	115.10	424.15
BLG_EVIC	14583	3.08×10^{4}	7.94×10^{4}	87.75	418.22	9.08×10^{3}	3.87×10^{5}	1.75×10^{6}
BLG_SALES	14583	7.20	1.45	1.76	3.97	7.17	10.58	11.77
BLG_PE	14583	33.02	175.11	0.10	3.23	17.48	269.24	1.44×10^{4}

Continued on next page

(Continued)

Parameter	count	mean	std	min	1%	50%	99%	max
BLG_PS	14583	7.24	258.02	0.03	0.12	1.32	15.90	1.67×10^4
BLG_PB	14583	1.18	0.65	0.10	0.29	1.04	3.44	6.04
BLG_CAP	14583	8.69	1.45	3.64	5.66	8.63	12.25	14.19
BLG_DE	14583	4.04	1.33	0.00	0.00	4.19	6.98	9.59
BLG_LT_DEBT	14583	7.01×10^3	2.15×10^4	0.00	0.00	1.46×10^3	1.08×10^5	3.21×10^5
BLG_PROFIT	14583	13.12	344.85	-3.72×10^3	-62.92	6.88	108.70	2.32×10^4
BLG_IC_RATIO	14583	156.17	934.54	-1.41×10^3	-25.95	11.37	2.84×10^3	7.75×10^4
BLG_RD_EXP	14583	1.98	2.03	0.00	0.00	1.67	7.24	8.83
BLG_FCF	14583	288.50	919.69	-6.10×10^3	-806.30	69.93	4.29×10^3	1.56×10^4
BLG_LT_INV	14583	1.00×10^4	6.09×10^4	0.00	0.00	66.49	2.68×10^5	1.24×10^6
FE_REVUP_PCT	14583	58.79	24.64	0.00	0.00	60.00	100.00	100.00
III. Ownership & Debt Structure								
BLG_SHARE_OUT	14583	0.61	0.88	0.00	0.00	0.21	3.83	4.44
BLG_INSTI_SHARE	14583	4.10	0.54	0.00	2.54	4.17	4.98	5.98
FS_GREEN_SHARE	14583	0.19	0.63	0.00	0.00	0.04	2.16	24.68
BLG_SHARE_REPUR	14583	81.32	373.74	-0.39	0.00	3.53	1.43×10^3	1.04×10^4
FS_GB_ISSUE	14583	0.03	1.14	0.00	0.00	0.00	0.27	79.59
FS_GB_PAID	14583	0.03	1.14	0.00	0.00	0.00	0.26	79.59
IV. Regional Indicators								
OECD_CROSS	14583	2.87	2.39	0.00	0.00	3.23	7.30	7.31
OECD_INT	14583	2.88	2.39	0.00	0.00	3.40	8.14	8.34
OECD_SEC	14583	3.45	2.48	0.00	0.00	4.85	6.30	6.30
WB_GDP	14583	1.98	3.92	-11.17	-10.36	2.22	13.39	24.48
IEA_COAL	14583	0.04	0.05	0.00	4.87×10^{-3}	0.02	0.25	0.36
IEA_CRUDE	14583	1.80×10^{-3}	3.88×10^{-3}	0.00	0.00	1.80×10^{-4}	0.03	0.04
IEA_ELEC	14583	0.23	0.05	0.05	0.16	0.22	0.46	0.48
IEA_HEAT	14583	0.02	0.03	0.00	0.00	4.09×10^{-3}	0.16	0.20
IEA_GAS	14583	0.18	0.08	0.00	0.02	0.17	0.34	0.35
IEA_OIL	14583	0.45	0.06	0.25	0.26	0.46	0.62	0.71
IEA_RENEWABLE	14583	0.07	0.05	0.00	0.02	0.06	0.27	0.64

Table 10: **Summary statistics of predictors - Modeled and Reported**

6.3 OLS with individual fixed-effects

	GHG (log)		ΔGHG		INT (log)		ΔINT	
	estimated	reported	estimated	reported	estimated	reported	estimated	reported
ISS_E	4.499	−0.011	4.443	−0.031	1.072	0.073	1.306	0.107
ISS_S	−4.753	1.665	−4.276	1.735	1.468	1.785	0.810	1.278
ISS_G	0.492	−5.234	0.413	−5.051	−3.860	−6.839	−3.606	−6.271
CBF	−0.441	−0.847	−0.968	−0.400	1.070	5.071	0.320	5.301
CBF_INT	2.372	2.045	1.863	1.558	−6.175	−5.338	−6.034	−5.521
BLG_RETURN	2.211	−3.787	2.334	−3.338	1.877	−5.239	0.378	−5.988
BLG_PE	−0.046	1.042	0.267	0.659	1.551	1.098	1.230	0.944
BLG_PS	−0.816	0.363	−1.159	0.450	0.237	0.321	0.251	0.623
BLG_PB	1.914	−0.005	2.205	0.086	1.121	0.189	1.642	1.025
BLG_DE	−2.904	0.697	−3.241	0.671	−5.426	−0.628	−5.819	−0.918
BLG_LT_DEBT	−0.847	0.989	−1.225	1.668	−1.302	−0.740	−1.569	−0.047
BLG_PROFIT	−0.705	−0.631	−1.150	−0.675	0.005	−0.058	0.150	−0.738
BLG_IC_RATIO	−1.325	1.462	−1.661	1.892	−2.280	−0.033	−2.646	0.115
BLG_RD_EXP	−0.043	1.262	−0.265	0.592	0.518	−0.073	0.601	−0.287
BLG_FCF	0.106	−0.126	−0.029	0.371	−0.432	−1.618	−0.256	−1.461
BLG_LT_INV	−0.049	−2.854	0.392	−2.615	0.476	−2.064	0.773	−1.993
FE_REVUP_PCT	2.297	0.066	1.466	0.184	−1.646	1.282	−2.278	0.569
BLG_SHARE_OUT	0.505	−0.103	0.141	−0.415	0.613	0.472	0.095	0.420
BLG_INSTI_SHARE	1.255	−1.517	1.417	−1.679	2.933	−1.012	3.256	−1.156
FS_GREEN_SHARE	0.195	0.815	0.149	0.872	−1.151	0.147	−0.893	0.336
BLG_SHARE_REPUR	0.095	1.455	0.446	1.539	0.497	0.748	0.759	1.253
FS_GB_PAID	−0.343	0.408	−0.484	0.356	−0.396	0.303	−0.156	0.280
WB_GDP	1.398	1.345	1.075	1.428	2.198	0.902	2.087	1.608
IEA_COAL	−2.662	−1.103	−3.154	−0.456	3.583	−0.520	2.595	−0.888
IEA_OIL	−0.592	2.832	−1.124	2.364	0.958	2.301	0.582	1.259
SBTI	1.739	−3.894	1.287	−3.010	1.805	−4.219	1.232	−3.764
GHG_S12_LOG	80.324	65.798	−17.109	−21.542	15.762	12.338	15.158	10.132
GHG_S12_PCTD	−6.476	−5.110	−7.330	−6.363	−2.712	−0.828	−1.760	−0.224
INT_S12_LOG	−2.953	−1.222	−3.264	−1.479	44.868	41.645	−31.088	−29.475
INT_S12_PCTD	−1.586	−3.760	−1.453	−3.083	−3.759	−6.317	−5.205	−7.419

Table 11: **Panel model.** 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.

6.4 LASSO surviving rates without fixed effects

		Raw footprint					Relative change			
		GHG (log)		INT (log)			ΔGHG		Δ INT	
rank		est.	rep.	est.	rep.		est.	rep.	est.	rep.
1	INT_S12_LAG	0	34	99	99	WB_GDP	94	99	90	77
2	GHG_S12_LAG	98	98	25	7	FE_REVUP_PCT	87	73	99	99
3	WB_GDP	35	30	30	23	EK_S	75	97	59	89
4	FE_REVUP_PCT	20	0	42	44	OECD_CROSS	75	88	79	77
5	EK_S	8	17	1	31	IEA_COAL	71	89	81	76
6	BLG_PB	30	0	0	21	BLG_PB	99	75	54	75
7	ISS_E	17	0	34	0	EK_E	52	96	64	86
8	IEA_COAL	0	20	7	23	BLG_LT_INV	66	76	79	72
9	OECD_CROSS	0	20	10	20	INT_S12_LAG	66	52	95	80
10	BLG_DAILY_RETURN	14	0	10	24	GHG_S12_LAG	95	84	60	49
11	CBF_GHG	15	0	24	9	ISS_E	80	67	88	54
12	BLG_FCF	0	0	15	23	CBF_GHG	78	60	80	68
13	BLG_LT_INV	0	0	17	19	BLG_FCF	66	63	78	78
14	BLG_INSTI_SHARE	0	16	0	17	FS_PE	66	75	75	68
15	BLG_RD_EXP	21	0	8	0	BLG_SHARE_OUT	63	81	63	72

Table 12: **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 **not** included in the model.

6.5 Characteristics of decarbonizing firms intensities

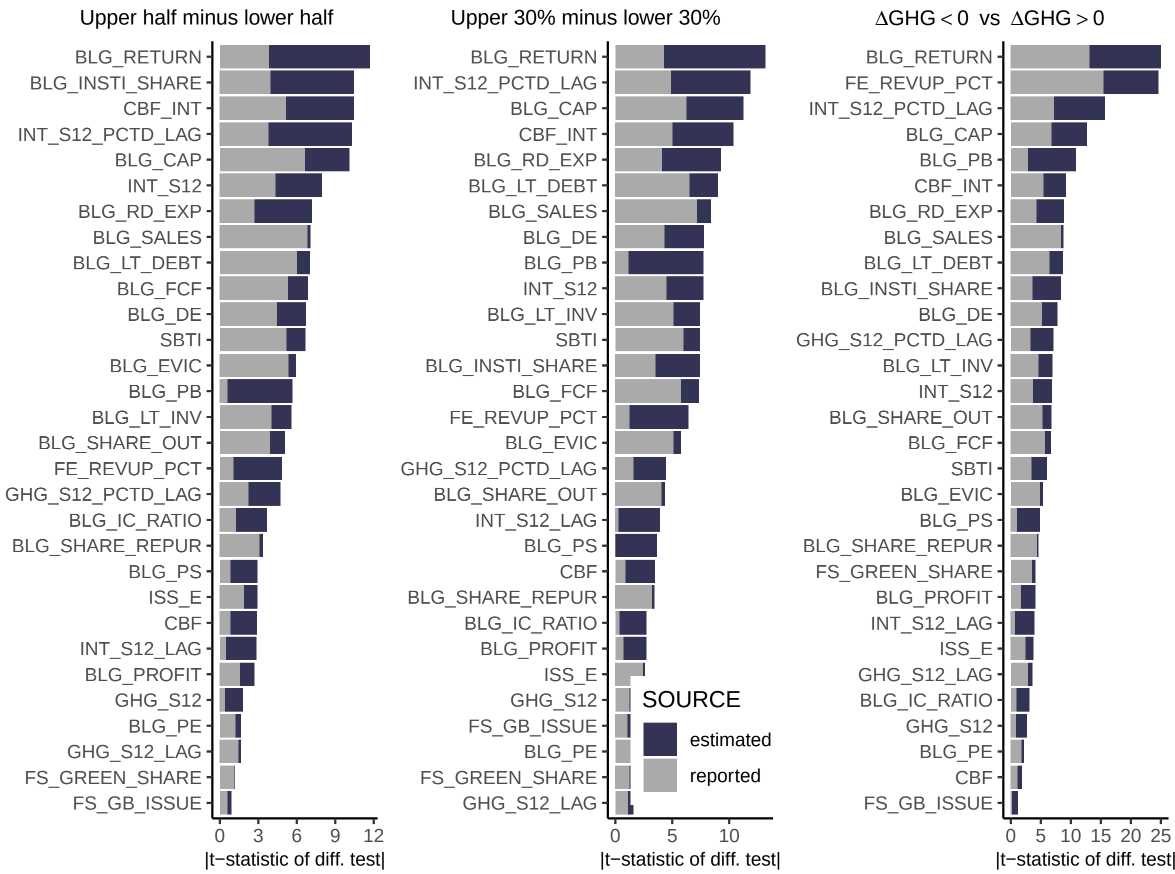

Figure 9: **Characteristics of (intensity) decarbonizing firms**. We plot the absolute values of the t-statistics of the tests in differences in means between characteristics with high versus low ΔINT.

6.6 Large errors from extrapolation

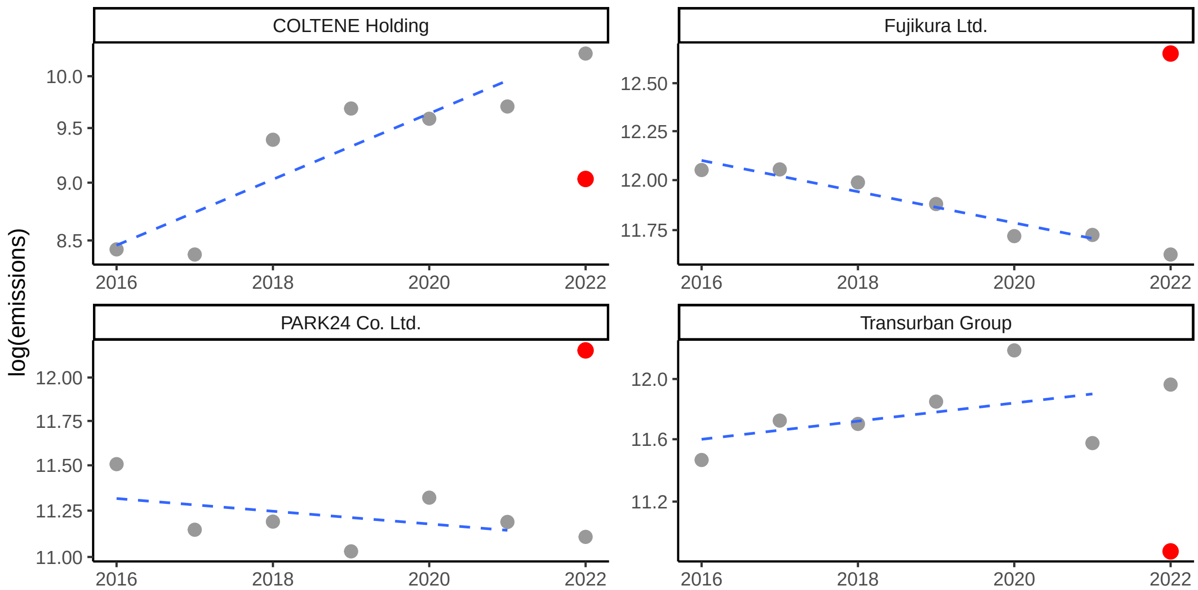

Figure 10: **Large errors from extrapolation**. We show the dynamics of log emissions for a sample of firms which have experience sharp breaks in 2022. The first six years are shown in grey and the last point is the extrapolation. The realized value for 2022 is shown in red.

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.

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.