Green Data or Greenwashing?

Do Corporate Carbon Emissions Data Enable Investors to Mitigate Climate Change?[A]

Vitali Kalesnik,[B] Marco Wilkens,[C] and Jonas Zink[D]

Research Affiliates, LLC

University of Augsburg

This draft: November 2, 2020

Abstract

Absent mandatory reporting, and although many companies report their carbon emissions, much of the emissions data are estimated by data providers. As we evaluate the forward-looking carbon scores from several popular data providers, we find no evidence that these scores predict future changes in emissions. Further, we find that data on estimated emissions are at least 2.4 times less effective than reported data in identifying the worst emitters and provide little information to identify green companies in brown sectors. Our results debunk the belief that third-party estimated emissions are a satisfactory substitute for company-reported emissions and call for mandatory and audited carbon emissions disclosure.

JEL Classification: G11, G14, Q51, Q54

Keywords: Carbon data, carbon databases, climate-related information, climate-related risks and opportunities, environmental impact of investments, carbon emissions, carbon intensities

---

[A] Version: October 25, 2020. First posted to SSRN: October 30, 2020. We appreciate the comments of Campbell Harvey and Alex Edmans as well as encouragement from Chris Brightman to explore this subject. Kay Jaitly and Jaynee Dudley provided editorial assistance.
[B] Vitali Kalesnik (corresponding author), Research Affiliates Global Advisors, London, Tel: +44 734 039 7975, E-mail: kalesnik@rallc.com
[C] Marco Wilkens, University of Augsburg, Faculty of Business Administration and Economics, Chair of Finance and Banking, Tel.: +49 821 598 4124, E-mail: marco.wilkens@wiwi.uni-augsburg.de.
[D] Jonas Zink, University of Augsburg, Faculty of Business Administration and Economics, Chair of Finance and Banking, Tel.: +49 821 598 4045, E-mail: jonas.zink@wiwi.uni-augsburg.de.

Strong evidence exists of increasing global temperature levels caused by the rise in greenhouse gas (GHG) emissions (IPCC, 2018). To reduce global warming, about 200 nations signed the Paris Agreement in an ambitious effort to combat climate change (United Nations, 2015). Investors also play an essential role in combating climate change. Investors, representing more than US$86 trillion assets under management (AUM), are signatories of the Principles for Responsible Investment (PRI), which aim, among other goals, to tackle climate change issues (PRI, 2019). Furthermore, investors who represent about US$35 trillion in AUM have joined Climate Action 100+, an organization more directly focused on coping with climate change (Climate Action 100+, 2019).

Investors employ multiple strategies to help mitigate climate change, such as switching investments from brown to green companies and engaging in activist measures. The efficacy of investor activity to incentivize the real economy to reduce carbon emissions vitally depends on the available GHG data. Some companies report their data voluntarily. For nonreporting companies, data providers attempt to close the availability gap by estimating carbon emissions. Estimated emissions by data providers often compose a large fraction of the data sets. Many investors view estimated emissions as a satisfactory substitute for company-reported emissions, thus revealing an implicit assumption in the status quo that data providers are successfully closing the data availability gap. In our study, we re-examine this assumption by analyzing the capacity of both reported and estimated corporate carbon emissions data to help mitigate climate change.

Our study refers to the ever-growing field of research on GHG emissions. Several studies analyze which factors contribute to a company's decision to disclose corporate GHG emissions (e.g., Prado-Lorenzo et al., 2009; Liesen et al., 2015) and how GHG emissions are priced in capital markets (e.g., Lee, Park, and Klassen, 2015, and Matsumura, Prakash, and Vera-Muñoz, 2014). In this context, some studies distinguish between mandatory or voluntarily disclosed GHG emissions (e.g., Busch and Lewandowski, 2018) and some analyze the effect of using different data providers (e.g., Busch et al., 2018, Berg, Koelbel, and Rigobon, 2019, and Li and Polychronopoulos, 2020). With this paper, we contribute to the existing literature strand by analyzing the various key GHG data available to investors and by examining how using these data can impact the efficiency of investor actions to combat climate change.

We develop a framework to evaluate carbon data. The framework stipulates five criteria in order for carbon data to be successful in mitigating climate change: 1) high data coverage, 2) comparability between companies, 3) consistency across data providers, 4) predictive power of forward-looking information, and 5) accuracy in reflecting true emissions. We use these criteria to compare the carbon data available to investors from four major carbon data providers. We label these providers as DP_A, DP_B, DP_C, and DP_D.[5] We show that these data sharply differ in terms of market-cap and emissions coverage. Moreover, the reported data do not follow a single standard, are often unaudited[6] and thus are not perfectly comparable between companies, and can be downward-biased. Voluntary reporting can lead to self-selection bias, because the likelihood of reporting rises with better carbon performance (Matsumura, Prakash, and Vera-Muñoz, 2014).

Beyond information on current and historical emissions, investors also gain valuable information that reveals future (planned) actions of companies to reduce or increase GHG emissions. Data providers attempt to capture this information from companies' forward-looking estimates. We examine if these forward-looking carbon data[7] are useful in forecasting companies'

[5] All data providers are anonymous in our study.
[6] Only about two-thirds of companies that externally report their emissions verify the data they report (Natixis, 2016).
[7] Data providers aggregate this information into carbon scores or carbon ratings, but do not estimate the amount of future carbon emissions in gigatons.

future changes in carbon emissions and find no predictability. This implies no evidence currently exists that these data sets capture sufficiently useful information for investors.

We also evaluate if estimated emissions by data providers accurately reflect true emissions. If investors use GHG data that lack accuracy, these data can lead to misidentification of brown companies as green companies and vice versa, and ultimately to greenwashing of investor activities (i.e., creating investment strategies that are likely to have only a limited impact). We demonstrate that currently available estimated data identify the worst 5% of emitters (currently responsible for 80% of overall emissions[8]), at least 2.4 times less efficiently than reported data. As a result, investor actions are at least 2.4 times less effective when based on the currently available estimated data. We further show that estimated emissions almost exclusively reflect the companies' operating industries and size-related information. This leads to an inability of investors to identify green companies in brown industries, which further reduces the efficacy of investor actions.

Overall, we argue that currently available estimated data are materially worse in quality compared to the reported data which reduces the effectiveness of investors in attempting to mitigate climate change. The prevailing data landscape leaves plenty of potential for greenwashing. We do not interpret our results as suggesting that data providers are sloppy estimators of various data sets. Instead, likely because of information asymmetry, these estimates are the best that data providers can make. Our results suggest needed change to the status quo, in which companies can voluntarily report their emissions forcing third-party data providers to close the data-availability gap with estimated data.

One solution could be an international regulatory initiative to make reporting of GHG emissions mandatory. Under such an initiative, we would advocate for the adoption of a single

[8] In our sample, the worst 300 emitters (5% of sample size) together account for 80% of total carbon emissions.

reporting standard (e.g., the GHG Protocol) for the measurement and reporting of current emissions (for current data) and ongoing investment projects (for forward-looking claims). To avoid greenwashing, carbon emissions should be audited to ensure data accuracy. Until GHG reporting becomes mandatory, investors should encourage companies to voluntarily report their emissions. One reporting method assumes the worst possible outcome, given the information available about the nonreporting company, and is consistent with the precautionary approach, Principle 15 of the 1992 Rio Declaration adopted by the United Nations (1992).

Description of Carbon Data Sets

We analyze carbon data from the four major carbon data providers (DP_A, DP_B, DP_C, and DP_D) for the seven-year period from 2010 to 2016.

The available data broadly fall into two categories: 1) historical (reported and estimated) GHG emissions and 2) carbon scores (and ratings)[9] that contain forward-looking information (e.g., emission reduction targets).[10] Both types of data are highly important for investors who wish to act to mitigate climate change. Historical emissions data help identify more- and less-GHG-efficient companies. Forward-looking data provide investors information about which companies are engaged in reducing emissions and which are increasing emissions.

We access three carbon databases that provide historical carbon emissions by emission category in accordance with the GHG Protocol (WBCSD and WRI, 2015b).[11] The fourth data provider, DP_D, solely provides carbon emission scores. The scores capture not just the raw emissions estimates, but several other types of (forward-looking) information that DP_D deems

[9] For example, an emission reduction score.
[10] More information about the forward-looking scores and ratings is in Appendix A.
[11] More information about the GHG Protocol is in Appendix B.

essential.[12] The score-based data of DP_D are not easily compared to the other data sets, so we exclude it from our analysis except when we examine the forward-looking data.

The GHG Protocol classifies reporting into three categories of carbon emissions: 1) direct GHG emissions (scope 1); 2) indirect GHG emissions (scope 2) composed of direct emissions from sources owned or controlled by the company and indirect emissions from the generation of purchased electricity consumed by the company; and 3) other indirect emissions (scope 3) for which reporting is optional. Scope 3 emissions arise from the activities of the company, but occur from sources not owned or controlled by the company (WBCSD and WRI, 2015a,b; and WBCSD and WRI, 2011). The scope 3 category often represents the largest source of GHG emissions, in some cases accounting for up to 90% of the total GHG impact.[13]

Very few companies report scope 3 emissions (only about 1,500 companies). For those that do report, scope 3 data are available on average for just 5 of the 15 scope 3 categories. Great variability exists in the number of categories companies report on, so that data comparability of scope 3 emissions is quite challenging. Moreover, scope 3 emissions are heavily subject to double counting. In certain cases, two or more companies may account for the same emissions within scope 3 (WBCSD and WRI, 2011).

Based on these considerations, we exclude scope 3 emissions from our analysis. Importantly, because GHG data offerings are rapidly evolving, the scope 3 data quality available to investors today could be of somewhat higher quality than the data we use in our study.[14] Our decision to exclude scope 3 data from our examination does not imply that investors should exclude it from their consideration. Our analysis focuses on the sum of scope 1 and scope 2 emissions,

[12] DP_D also publishes emissions data, but only in the form of scores, which are not comparable to the data provided by the other data providers.
[13] https://ghgprotocol.org/sites/default/files/standards_supporting/FAQ.pdf.
[14] We restrict our survey period from 2010 to 2016 because we have overlapping data for this period.

although many of the concerns we raise for these two emission types also apply to scope 3 emissions. Further, we only focus on carbon data from publicly listed companies.

Most data providers rely on self-disclosed information (e.g., from company websites or company sustainability reports) for gauging company-level GHG emissions. DP_C follows a different approach and annually sends an electronic questionnaire for companies to voluntarily complete and does not gather information from any other sources. Because the information collected through the DP_C questionnaire is very comprehensive and detailed, DP_A and DP_B also use DP_C data as a source for their own data sets. DP_A and DP_B complement the DP_C data with publicly available information. In Appendix C, we provide summary statistics for the data available from the various providers. DP_A only estimates carbon emissions jointly for the two emission categories (scope 1 and 2), whereas DP_B provides estimates separately for the categories. Companies reporting to DP_C are on average larger as indicated by higher net sales and higher number of employees.

Framework for Carbon Data Evaluation

What GHG emission-related data do investors need in order to mitigate climate change? We argue the GHG data should satisfy the following five criteria: 1) wide availability, 2) comparability between companies, 3) consistency across data providers, 4) forward-looking information should have predictive power, and 5) all data should accurately reflect true emissions. We use these criteria to examine the GHG data available to investors.

As with financial data, nonfinancial information needs to be *available for all investment opportunities* for investors to make sound investment decisions. Using four major carbon data sets, we test if GHG data are widely available for listed companies. We find that due to voluntary reporting, GHG reporting is still in its infancy.

For the reported GHG emissions data that are available, the data *need to be comparable between companies* to be of value for investors. We find that due to the lack of a uniform reporting standard and generous leeway within the existing standards, reported GHG emissions are not perfectly comparable between companies.

Usually, investors obtain ESG-related nonfinancial data from a specialized ESG data provider. Preferably, reported GHG emissions *would not differ among the data providers*. We find some discrepancies, however, between reported emissions from various data providers. Because every data provider applies its own estimation model, estimated emissions differ more strongly.

Many investors are interested not only in current levels of emissions, but also how companies will change their carbon emissions in the future. To provide this information, data providers have started to collect information about a corporation's anticipated future carbon performance (e.g., Does a company have emission reduction targets?). To be helpful for investors, this forward-looking information should have *power to predict future changes in carbon emissions*.

Finally, we argue that the GHG data should *accurately reflect the true emissions*. If investors act on noisy data they identify the polluting companies with less accuracy, which leads to investor actions that are ineffective.

We illustrate this by examining the ways investors can impact the real economy. Investors follow two main approaches to realize their goal of having a positive impact on the environment: 1) shifting capital from brown (*exit*) to green companies (e.g., selling brown stocks and buying green ones) and 2) using a shareholder activist approach (*voice, stewardship, engagement*). Dordi and Weber (2019) provide evidence that divestment announcements by investors targeting brown firms result in lower share prices for those companies.[15] Heinkel, Kraus, and Zechner (2001) argue that higher prices for brown companies result in a higher cost of capital that stifles brown activities. Further, Rohleder, Wilkens, and Zink (2020) show that concentrated divestment actions by mutual funds toward brown firms result in a change in the carbon-emitting behavior of affected firms. Conversely, providing cheaper capital to green companies has the effect of lowering the cost of capital and helps to close funding gaps. Further, because management compensation is linked to share prices (Edmans, Gabaix, and Jenter, 2017), the downward price pressure incentivizes managers to align their company's operations with investors' interests.

The second approach of investors who wish to drive positive change in the environment involves shareholder activism (*voice*). Investors have begun to engage with a company's management to push them toward a lower negative impact on the environment.[16] In contrast to the first approach, these investors aim to invest in brown companies and then try to exert pressure on the management from inside the company. Akey and Appel (2019) show that hedge fund activism targeting companies' environmental behavior is associated with a 17% drop in chemical emissions at the plants of targeted firms. The findings of Naaraayanan, Sachdeva, and Sharma (2019) support these results.

The impact channels of investors are severely limited. For instance, when investors sell shares of brown companies to purchase the shares of green ones, they are not denying the brown company access to capital. Given that for every seller there is a buyer, these actions simply reallocate ownership. Only if the selling pressure is extremely high, putting downward pressure on

[15] Bolton and Kacperczyk (2020) observe significant divestment transactions of institutional investors. Similarly, Boermans and Galema (2019) show that a sample of Dutch pension funds have been actively decarbonizing their portfolios.

[16] Often, investors must push companies to even disclose their environmental practices (Cotter and Najah, 2012, and Flammer, Toffel, and Viswanathan, 2019).

the share price, will the brown company face a higher cost in raising new capital. In contrast, green firms benefit from access to cheaper capital. Because substantial reallocation is required to move the cost of capital, more-concentrated and higher-volume investor actions are necessary to achieve a strong impact. Further, GHG emissions are highly concentrated. According to CDP (2017), 71% of all global GHG emissions since 1988 can be traced to just 100 fossil fuel producers, whereas only 5% of the entire universe of emitters is responsible for 80% of emissions. Therefore, because the worst emitters are responsible for the lion's share of emissions and because investor actions quickly lose efficacy if applied in an unfocused manner, it is critical for investors to have accurate emissions data.

Another important consideration is the type of information the GHG emissions data are based on. Although emissions are concentrated in certain industries, investors need to distinguish between companies within a brown industry and to identify the green(er) companies. When investors act on industry information alone, they indiscriminately penalize the whole industry and thus punish green(er) companies within the industry. This stifles the green projects in the brown sectors—the opposite of what investors desire. Consequently, companies may be discouraged from pursuing green investment if the investment is not rewarded—and may even be punished—by investors.

GHG Data Coverage

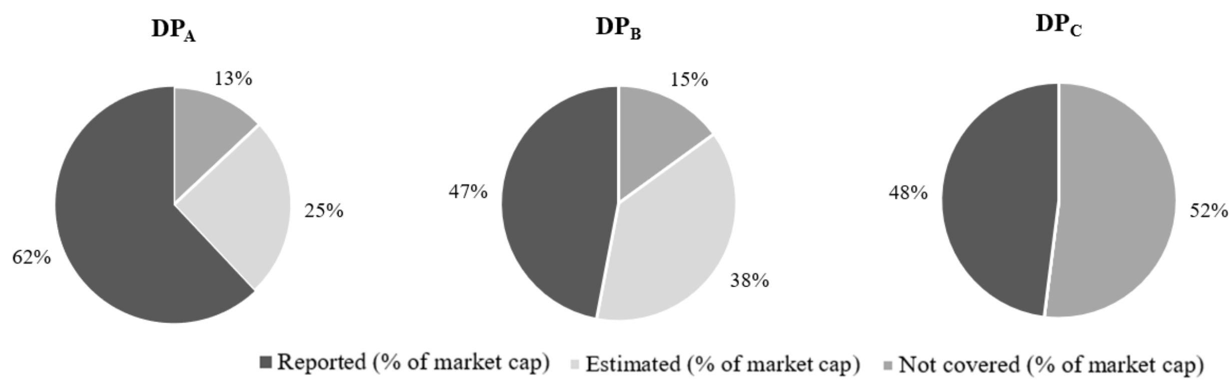
Until now, GHG reporting has been voluntary in most jurisdictions. Only in very few instances is reporting mandatory (e.g., the company falls under the scope of an emission trading system).[17] If companies do report to regulators, the data are often hard to access and not investor friendly.[18] The

[17] The Canadian government requires large businesses affected by COVID-19 to publish an annual climate-related disclosure report consistent with TCFD if they obtain state aid (Trudeau, 2020).
[18] For instance, companies that fall under the EU Emissions Trading System report their carbon emissions at the plant level, not at the firm level.

voluntary basis for reporting significantly lowers the data coverage and introduces a potential self-selection bias (Matsumura, Prakash, and Vera-Muñoz, 2014). We argue that carbon emissions are the most useful for investors if they are widely available in the investment universe. To test this, we compare the GHG data coverage of different carbon data providers by 1) market capitalization and 2) amount of carbon emissions.

Market-Capitalization Coverage. In **Figure 1**, we display the breakdown of the market-capitalization coverage of GHG data for the three data providers that provide carbon emissions for the seven-year period 2010–2016 (Appendix D provides more detail by year).

Figure 1. Comparison of Market-Capitalization Coverage with GHG Emissions, 2010–2016

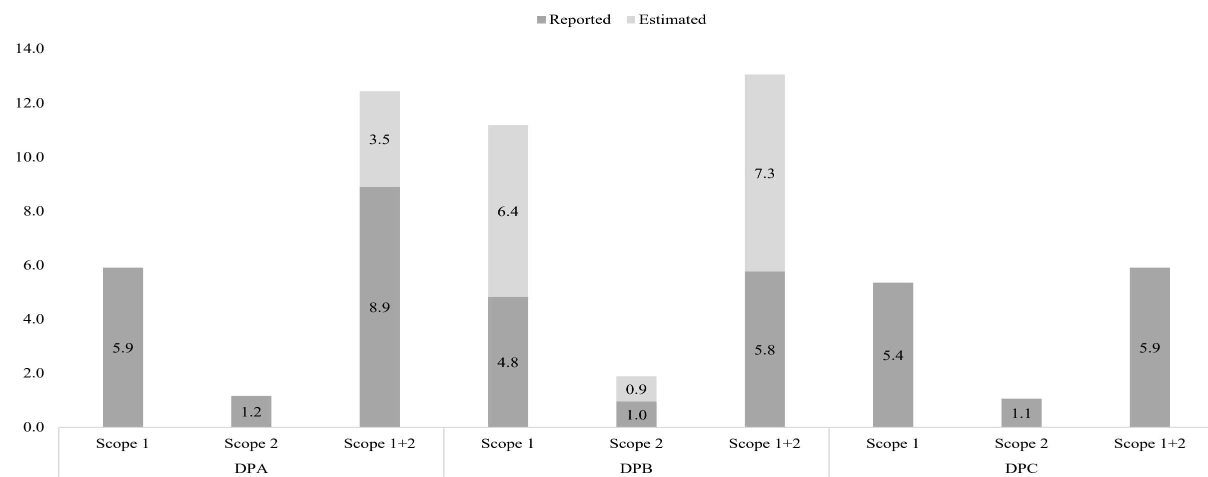


Source: Research Affiliates, LLC, and University of Augsburg, based on anonymized data from GHG emissions data providers. Note: This figure compares the market-capitalization coverage across the GHG data providers. *Reported (% of market cap)* represents the fraction of market capitalization that is on average covered with reported GHG data in the respective data set. *Estimated (% of market cap)* represents the fraction of market capitalization that is on average covered with estimated GHG data in the respective data set. *Not covered (% of market cap)* represents the fraction of market capitalization that is on average *not* covered with any GHG data in the respective data set. The numbers reflect the time-series average of cross-sectional means for the period 2010–2016.

Overall, DP_A captures reported carbon emissions for around 62% of all listed companies by market capitalization. DP_B's and DP_C's coverage is significantly smaller at 47% and 48%, respectively. To complement the data sets, DP_A and DP_B use models to estimate emissions for 25% and 38% of the companies, respectively. (We analyze the quality of the estimated emissions later in the paper.) We do not have estimated data for DP_C.[19] Consequently, the remaining fractions of market capitalization not covered with GHG data by DP_A and DP_B are 13% and 52%, respectively. The insufficient coverage of GHG data can be attributed to the voluntary reporting of climate-related information.

Carbon Emissions Coverage. Green investors who aim to mitigate climate change desire to reduce the overall GHG emissions of their investments. To achieve the greatest impact possible, GHG data must be available for most investments, and especially for the largest emitters. Heavily emitting companies have the greatest potential for emission reductions. We compare the data sets by their respective carbon emissions coverage (in metric gigatons[20]). In **Figure 2**, we display the average annual amount of covered reported and estimated emissions by emission category (scope) for each data provider. We exclude from the figure those companies not covered with GHG data.

Figure 2. Coverage Comparison by Total GHG Emissions, 2010–2016



Source: Research Affiliates, LLC, and University of Augsburg based on anonymized data from GHG emissions data providers.
Note: This figure compares the total amount of covered carbon emissions by data set and emission category. *Reported* represents the total amount of covered reported carbon emissions (in gigatons) and *Estimated* represents the total amount of covered estimated carbon emissions (in gigatons). The numbers reflect the time-series average of annual total covered emissions for the period 2010–2016.

---

[19] DP_C began to estimate emissions after our period of analysis.

[20] A gigaton is 10^9 tons.

DP_A and DP_C cover a similar amount of reported scope 1 emissions at 5.9 and 5.4 gigatons, respectively. Recall that DP_B separately estimates scope 1 and scope 2 emissions. Therefore, DP_B reports much higher total covered direct emissions at 11.2 gigatons. The reported scope 2 emission coverage is similar for all three data sets, but slightly higher for DP_B after we include estimated emissions. Considering both emission categories together (scope 1 and 2), DP_A displays the highest reported carbon emissions coverage (50% more than DP_C), whereas DP_C and DP_B show similar coverage of 5.9 and 5.8 gigatons, respectively. Finally, DP_B estimates the most scope 1 and 2 emissions at 7.3 gigatons.

To conclude, green investors desire access to a data set that covers as many emissions as possible. From this perspective, DP_A is superior to the other analyzed data sets if we exclude estimated emissions, which we examine later in more detail.

Comparability between Companies

No universally accepted reporting standard for GHG emissions exists. The Task Force on Climate-Related Financial Disclosures (TCFD) was founded to provide guidance on which information is relevant to investors, lenders, insurers, and other stakeholders, and how the information should be disclosed by reporting companies.[21] In terms of the measurement and disclosure of carbon emissions, the TCFD refers to the GHG Protocol, which for many years has been one of the most commonly followed standards, although other reporting standards exist (e.g., US EPA Center for Corporate Climate Leadership GHG Inventory Guidance, UK Environmental Reporting Guidelines, and China NDRC GHG accounting and reporting guidelines).[22] The ability to choose

[21] https://www.fsb-tcfd.org/

[22] For the US EPA Center for Corporate Climate Leadership GHG Inventory Guidance see https://www.epa.gov/climateleadership/center-corporate-climate-leadership-greenhouse-gas-inventory-guidance. For the UK Environmental Reporting Guidelines see: https://www.gov.uk/government/publications/environmental-reporting-guidelines-including-mandatory-greenhouse-gas-emissions-reporting-guidance. For the China NDRC GHG

the reporting standard potentially introduces a self-serving bias in the reported data and hinders comparability between companies. In 2018, only 33% of companies reported GHG emissions in line with TCFD recommendations (TCFD, 2019). Furthermore, the academic ClimateDisclosure100.info initiative recognizes only 21 firms worldwide to have reported 100% of their scope 1 GHG emissions.[23]

Finally, the standards themselves often give companies significant leeway in measuring and disclosing their emissions. For instance, the GHG Protocol allows for reporting carbon emissions using either the equity share or the financial control approach.[24] As a result, reported carbon emissions can significantly differ based on the approach used.

Consistency across Data Providers

Most investors obtain their carbon data from a specialized carbon data provider. Each data provider has unique features, which can potentially lead to different investment decisions. For instance, data providers deal differently with corporate events, such as mergers and acquisitions. Some adjust carbon emissions for corporate actions (e.g., DP_A and DP_B claim to do so), while others do not (DP_C). Further, some data providers correct obvious reporting errors (e.g., use of the wrong unit), while others do not.

More discrepancies arise due to the variation in treatment of reported scope 2 emissions. Specifically, the GHG Protocol allows companies to use the market-based or location-based

accounting and reporting guidelines see: https://www.wri.org/our-work/top-outcome/china-moves-toward-mandatory-corporate-ghg-reporting.

23 https://climatedisclosure100.info/

24 Under the equity share approach, a company accounts for GHG emissions from operations according to its share of equity in the operation. Under the control approach, a company accounts for 100% of the GHG emissions from operations over which it has control (WBCSD and WRI, 2015b).

approach to calculate their scope 2 emissions.[25] Often, companies report emissions compliant to both approaches, but some data providers report only one approach to investors.[26] The choice of reporting appears arbitrary and leads to considerable differences in reported scope 2 emissions between data providers.[27]

Our access to multiple carbon data sets allows us to test if reported carbon emissions data are consistent across data providers. Although consistency itself does not guarantee good quality data, it serves as an indirect proxy for data accuracy. To evaluate the consistency of emissions across data sets, we compare rank correlations of emission levels. Ranks are very important because many investment companies and initiatives encourage the exclusion of, for instance, the 200 most polluting companies from their portfolios (negative screening).

In **Table 1**, we compare emissions pairwise from two data providers if both cover the respective company in the respective year. In addition, we distinguish between the scope of emissions and if the emissions were self-reported by the companies or estimated by the data providers (Panel D).

[25] The location-based method quantifies scope 2 emissions based on average energy-generation emission factors for defined geographic locations. The market-based method quantifies scope 2 emissions based on GHG emissions emitted by the generators from whom the reporting company contractually purchases electricity (WBCSD and WRI, 2015a).
[26] Our analysis suggests that DP_A is more likely to report scope 2 emissions in accordance with the market-based approach as opposed to DP_B, which uses the location-based approach.
[27] We compare scope 2 emissions for random observations and find the reported emissions differ by more than 30%.

Table 1. Correlation of Reported and Estimated GHG Emissions between Data Providers, 2010–2016

Panel A: Reported scope 1 emissions

	DP_A	DP_B	DP_C
DP_A	1		
DP_B	0.987	1	
DP_C	0.986	0.995	1

Panel B: Reported scope 2 emissions

	DP_A	DP_B	DP_C
DP_A	1		
DP_B	0.976	1	
DP_C	0.979	0.993	1

Panel C: Reported scope 1+2 emissions

	DP_A	DP_B	DP_C
DP_A	1		
DP_B	0.984	1	
DP_C	0.984	0.996	1

Panel D: Estimated scope 1+2 emissions

	DP_A	DP_B	DP_C
DP_A	1		
DP_B	0.847	1	
DP_C	--	--	1

Source: Research Affiliates, LLC, and University of Augsburg based on anonymized data from GHG emissions data providers.
Note: This table shows rank correlations between various emission categories and between carbon data sets for the period 2010–2016. The number of observations varies within and across each panel.

Table 1 suggests that the ranks are relatively consistent across GHG data sets for the reported data: the rank correlations are almost 1 between DP_B and DP_C across all emission categories. The rank correlations of scope 2 emissions between DP_A and DP_C and between DP_A and DP_B are lower due to their different treatments of scope 2 emissions (location based versus market based). Estimated carbon emissions are more inconsistent across data providers (0.847) than reported emissions and can be attributed to the differences in their estimation models.[28] Busch et al. (2018) support our findings as they observe similar rank correlations.[29]

In summary, reported data are largely consistent across data providers. Estimated data exhibit large inconsistencies between data providers and the information content of estimates deserves careful examination.

[28] Many studies support our results that data providers often disagree in their examination of a company's sustainability (e.g., Dimson, Marsh, and Staunton, 2020, and Berg et al., 2019).
[29] In Appendix E, we conduct another analysis in which we calculate percentage deviations between the data providers. In this analysis, we find some inconsistencies, but they do not lead to significant changes in rank.

Predictive Power of Forward-Looking Information

Investors who desire to mitigate climate change are particularly interested in which companies plan to reduce their future emissions. Access to a company's forward-looking information is therefore precious. Data providers have also started to supply data with information on how companies are expected to change their carbon-emitting behavior in the future (e.g., based on corporate emissions reduction targets). This information is not supplied to investors in terms of future estimates of carbon emissions (in metric gigatons), but rather through a score or rating. We examine if these data are helpful for investors in predicting future *levels* and *changes* in emissions and if carbon scores with forward-looking information have any predictive power, as often claimed by data providers (e.g., emission reduction score). (Appendix A describes the carbon scores we use in our analysis).

Predictability of Future Reported Emission *Levels.* We expect carbon emission levels to be very persistent over time. We test this expectation by running regressions on future reported carbon emissions, which are explained by simple fixed effects as in predictive Models (1) and (2) and by historical emissions (autoregressive model) as in predictive Models (3) and (4). After calibrating the regression models, we studied how well investors can forecast future emission levels by examining rank correlations between model-predicted future emissions and real reported future emissions. In **Table 2**, we display the results.

Table 2. Predictability of Future Reported Carbon Emission Levels, 2010–2016

	(1)	(2)	(3)	(4)
	Log(Emissions$_{t+1}$)	Log(Emissions$_{t+1}$)	Log(Emissions$_{t+1}$)	Log(Emissions$_{t+1}$)
Log(Emissions$_t$)			0.98**	0.94**
Constant	14.37**	14.79**	0.27**	0.77**
	(27.55)	(17.45)	(8.15)	(6.27)
Industry fixed effects	Yes	Yes	No	Yes
Year fixed effects	No	Yes	No	Yes
Country fixed effects	No	Yes	No	Yes
No. of observations	11,984	11,984	11,984	11,984
Adjusted R^2	60.4%	65.4%	97.1%	97.2%
The rank correlation between predicted future emissions and real reported future emissions.				
	78.0%	81.0%	98.7%	98.7%

Source: Research Affiliates, LLC, and University of Augsburg based on anonymized data from GHG emissions data providers.
Note: This table reports regression results of past reported carbon emissions on real reported future carbon emissions (scope 1 and 2). All variables are logarithmized due to their skewness. *Emissions* reflect a company's scope 1 and 2 emissions. Industry fixed effects are based on a company's TRBC activity code. Country fixed effects are based on a country's ISO 3166 code. ∗∗ and ∗ indicate significance at the 1% and 5% level, respectively. Standard errors are clustered at the company level.

The adjusted R^2 of the regression using both fixed effects and historical emissions is 97.2%; that is, we can almost perfectly predict future emissions by a simple regression model. In addition, the rank correlation between predicted future emissions and reported future emissions is very high at 98.7%. Overall, our results suggest that carbon emissions are stable from one year to another and that investors are well served by taking the latest reported emissions as the proxy for next year's emissions. In this case, forward-looking carbon scores from the data providers are not needed.

Predictability of Future Reported Emission *Changes* Using Forward-Looking Information. Next, we test whether future changes in emissions are predictable and whether carbon scores from the data providers have explanatory power in predicting future changes in emissions.[30] Again, we use a regression approach, as shown in **Table 3,** to explain future changes in emissions by past changes (autoregressive model) and to add provider-specific carbon scores[31] with forward-looking information to the model.

Table 3. Predictability of Future Reported Carbon Emission Changes, 2010–2016

	Δ% Emissions$_{t+1,t}$	Δ% Emissions$_{t+1,t}$	Δ% Emissions$_{t+1,t}$	Δ% Emissions$_{t+1,t}$	Δ% Emissions$_{t+1,t}$	Δ% Emissions$_{t+1,t}$
Δ%Emissions$_{t,t-1}$		-0.04*	-0.04*	-0.04*	-0.04*	-0.04*
DP$_C$ A list$_t$			-0.09			-0.02
DP$_A$ emission reduction score$_t$				-0.00		-0.00
DP$_D$ carbon emission score$_t$					-0.08	-0.07
Constant	0.72	0.74	0.74	1.02*	0.94	1.04*
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	5,917	5,917	5,917	5,917	5,917	5,917
R^2	3.6%	3.7%	3.7%	3.7%	3.8%	3.8%
Adjusted R^2	-3.0%	-2.9%	-2.9%	-2.9%	-2.9%	-2.9%
Rank correlation between predicted changes in future emissions and real reported changes in future emissions.						
	6.6%	7.0%	7.2%	6.6%	7.0%	6.9%

Source: Research Affiliates, LLC, and University of Augsburg based on anonymized data from GHG emissions data providers.
Note: This table reports regression results of past reported carbon emission changes on real reported future carbon emission changes (scope 1 and 2). *Δ% Emissions$_{t+1,t}$* reflect the percentage change of a company's scope 1 and 2 emissions from one year to another. *DP$_C$ A list* reflect a binary variable, which specifies if a company was in the DP$_C$ A list in the respective year. Industry fixed effects are based on a company's TRBC activity code. Country fixed effects are based on a country's ISO 3166 code. ∗∗ and ∗ indicate significance at the 1% and 5% level, respectively. Standard errors are clustered at the company level.

Independent of the model specification, it is difficult to predict future changes in emissions as indicated by a low (negative) adjusted R^2. Even data provider–specific carbon emission scores with forward-looking content cannot overcome this problem. The rank correlation between model-predicted future changes in emissions and real reported future emissions changes for any model

[30] In Appendix F, we examine if changes in carbon emissions (Panel A) and carbon intensities (Panel B) are persistent over time. We find the likelihood of last year's trend (e.g., strong emissions reduction) continuing in the future is low.
[31] Appendix A gives an overview of carbon scores with forward-looking information.

specification is very low (below 8%). We repeat the analysis focusing on future changes in carbon *intensities* (Appendix G) and get similar results.

The bottom line is that future changes in carbon emissions are very hard to forecast. We find no evidence that the forward-looking scores from the data providers carry any useful forecasting information. We do not doubt that data providers are doing an honest job collecting data from companies. The lack of predictability is likely driven by the use of nonscientifically verified estimation methods and by "cheap talk" from companies engaged in greenwashing. We argue that this kind of information should be externally verified before being published. Until scientific verification demonstrates that a specific type of forward-looking data is predictive of future emissions changes, investors would be prudent to operate under the presumption these data contain no useful forward-looking information.

Accuracy of Estimated Emissions

Voluntary reporting results in reported data on GHG emissions being available for only 48−62% of companies based on market capitalization (see Figure 1). The data providers that estimate the missing data have limited visibility into companies' activities, implying that the estimated emissions are less likely to contain detailed company-specific information about companies' green efforts and activities. Nevertheless, data providers claim their estimated emissions are based on highly sophisticated estimation models and some have even patented their models.[32] These representations lead to a widespread misunderstanding that estimated emissions are of similar quality as reported emissions and can be used with equal confidence in decision making. In this section, we analyze the accuracy of estimated emissions.

[32] Appendix H gives an overview of the carbon estimation models.

We run several tests to assess the information content of the estimated data using reported data as the best available proxy for the true emissions. Estimated emissions are by definition a noisy proxy of true emissions, which are unobserved. Data providers often base their estimates on broad business metrics and industry affiliations, and to our best knowledge, do not engage with companies in estimating emissions. As described in Appendix H, data providers typically base their estimates on very simple financial information (e.g., net sales or number of employees).

Calibrating Model Accuracy. We test how much power simple financial business metrics and fixed effects have in explaining reported carbon emissions. We use the model R^2—a statistical metric representing the percentage of total emissions explained by the model-independent variables—as the measure of model accuracy. If the explanation power is very high (R^2 close to 1), this suggests that estimated emissions are likely to adequately represent a company's true emissions. We later assess how different levels of R^2 translate into investors' ability to identify the worst emitters and use this measure to assess how model accuracy can impact the efficacy of investor actions.

We set up various specifications of regression models to test if simple business metrics and fixed effects have any power in explaining reported emissions. We use DP_C's emissions in the model calibration as the dependent variable because DP_C collects the bulk of the reported data first hand and the other data providers use these data as the main source for their own data sets. We use quite simple financial indicators, such as the logarithm of net sales (to proxy for the size of a company's output), as independent variables. Next, we use a company's industry affiliation, country of domicile, and the year as fixed effects. (Appendix I shows that carbon emissions differ highly across countries and industries.) In Appendix J, we provide a detailed definition and rationale for each variable we use. We take logs of carbon emissions and business metrics due to their high skewness. We use an ordinary least squares regression to calibrate the models. Equation 1 is our baseline regression model,

$$\text{Log(Carbon emissions}_{i,t}^{\text{Reported}}) = \alpha + \sum_{j}^{n} \beta_j \text{ Log(Business metric}_{i,t}) + \sum_{k}^{m} \beta_k \text{ Fixed and time effects}_{i,t} + \varepsilon_{i,t} \qquad (1)$$

Table 4. Calibrating Model Accuracy Using Reported Carbon Emissions, 2010–2016

	(1) Log(Emissions$_t$)	(2) Log(Emissions$_t$)	(3) Log(Emissions$_t$)	(4) Log(Emissions$_t$)	(5) Log(Emissions$_t$)
Log(Net sales$_t$)	1.06**			1.10**	1.11**
Log(Employees/NS$_t$)					0.29**
Log(Market Cap/NS$_t$)					-0.18**
Log(EBT/NS$_t$)					-0.03
Log(FFO/NS$_t$)					0.28**
Log(OI/NS$_t$)					-0.08
Log(PP&E/NS$_t$)					0.43**
Log(COGS/NS$_t$)					0.18**
Constant	-10.96**	13.99**	13.94**	-9.94**	-6.47**
Industry fixed effects	No	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes
Country fixed effects	No	No	Yes	Yes	Yes
No. of observations	7,227	7,227	7,227	7,227	7,227
Adjusted R^2	45.7%	54.2%	61.2%	83.8%	87.0%

Source: Research Affiliates, LLC, and University of Augsburg based on anonymized data from GHG emissions data providers.
Note: This table displays panel regression results of various company characteristics on carbon emissions (scope 1 and 2). All variables are logarithmized due to their skewness. *Emissions* reflect a company's scope 1 and 2 emission reported to DP$_C$ for the period 2010–2016. The independent variables are described in Appendix J. Industry fixed effects are based on a company's TRBC activity code. Country fixed effects are based on a country's ISO 3166 code. ** and * indicate significance at the 1% and 5% level, respectively. Standard errors are clustered at the company level.

We display the model calibration results in **Table 4**. Models (1) through (4) show that company size and industry explain most of the company variations in reported emissions. Specifically, Model (1), which uses the logarithm of net sales as the proxy for company size, explains around half of the emission variation; the model adjusted R^2 is 45.7%. Similarly, Model (2), which uses only the industry fixed effects, has an adjusted R^2 of 54.2%. When we combine the measure of size, industry, and other fixed effects in Model (4), the adjusted R^2 rises to 83.8%.

In Model (5), we add a few more company-specific financial ratios as explanatory variables. The most important variables, as reflected by their *t*-stats, are 1) property, plant, and equipment (PP&E) value scaled by net sales (Log(PP&E/NS)), which proxies for the equipment intensiveness of production, and 2) number of employees scaled by net sales (Log(Employees/NS)), which measures the employee intensiveness of production. Both variables positively predict the variations in emissions, suggesting that the more-innovative and greener companies are also less well equipped and less employee intensive. Model (5) captures some information beyond the broad correlates: the adjusted R^2 increases from 83.8% for Model (4) to 87.0% for Model (5).

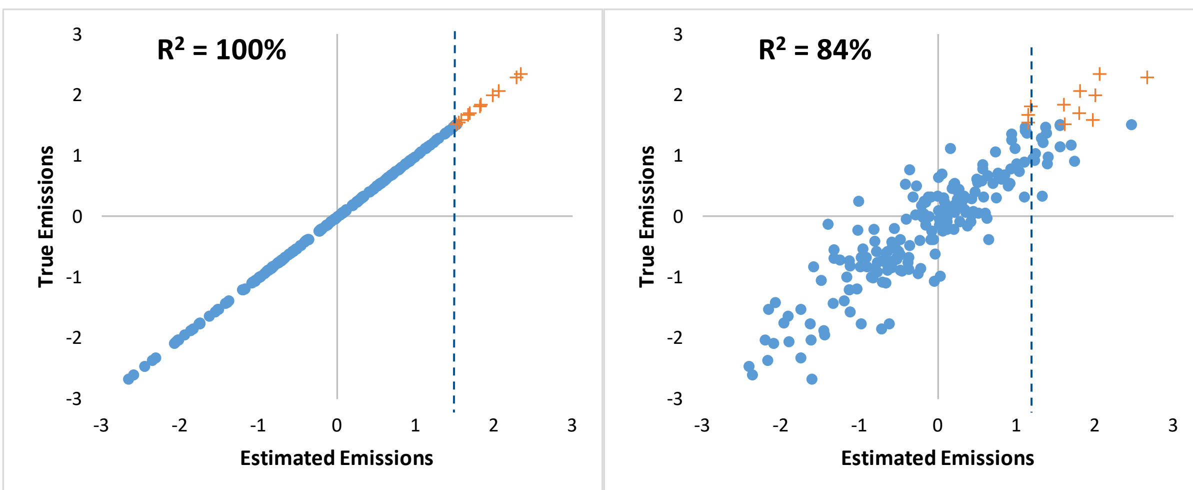
We conclude that industry and size characteristics play a major role in explaining carbon emissions. Models incorporating this type of information as well as information on a company's efficiency have excellent explanatory power with an R^2 approaching 100%. At the same time, we do not observe perfect explanatory power, which is relevant for the comparison with the reported data. The missing data may hold extremely relevant and valuable information, such as whether a company is greener than its peers in a specific industry (e.g., it uses green electricity). In the following section, we illustrate how using estimated emissions based on an R^2 lower than 100% can affect the impact of investor actions.

Model Accuracy and Investor Ability to Identify the Worst Emitters. Carbon emissions are highly concentrated. Only 5% of companies are responsible for 80% of total emissions. We raise the question of how precisely investors can identify these 5% largest emitters using noisy estimated data.

To understand the exercise, consider the following example. Suppose an investor wants to avoid investing in the 5% largest emitters in a universe of 10,000 companies and wishes to use estimated data to identify these 5% of heavy emitting companies. With perfect data (reported or estimated with an R^2 of 100%), an investor would be able to exactly identify the 500 worst companies (100% confidence). **Panel A** of **Figure 3** illustrates a simulated distribution of true and estimated emissions with an R^2 of 100%. In this case, estimated emissions allow the perfect identification of all 5% of the worst emitters (data points marked with crosses). If the data are noisy and do not perfectly reflect a company's true emissions (R^2 below 100%), an investor would need to exclude more companies from their portfolio in order to ensure that all of the 500 largest emitters would be excluded.

Figure 3. Simulated Distribution of True (Unobservable) and Estimated Emissions for Different Levels of R^2

Panel A **Panel B**

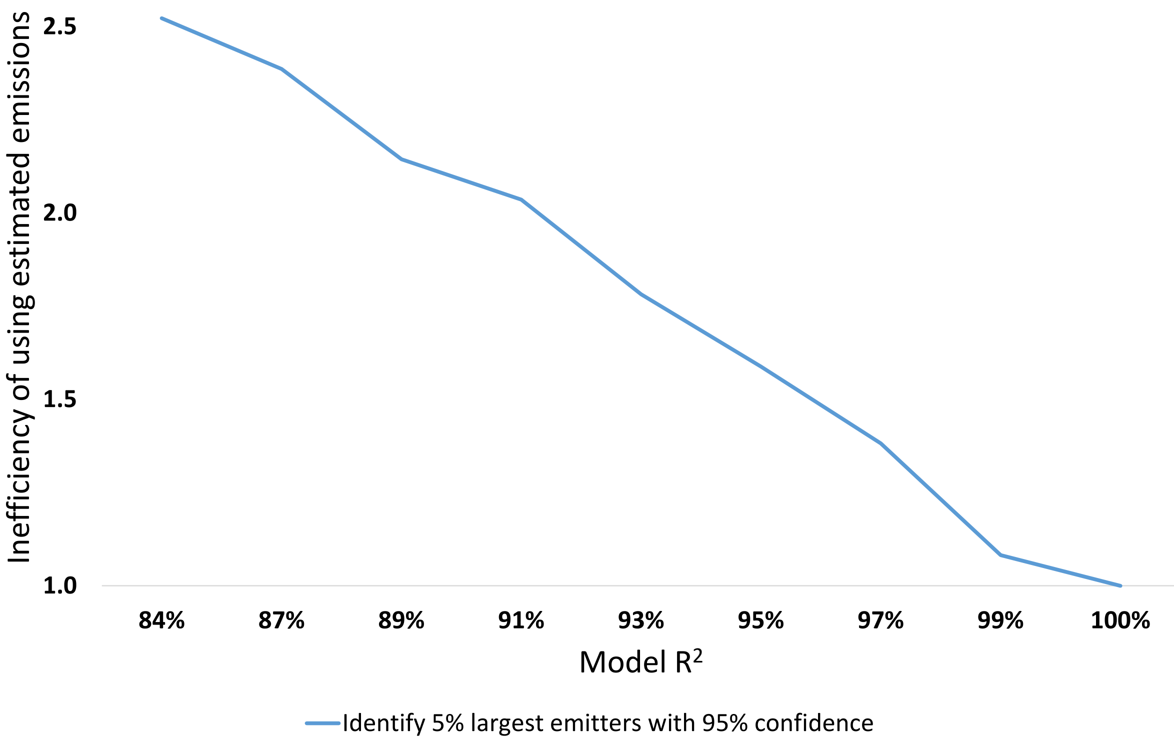


Source: Research Affiliates, LLC, and University of Augsburg.
Note: The figure shows the simulated distribution of normally randomly distributed true (unobservable) and estimated (observable) emissions. We mark with crosses the 5% worst emitters based on the true emissions. We also mark with the dashed line the cut-off value of the estimated emissions necessary to select at least 95% of the 5% worst emitters based on the estimated emissions.

To show the effects of using estimated data, we perform a Monte-Carlo simulation. We assume an R^2 of 84% as in Model (4) and calculate that an investor would need to exclude 1,250 companies to remove at least 95% of the largest emitters. In **Panel B** of Figure 3, to the right of the dashed line, the dots represent the less-severe emitters together with the crosses that represent the 5% worst emitters. In this case, the number of noninvestable stocks increases by 2.5 relative to the

500 stocks that would be excluded with perfect insight, reducing the effectiveness of the investor action by 2.5 times (i.e., the investor does not invest in 1,250 companies versus 500). We call 2.5 the *inefficiency of using estimated emissions* in comparison to reported emissions. We repeat our simulation for various levels of R^2 and display the results in **Figure 4**.

Figure 4. Inefficiency of Targeting the Largest Emitters Using Estimated Emissions



Source: Research Affiliates, LLC, and University of Augsburg.
Note: The figure shows the results of Monte-Carlo simulations with the sample size of the 10,000 using normally distributed random variables in all parts of the simulations. For a model with a given R^2, we estimate the size of the sample sufficiently large to identify the 5% largest emitters with the 95% confidence.

Compared to Model (4), using Model (5) with an R^2 of 87.0%, the estimates are not much better. An investor would need to pick about 2.4 times (compared to 2.5 times) as many stocks to exclude the 5% worst stocks, about 1,190 versus 500 using reported data. The higher R^2 of Model (5) lowers the inefficiency of the investor action, but only minimally. A model R^2 of 98% or higher is likely the point at which the multipliers begin to converge to 100%, and investors would need a sample of no more than 1.2 times to target the 5% worst emitting companies.

Accuracy and Information Content of Estimated Emissions by Data Providers. We use a unique subset of companies to run the horse race between the data of the two data providers, DP_A and DP_B, and the estimates obtained from Models (4) and (5). This subset of companies began reporting after a period of nonreporting, during which the data providers had been estimating the companies' emissions. Given the high persistence of emissions as described earlier in the paper, the subsequently reported data are an excellent proxy for historical emissions.[33]

We argue that if one data provider has superior information compared to the other, its estimates would be more highly correlated to the (later) reported emissions. To test this expectation, we compare correlations between the last known estimated emissions by the data providers and the first reported emissions by the companies. In **Table 5**, we report the rank correlations and log-level correlations between the different types of estimates. Because estimating emissions may be harder for smaller companies due to less available information, we show results for two different samples. Panel A shows correlations for observations in which both data providers switch from estimated to reported emissions at the same time (i.e., we compare the same companies). Panel B relaxes this condition and shows results for all observations within a data set. Our Model (4) and Model (5) predictions are benchmarked against reported DP_C emissions.

[33] We expect some variation between estimated and reported emissions because of the one-year time lag between the estimated (t) and the reported emissions ($t + 1$). Therefore, we focus on correlations rather than absolute deviations.

Table 5. Accuracy of Estimated and Model-Predicted Emissions, 2010–2016

	N	Rank correlations	Log-level correlations
Panel A: Data provider switches from estimated to reporting emissions at the same time			
DP_A	100	0.84	0.79
DP_B	100	0.78	0.75
Model (4)	100	0.83	0.79
Model (5)	100	0.87	0.84
Panel B: Data provider–specific accuracy			
DP_A	860	0.69	0.67
DP_B	1,479	0.79	0.79
Model (4)	1,900	0.78	0.75
Model (5)	1,900	0.83	0.80

Source: Research Affiliates, LLC, and University of Augsburg based on anonymized data from GHG emissions data providers.
Note: This table reports both rank correlations and log-level correlations between data providers' estimates on carbon emissions (or predicted values) (in t) and company-reported emissions (in $t + 1$) for the period 2010–2016. Panel A reports correlations for observations when DP_A and DP_B simultaneously switched from estimated emissions (in t) to reported emissions (in $t + 1$) in their data sets. Panel B ignores other databases and reports correlations for all observations in which a switch occurred from estimated to reported emissions in the respective data set.

All models have roughly similar correlations. Relative to Model (4), the estimates from data providers on average display similar, and in rare cases, higher correlation levels with subsequently reported emissions. Relative to Model (5), they display correlations a notch lower in all cases, but the difference is very small.

We cannot directly translate the correlations with the subsequently reported emissions into the model R^2 for the current emissions, which serves as our measure of model accuracy. We have, however, estimated the R^2 for Models (4) and (5) at 83.7% and 87.0%, respectively. Given that the correlations of data from DP_A and DP_B are on average similar to the correlations of Model (4) and lower than for Model (5), we can conclude that the equivalent R^2 should be on par with 83.7% and lower than 87.0%. That means that the estimates from data providers do capture a significant portion of the return variation. Our estimates in Figure 4 show that Model (5) estimates with an R^2

of 87% are about 2.4 times less effective in identifying the worst 5% emitters compared to the reported emissions. It follows that the estimates from the data providers are at least 2.4 times less effective in identifying the worst emitters compared to the reported data.

Estimate consistency. We observe the rank correlations of 0.85 for the estimates from DP_A and DP_B and compare these estimates with those we obtain from Models (4) and (5). In **Table 6**, we display additional rank correlations.

Table 6. Cross-Rank Correlation of Estimated Data vs. Model Predictions, 2010–2016

Rank correlations	DP_A	DP_B	Model (4)	Model (5)
DP_A	1.00			
DP_B	0.85	1.00		
Model (4)	0.78	0.82	1.00	
Model (5)	0.82	0.86	0.95	1.00

Source: Research Affiliates, LLC, and University of Augsburg based on anonymized data from GHG emissions data providers.
Note: The table reports rank correlations between data providers' estimates on carbon emissions and the predicted values from the regression models we use in our analysis. We also analyze log-level correlations and find similar results, which are available upon request.

The rank correlation of 0.85 between the DP_A and DP_B estimates is a notch higher than the Model (4) correlations, 0.78 and 0.82, with data from DP_A and DP_B, respectively. The Model (5) correlation with DP_B estimated data is 0.86, a notch higher than 0.85; the correlation with DP_A estimated data is 0.82, a notch lower than 0.85. Overall, the correlations between the estimates from data providers and Model (5) are similar when compared to our model predictions. The fact that we observe similar correlations between the DP_A and DP_B estimates versus their correlations with Model (5), combined with the fact that Model (5) scores quite high in terms of accuracy, suggest that the Model (5) estimates could serve as the simple and transparent substitute for estimates from data providers for academic purposes.

Information content of the estimates. In addition to the need for accuracy, emissions data should capture information beyond simple correlates such as industry or size. Failure to capture this information stops investors from identifying the green companies in brown sectors and leads to counterproductive investor actions. By construction, estimates from Models (4) and (5) almost exclusively rely on information about company size and industry as well as a few other simple correlates; for Model (4) company size and industry are the only information used in the estimates. This implies the estimates from these models will not help investors identify the green companies in brown sectors. Will the estimates from DP_A and DP_B be more successful in this regard?

To analyze the information content of the estimates from DP_A and DP_B, we use the same regression, equation (1), as in Models (4) and (5), but in equation (2) we use estimated emissions instead of reported emissions,

$$\text{Log}(\text{Carbon emissions}_{i,t}^{\text{Estimated}}) = \alpha + \sum_{j}^{n} \beta_j \,\text{Log}(\text{Business metric}_{i,t}) + \sum_{k}^{m} \beta_k \,\text{Fixed effects}_{i,t} + \varepsilon_i \quad (2)$$

In **Table 7,** we provide the results of the regressions.

Table 7. Information Content in Estimated Data, 2010–2016

	Model (2)		Model (4)		Model (5)	
	DP_A	DP_B	DP_A	DP_B	DP_A	DP_B
	Log(Emissions$_t$)	Log(Emissions$_t$)	Log(Emissions$_t$)	Log(Emissions$_t$)	Log(Emissions$_t$)	Log(Emissions)
Log(Net sales$_t$)			0.95**	0.95**	0.98**	0.97**
Log(Employees/NS$_t$)					0.33**	0.24**
Log(Market Cap/NS$_t$)					-0.01	-0.03**
Log(EBT/NS$_t$)					0.01	-0.00
Log(FFO/NS$_t$)					0.03	0.05**
Log(OI/NS$_t$)					-0.09**	-0.03**
Log(PP&E/NS$_t$)					0.08**	0.21**
Log(COGS/NS$_t$)					-0.04*	0.06**
Constant	15.10**	13.73**	-5.85**	-5.86**	-2.78**	-3.54**
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes	Yes
Country fixed effects	No	No	Yes	Yes	Yes	Yes
Observations	8,665	19,365	8,665	19,365	8,665	19,365
Adjusted R^2	64.9%	55.6%	87.7%	87.6%	89.2%	89.6%

Source: Research Affiliates, LLC, and University of Augsburg based on anonymized data from GHG emissions data providers. Note: This table displays panel regression results of various company characteristics on carbon emissions (scope 1and 2). All variables are logarithmized due to their skewness. *Emissions* reflect a company's scope 1and 2 emissions. Column DP_A reports the estimates from the DP_A data set, whereas column DP_B reports the estimates from the DP_B data set. *Models* refer to Table 5. The independent variables are described in Appendix J. Industry fixed effects are based on a company's TRBC activity code. Country fixed effects are based on a country's ISO 3166 code. ∗∗ and ∗ indicate significance at the 1% and 5% level, respectively. Standard errors are clustered at the company level.

Company size, as well as industry, year, and country fixed effects, capture most of the variations in estimates. The specification of the model that captures only industry fixed effects, Model (2), explains 64.9% (55.6%) of DP_A's (DP_B's) estimated data variation. As reported in Table 4, industry fixed effects explain 54.2% of the reported data variation, the adjusted R^2 of Model (2). Because the adjusted R^2 is significantly higher for the DP_A estimates, DP_A seems to rely more heavily on simple industry affiliations in estimating carbon emissions. The specification of Model (4), which captures the company size and fixed effects (largely capturing the between-industry variation), draws a similar picture. The fraction of explained variation is higher using estimated emissions as the dependent variable as opposed to using reported emissions.

When we add the Model (5) extra variables, which mostly capture the employee and equipment intensity of production, the adjusted R^2 rises to 89–90% for both data providers. This estimate is higher than the Model (4) estimate of 87–88%. The increase in the R^2 is good news, indicating that these data sets do capture some, albeit very little, information beyond the general correlates such as industry or size.

More generally, the simple correlates, such as industry and size effects, explain the bulk of the variation in the estimated data. These models' accuracy is similar to Model (4) and lower compared to Model (5). Given that both Models (4) and (5) almost exclusively rely on and therefore reflect industry and size information, the implication is that the DP_A and DP_B estimates are equally unlikely to help investors differentiate the green companies in brown sectors, further reducing the efficacy of investor climate-mitigation activity. Overall, our results suggest that estimated emissions from data providers do explain the majority of the variation in the emissions, yet also suggest significant inferiority of the estimated data compared to the reported data.

Conclusion

Investors play a major role in combating climate change. To conduct effective investor actions and to have an impact on the real economy, they need high-quality GHG data. In our study, we investigate the quality of the currently available GHG data from an investor's perspective.

We argue that the following five criteria are necessary to help investors successfully mitigate climate change: 1) GHG data need to be widely available, 2) GHG data should be comparable between companies, 3) GHG data should be consistent across data providers, 4) forward-looking GHG information should have predictive power, and 5) GHG data should accurately reflect true company emissions.

We show that about half of the current emissions data are reported directly by companies. The concerns about the quality of these data include: 1) reporting is voluntary, which lowers data availability and introduces a potential self-reporting bias; 2) no single reporting standard has been adopted, which leads to incomparability of GHG emissions between companies; and 3) reported data are not perfectly consistent across data providers. Despite these drawbacks, the reported data are the best quality information currently available.

In addition to historical and current emissions, we also analyze data provider–specific carbon ratings and scores, which claim to capture forward-looking information. To be valuable to investors, these carbon ratings and scores should be able to explain future changes in emissions, but we find they have no predictive power.

Finally, we analyze the accuracy of the estimated GHG data. Overall, the estimated data do capture most of the emissions variation. At the same time, the conservative estimates suggest that investor actions are at least 2.4 times more diluted when investors use estimated emissions compared to reported emissions. Further, we show that estimated emissions data are based mainly on industry and size information. Thus, using estimated data may not accurately identify the green projects in brown sectors and lead to the opposite of the desired outcome for green investors. Our study uncovers the misconception that estimated emissions can be equally as useful to investors as reported emissions. We do not interpret our results as suggesting that data providers are sloppy estimators of various data sets. Instead, likely due to information asymmetry, these estimates are the best estimates data providers can make.

Our findings suggest that the status quo, in which GHG reporting is voluntary and data providers estimate the missing data, is inadequate for investors. This inadequacy leads to a significant reduction in the impact of investors' actions as they attempt to mitigate climate change.

The potential for greenwashing is high. Our analysis suggests the solution may be the introduction of an international regulatory initiative with mandatory reporting of GHG data. This should be accompanied by the establishment of a single standard, such as the GHG Protocol, and the introduction of mandatory auditing to ensure data quality. In the interim, while reporting remains voluntary, data providers should strive to increase the percentage of the reported data while investors should push companies to report their GHG emissions. One way for investors to incentivize companies to start reporting is to assume the worst possible outcome given the information available about the nonreporting company, as proposed in the United Nations (1992) precautionary principle.

References

Akey, Pat, and Ian Appel. 2019. "Environmental Externalities of Activism." Working paper (November 18). Available at SSRN.

Berg, Florian, Julian Koelbel, and Roberto Rigobon. 2019. "Aggregate Confusion: The Divergence of ESG Ratings." MIT Sloan Working Paper 5822-19. MIT Sloan School of Management (August).

Boermans, Martijn A., and Rients Galema. 2019. "Are Pension Funds Actively Decarbonizing Their Portfolios?" *Ecological Economics*, 161 (July): 50–60.

Bolton, Patrick, and Marcin T. Kacperczyk. 2020. "Carbon Premium Around the World." Working Paper (March 13). Available on SSRN.

Busch, Timo, Matthew Johnson, Thomas Pioch, and Matthias Kopp. 2018. "Consistency of Corporate Carbon Emission Data." Hamburg: University of Hamburg / WWF Deutschland. https://ec.europa.eu/jrc/sites/jrcsh/files/paper_timo_busch.pdf.

Busch, Timo, and Stefan Lewandowski. 2018. "Corporate Carbon and Financial Performance: A Meta-Analysis." *Journal of Industrial Ecology*, 22 (4): 745–759.

CDP. 2017. "The Carbon Majors Database: CDP Carbon Majors Report 2017."

Climate Action 100+. 2019. 2019 Progress Report. http://www.climateaction100.org/.

Cotter, Julie, and Muftah M. Najah. 2012. "Institutional Investor Influence on Global Climate Change Disclosure Practices." *Australian Journal of Management*, 37 (2): 169–187.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2020. "Divergent ESG Ratings." *Journal of Portfolio Management*, 47 (9).

Dordi, Truzaar, and Olaf Weber. 2019. "The Impact of Divestment Announcements on the Share Price of Fossil Fuel Stocks." *Sustainability*, 11 (11): 1–20.

Edmans, Alex, Xavier Gabaix, and Dirk Jenter. 2017. "Chapter 7—Executive Compensation: A Survey of Theory and Evidence." In *The Handbook of the Economics of Corporate Governance*, 1: 383–539.

Flammer, Caroline, Michael W. Toffel, and Kala Viswanathan. 2019. "Shareholder Activism and Firms' Voluntary Disclosure of Climate Change Risks." Working paper (October 12). Available on SSRN.

Heinkel, Robert, Alan Kraus, and Josef Zechner. 2001. "The Effect of Green Investment on Corporate Behavior." *Journal of Financial and Quantitative Analysis*, 36 (4): 431–449.

IPCC. 2018. "Summary for Policymakers." In *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*.

Lee, Su-Yol, Yun-Seon Park, and Robert D. Klassen. 2015. "Market Responses to Firms' Voluntary Climate Change Information Disclosure and Carbon Communication." *Corporate Social Responsibility and Environmental Management*, 22 (1): 1–12.

Li, Feifei, Polychronopoulos, Ari. 2020. "What a Difference an ESG Ratings Provider Makes!" Working paper (January 2020). Available at researchaffiliates.com

Liesen, Andrea, Frank Figge, Andreas Hoepner, and Dennis M. Patten. 2015. "Climate Change and Asset Prices: Are Corporate Carbon Disclosure and Performance Priced Appropriately?" *Journal of Business Finance & Accounting*, 44 (1-2): 35–62.

Matsumura, Ella Mae, Rachna Prakash, and Sandra C. Vera-Muñoz. 2014. "Firm-Value Effects of Carbon Emissions and Carbon Disclosures." *The Accounting Review*, 89 (2): 695–724.

———. 2017. "To Disclose or Not to Disclose Climate-Change Risk in Form 10-K: Does Materiality Lie in the Eyes of the Beholder?" Working paper (June 8). Available on SSRN.

Naaraayanan, S. Lakshmi, Kunal Sachdeva, and Varun Sharma. 2020. "The Real Effects of Environmental Activist Investing." Working paper (September 18). Available at SSRN.

Natixis. 2016. "Challenges and Tools for Incorporating Climate Themes Into Divestment Strategies." Working paper (September 7). Available at southpole.com.

Prado-Lorenzo, Jose-Manuel, Isabel Gallego-Alvarex, and Isabel M. Garcia-Sanchez. 2009. "Stakeholder Engagement and Corporate Social Responsibility Reporting: The Ownership Structure Effect." *Corporate Social Responsibility and Environmental Management*, 16 (2): 94–107.

PRI. 2020. "Principles for Responsible Investment." https://www.unpri.org/download?ac=10948.

Rohleder, Martin, Marco Wilkens, and Jonas Zink. 2020. "The Effects of Decarbonizing Institutional Portfolios on Stock Prices and Carbon Emissions." Working paper (June 23). Available at SSRN.

TCFD. 2019. "Task Force on Climate-Related Financial Disclosures: Status Report." Available at https://www.fsb.org/wp-content/uploads/P050619.pdf.

Trudeau, Justin. 2020. "Prime Minister Announces Additional Support for Businesses to Help Save Canadian Jobs." Available at https://pm.gc.ca/en/news/news-releases/2020/05/11/prime-minister-announces-additional-support-businesses-help-save.

United Nations. 1992. "Report of the United Nations Conference on Environment and Development." A/CONF.151/26, vol. 1. Available at https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_CONF.151_26_Vol.I_Declaration.pdf.

———. 2015. "Paris Agreement." FCCC/CP/2015/L.9/ Rev.1 Available at https://unfccc.int/files/meetings/paris_nov_2015/application/pdf/paris_agreement_english_.pdf.

WBCSD and WRI. 2011. "Corporate Value Chain (Scope 3) Accounting and Reporting Standard." Available at https://ghgprotocol.org/sites/default/files/standards/Corporate-Value-Chain-Accounting-Reporing-Standard_041613_2.pdf.

———. 2015a. “GHG Protocol Scope 2 Guidance.” Available at https://ghgprotocol.org/sites/default/files/standards/Scope%202%20Guidance_Final_Sept26.pdf.

———. 2015b. “The Greenhouse Gas Protocol - A corporate accounting and reporting standard.” Available at https://ghgprotocol.org/sites/default/files/standards/ghg-protocol-revised.pdf.

Appendix A: Description of Carbon Scores by Data Provider

Data provider	Name of score	Description	Range	Scored companies in 2016
DP_A	Emission reduction score	Measures a company's commitment and effectiveness in reducing environmental emissions in the production and operational processes.	0–100	5,968
DP_C	Climate change score	Measures the comprehensiveness of disclosure, awareness, and management of environmental risks and best practices associated with environmental leadership, such as setting ambitions and meaningful targets.	From A to D- (eight distinct categories)	2,112
DP_D	Carbon emission score	Evaluates the extent to which companies may face increased costs linked to carbon pricing or regulatory caps. Scores are based on exposure to GHG-intensive businesses and emerging regulations; carbon reduction targets and mitigation programs; and carbon intensity over time and versus peers.	0–10	14,798

Appendix B: The GHG Protocol[1]

The Greenhouse Gas (GHG) Protocol Initiative is a multi-stakeholder partnership of businesses, nongovernmental organizations (NGOs), governments, and others convened by the World Resources Institute (WRI), a US-based environmental NGO, and the World Business Council for Sustainable Development (WBCSD), a Geneva-based coalition of 170 international companies. Launched in 1998, the Initiative's mission is to develop internationally accepted greenhouse gas accounting and reporting standards for business and to promote the standards broad adoption.

This GHG Protocol Corporate Standard provides standards and guidance for companies and other types of organizations preparing a GHG emissions inventory. It addresses the accounting and reporting of the six greenhouse gases included in the Kyoto Protocol—carbon dioxide (CO2), methane (CH4), nitrous oxide (N2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), and sulphur hexafluoride (SF6).

The standard and guidance are offered with the following objectives in mind:

1) To help companies prepare a GHG inventory that represents a true and fair accounting of their emissions through the use of standardized approaches and principles.
2) To simplify and reduce the costs of compiling a GHG inventory.
3) To provide businesses with information that can be used to build an effective strategy to manage and reduce GHG emissions.
4) To provide information that facilitates participation in voluntary and mandatory GHG programs.
5) To increase consistency and transparency in GHG accounting and reporting among various companies and GHG programs.

[1] Information was obtained from WBCSD and WRI (2015b).

Scope of Emissions

One of the most important aspects of the GHG Protocol is to classify corporations' emissions into three areas, or scopes, as illustrated in the figure below:

Scope 1: Direct GHG emissions

Companies report GHG emissions from sources they own or control:

- Generation of electricity, heat, or steam
- Physical or chemical processing
- Transportation of materials
- Fugitive emissions

Scope 2: Electricity indirect GHG emissions

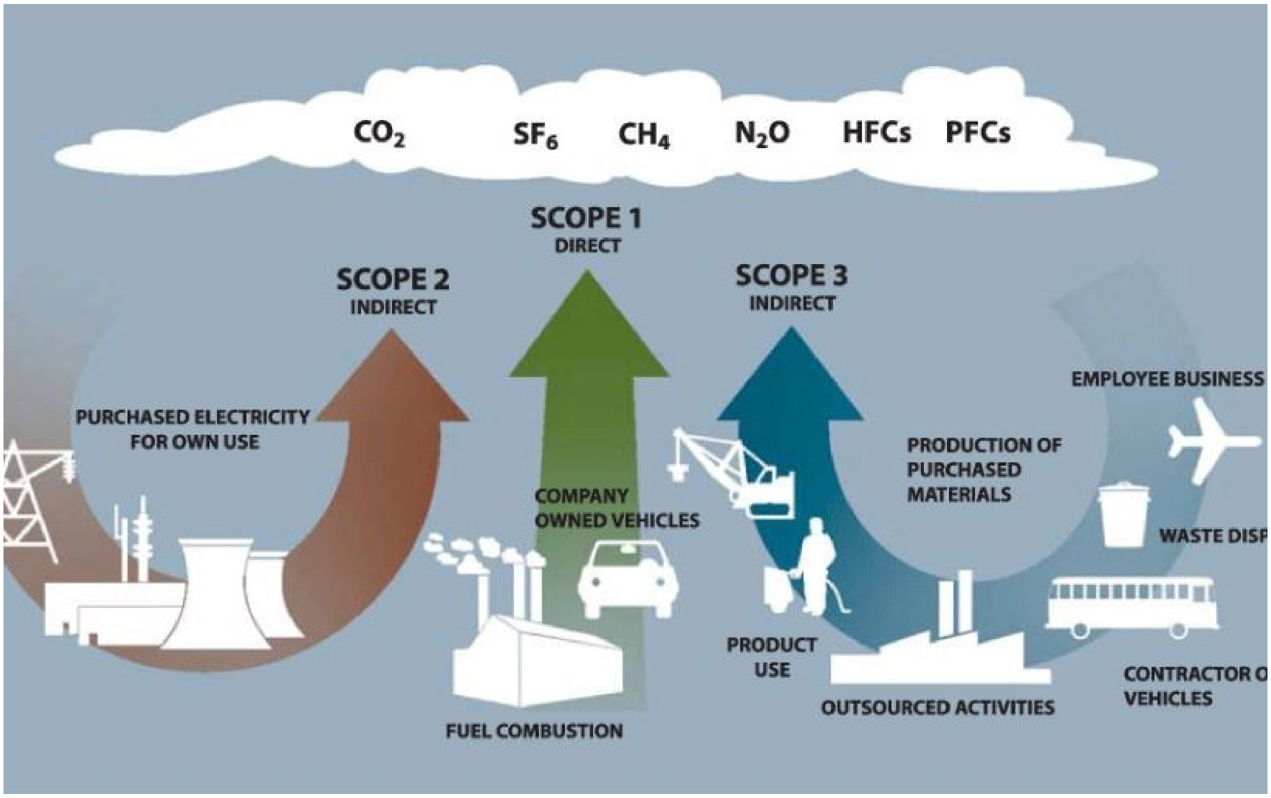
Companies report the emissions from the generation of purchased electricity consumed in its owned or controlled equipment or operations. For many companies, purchased electricity represents one of the largest sources of GHG emissions and the most significant opportunity to reduce these emissions.

Scope 3 (Optional): Other indirect emissions

This scope provides an opportunity to be innovative in GHG management and includes the following emissions:

- Extraction and production of purchased materials and fuels
- Transport-related activities
- Electricity-related activities not included in scope 2
- Leased assets, franchises, and outsourced activities
- Use of sold products and services
- Waste disposal

Overview of Scopes and Emissions across a Value Chain



Source: Available at http://synergyfiles.com/2017/04/scope-ghg-emissions/. This slide may contain copyrighted material. Such material is being made available for research and educational purposes only and should be considered a fair use.

Appendix C: Summary Statistics of Carbon Data and Company Characteristics by Data Set, 2010–2016

	DP$_A$			DP$_B$			DP$_C$			DP$_D$		
	# obs.	Mean	p50	# obs.	Mean	p50	# obs.	Mean	p50	# obs.	Mean	p50
Panel A: Carbon data												
Reported scope 1 emissions	12,629	3,274,450	71,410	11,135	3,034,741	59,606	12,194	3,074,414	58,576	--	--	--
Reported scope 2 emissions	12,477	651,793	111,240	11,119	602,332	107,140	12,099	614,282	105,616	--	--	--
Reported scope 1+2 emissions	15,708	3,962,170	254,989	11,119	3,633,120	240,055	11,986	3,449,618	231,695	--	--	--
Estimated scope 1 emissions	--	--	--	49,991	889,359	12,748	--	--	--			
Estimated scope 2 emissions	--	--	--	49,991	130,546	18,146	--	--	--			
Estimated scope 1+2 emissions	17,101	1,452,074	48,181	49,991	1,019,905	38,852	--	--	--	--	--	--
Carbon intensity	32,142	12.35	0.41	10,256	182.13	0.43	10,864	229.89	0.43	--	--	--
DP$_A$ emission reduction score	31,588	50.6	50.8	--	--	--	--	--	--	--	--	--
DP$_D$ carbon emission score	--	--	--	--	--	--	--	--	--	52,074	7.3	7.6
Panel B: Company characteristics												
Net sales (in US$ million)	32,316	9,100	2,590	54,469	5,490	1,010	10,877	15,500	5,340	19,844	11,100	3,170
BTM	32,168	0.71	0.55	51,243	0.56	0.56	10,761	0.70	0.56	19,702	0.56	0.52
Employees	28,035	26,735	8,100	44,800	17,263	3,872	10,063	39,931	14,529	17,944	29,658	8,700
Total assets / net sales	32,134	125.79	1.74	53,471	14.63	1.62	10,821	16.31	1.44	19,672	16.53	1.63
R&D expenses / net sales	12,750	230.9%	1.8%	21,593	328.4%	2.1%	5,391	4.1%	2.0%	8,317	167.4%	2.0%
PP&E / net sales	31,756	4.02	0.28	52,761	3.33	0.26	10,754	0.92	0.26	19,474	2.36	0.29
EBT / net sales	32,124	-2.14	0.10	53,609	-2.10	0.09	10,820	0.04	0.09	19,671	-1.28	0.10
FFO / net sales	32,128	-1.86	0.15	53,588	-1.39	0.13	10,820	0.43	0.13	19,664	-0.48	0.15
OI / net sales	32,102	-2.47	0.12	53,549	-1.90	0.10	10,810	0.05	0.10	19,653	-1.02	0.12
COGS / net sales	28,217	0.22	0.60	47,262	0.77	0.61	9,572	0.60	0.62	17,584	0.64	0.60

Source: Research Affiliates, LLC, and University of Augsburg based on anonymized data from GHG emissions data providers.

Note: This table provides summary statistics of variables used in this study for the period 2010–2016. Panel A shows yearly carbon-related data and Panel B lists yearly company characteristics. Carbon data are explained in the text in detail. *Carbon intensity* is displayed in (tons/$) scaled by 10,000. Company characteristics are obtained from Refinitiv Datastream. *# obs.* reflect the number of company-year observations. *BTM* reflects the book-to-market ratio. *R&D expense* represents all direct and indirect costs related to the creation and development of new processes, techniques, applications, and products with commercial possibilities. *PP&E* represents gross property, plant, and equipment less accumulated reserves for depreciation, depletion, and amortization. *EBT* reflects the pre-tax margin. *FFO* represents the sum of net income and all noncash charges or credits. *OI* represents the difference between sales and total operating expenses. *COGS* reflects the cost of goods sold excluding depreciation.

Appendix D: Market-Capitalization Coverage by Year, 2010−2016

	DP_A				DP_B				DP_C			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Year	No. of covered companies	Reported (%-age of market cap)	Estimated (%-age of market cap)	Not covered (%-age of market cap)	No. of covered companies	Reported (%-age of market cap)	Estimated (%-age of market cap)	Not covered (%-age of market cap)	No. of covered companies	Reported (%-age of market cap)	Estimated (%-age of market cap)	Not covered (%-age of market cap)
2010	4,012	59%	26%	15%	7,959	38%	43%	19%	1,479	45%	0%	55%
2011	4,155	65%	27%	8%	8,246	43%	45%	13%	1,584	50%	0%	50%
2012	4,249	63%	25%	12%	8,460	41%	44%	15%	1,754	50%	0%	50%
2013	4,356	61%	25%	14%	8,576	51%	34%	15%	1,758	48%	0%	52%
2014	4,462	59%	27%	14%	8,614	50%	35%	15%	1,795	48%	0%	52%
2015	5,322	61%	26%	13%	8,564	51%	34%	15%	1,812	47%	0%	53%
2016	6,216	63%	24%	13%	10,691	51%	36%	13%	1,804	48%	0%	52%
Mean	4,682	62%	25%	13%	8,730	47%	38%	15%	1,712	48%	0%	52%

Source: Research Affiliates, LLC, and University of Augsburg based on anonymized data from GHG emissions data providers.

Note: This table compares the coverage of companies with GHG data across data providers and over time. Column (1) reflects the number of companies covered with GHG data. Column (2) reflects the fraction of total world market capitalization covered with reported GHG data. Column (3) reflects the fraction of total world market capitalization covered with data providers' estimates on carbon emissions. Column (4) reflects the fraction of world total market capitalization not covered with any GHG data. Data on world market capitalization of listed companies is from the World Bank (https://data.worldbank.org/indicator/CM.MKT.LCAP.CD).

Appendix E: Relative Emission Deviation

The rank correlations are important but may mask big deviations in individual observations. We dig deeper into the percentage deviation of emissions and compare pairwise relative emission deviations. For each company, we compute relative emission deviations in emission levels of observations for which we have GHG data from multiple data providers and in **Table E1** display the percentage of companies that fall into various deviation ranges. Equation (E1) displays the calculation formula of relative emission deviation (*K* and *L* reflect the respective carbon data provider),

$$\text{Relative emission deviation}_{i,t,K,L} = \left| \frac{\text{Carbon emissions}_{i,t,K} - \text{Carbon emissions}_{i,t,L}}{\text{Carbon emissions}_{i,t,L}} \right| \tag{E1}$$

Table E1

			Relative deviation (in %)					
		Joint observations	|0-1%|	|1-5%|	|5-10%|	|10-30%|	|30-70%|	|>70%|
Panel A: Reported scope 1 emissions								
DP_A	DP_C	7,804	60%	12%	7%	11%	5%	5%
DP_A	DP_B	7,566	61%	12%	7%	10%	5%	5%
DP_C	DP_B	9,052	88%	3%	3%	3%	1%	1%
Panel B: Reported scope 2 emissions								
DP_A	DP_C	7,796	57%	12%	9%	12%	7%	4%
DP_A	DP_B	7,543	60%	11%	8%	10%	6%	5%
DP_C	DP_B	9,111	90%	2%	2%	2%	2%	1%
Panel C: Reported scope 1+2 emissions								
DP_A	DP_C	8,637	50%	16%	10%	13%	7%	4%
DP_A	DP_B	8,353	54%	15%	9%	12%	6%	4%
DP_C	DP_B	9,114	86%	5%	3%	3%	2%	1%
Panel D: Estimated scope 1+2 emissions								
DP_A	DP_B	12,452	1%	4%	5%	18%	25%	47%

Source: Research Affiliates, LLC, and University of Augsburg based on anonymized data from GHG emissions data providers.
Note: This table reports relative deviations in percentage terms for each emission category. Relative deviations reflect the percentage difference between the emissions from two different data sets for the same company in the same year. The percentage values within the table reflect the fraction of joint observations in the respective relative emission deviation class. Numbers may not add up to 100% due to rounding.

The relative deviation between DP_C's and DP_B's reported emissions is relatively low. In the reported scope 1 and 2 levels, 86% of all observations deviate less than 1%, and 94% deviate less than 10%, however, there are a few outliers. About 3% of all observations deviate by more than 30%, and 1% deviate by more than 70%. The reported emissions between DP_A and DP_C and between DP_A and DP_B share more inconsistencies. For instance, in more than 24% of cases, scope 1 and 2 carbon emissions differ by more than 10% between DP_A and DP_C.

In the case of estimated emissions, we identify even larger differences that may be masked by relatively large correlation levels. For DP_A and DP_B, only 10% of the sample deviates by less than 10%. For the vast majority of the sample (72%) the difference is larger than 30%, and for almost half of the sample (47%) the estimates from the two data providers disagree by more than 70%. These differences are large and caused by the different carbon emission estimation models.

Appendix F: Persistence of Carbon Emissions/Intensities Trends

Stability of Carbon Emission Trends

t/t+1	Strong reducer$_{t+1}$	Reducer$_{t+1}$	No changer$_{t+1}$	Increaser$_{t+1}$	Strong increaser$_{t+1}$
Strong reducer$_t$	27.2%	23.9%	6.4%	17.6%	25.0%
Reducer$_t$	17.0%	36.0%	9.3%	22.9%	14.8%
No changer$_t$	14.5%	33.6%	12.3%	24.4%	15.2%
Increaser$_t$	13.8%	29.9%	11.0%	28.1%	17.2%
Strong increaser$_t$	20.6%	21.6%	7.2%	23.3%	27.4%

Source: Research Affiliates and University of Augsburg, based on anonymized data from GHG emissions data providers.

Note: The migration matrix shows from which category companies migrated from one year to the next year. We distinguish between strong reducers (emissions reduction of more than 10%), reducers (emissions reduction between 1% and 10%), no changer (overall emissions change smaller than 1%), increasers (emissions rise between 1% and 10%), and strong increasers (emissions rise more than 10%). Changes in emissions reflect the percentage deviation in companies' scope 1 and 2 emissions for the period 2010–2016.

Stability of Carbon Intensity Trends

t/t+1	Strong reducer$_{t+1}$	Reducer$_{t+1}$	No changer$_{t+1}$	Increaser$_{t+1}$	Strong increaser$_{t+1}$
Strong reducer$_t$	24.3%	20.8%	4.9%	16.0%	34.0%
Reducer$_t$	21.3%	28.1%	5.5%	19.0%	26.0%
No changer$_t$	17.9%	26.6%	5.4%	24.1%	26.1%
Increaser$_t$	19.8%	23.7%	5.5%	21.8%	29.2%
Strong increaser$_t$	24.6%	16.2%	5.0%	18.5%	35.7%

Source: Research Affiliates, LLC, and University of Augsburg based on anonymized data from GHG emissions data providers.

Note: The migration matrix shows from which category companies migrated from one year to the next year. We distinguish between strong reducers (emissions reduction of more than 10%), reducers (emissions reduction between 1% and 10%), no changer (overall emissions change smaller than 1%), increasers (emissions rise between 1% and 10%), and strong increasers (emissions rise more than 10%). Changes in emissions reflect the percentage deviation in companies' carbon intensities. Carbon intensities reflect the ratio of carbon emissions to net sales.

Appendix G: Predictability of Future Reported Carbon Intensity Changes

	Δ% Carbon intensity$_{t+1,t}$	Δ% Carbon intensity$_{t+1,t}$	Δ% Carbon intensity$_{t+1,t}$	Δ% Carbon intensity$_{t+1,t}$	Δ% Carbon intensity$_{t+1,t}$	Δ% Carbon intensity$_{t+1,t}$
Δ% Carbon intensity$_{t1,t-1}$		-0.04*	-0.04*	-0.04*	-0.04*	-0.04*
DP_C A list$_t$			-0.08			-0.00
DP_A emission reduction score$_t$				-0.00		-0.00
DP_D carbon emission score$_t$					-0.08	-0.07
Constant	0.66	0.68	0.68	0.94	0.87	0.96
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	5,895	5,895	5,895	5,895	5,895	5,895
R^2	3.5%	3.6%	3.6%	3.6%	3.7%	3.7%
Adjusted R^2	-3.2%	-3.1%	-3.1%	-3.1%	-3.0%	-3.0%
Rank correlation between predicted changes in future emissions and reported changes in future emissions	10.9%	11.4%	11.3%	10.9%	10.2%	10.1%

Source: Research Affiliates, LLC, and University of Augsburg based on anonymized data from GHG emissions data providers. Note: This table reports regression results of past reported carbon emissions changes on future reported carbon emissions changes (scope 1 and 2). *Δ% Carbon intensity$_{t+1,t}$* reflects the percentage change of a company's carbon intensity (scope 1 and 2 emissions divided by net sales) from one year to another. Industry fixed effects are based on a company's TRBC activity code. Country fixed effects are based on a country's ISO 3166 code. ** and * indicate significance at the 1% and 5% level, respectively. Standard errors are clustered at the company level.

Appendix H: Carbon Estimation Models by Data Providers

DP_A: Only half of all companies in the DP_A ESG coverage report on carbon emissions. Therefore, DP_A developed a patented model that estimates emissions when a reported value is not available. DP_A jointly estimates scope 1 and scope 2 issues and therefore does not provide estimates for single emission categories. DP_A uses a three-stage approach to estimate carbon emissions:

1) **CO_2 model:** If the company has reported carbon emissions in the past, but does not report in the current year, historical emissions are estimated based on the current number of employees and net sales (*CO_2 model)*.

   Variables used: Historical carbon emissions (enerdp023), total revenue (wc01001), and number of employees (wc07011).

2) **Energy model**: If the company has not reported carbon emissions in the past, emissions are estimated based on the amount of energy consumed using peer group information (*energy model*).[35]

   Variables used: total energy use (enrrdp033), direct energy produced (enrrdp0342), total revenue (wc01001), number of employees (wc07011), and industry classification.

3) **Median model:** If no information about the amount of consumed energy is available, DP_A calculates median carbon intensities using number of employees and net sales (*median model*).

   Variables used: Total revenue (wc01001), number of employees (wc07011), and industry classification.

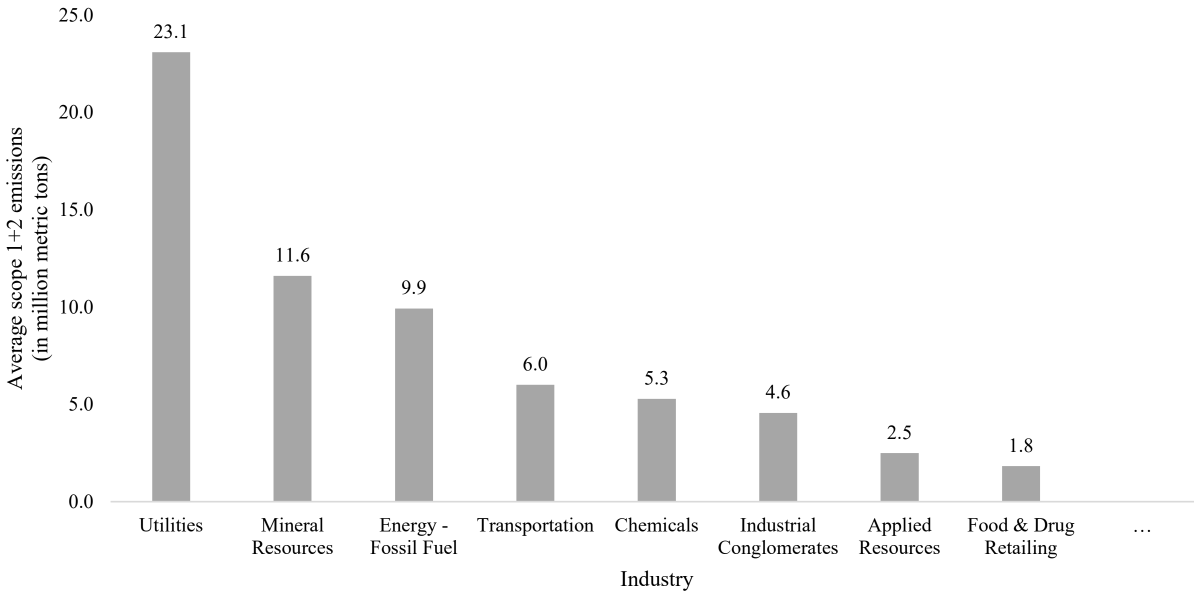
[35] For utilities, we use the amount of energy produced.

DP_B: In contrast to DP_A, DP_B separately estimates scope 1 and scope 2 emissions for companies that do not report. Through regression analysis, DP_B generates estimates for nonreporting companies using company-reported carbon emissions to calibrate the model. Similar to DP_A, DP_B estimates carbon emissions in relation to companies' net sales, number of employees, and gross PP&E. The final estimate is derived by averaging the predicted outcomes.

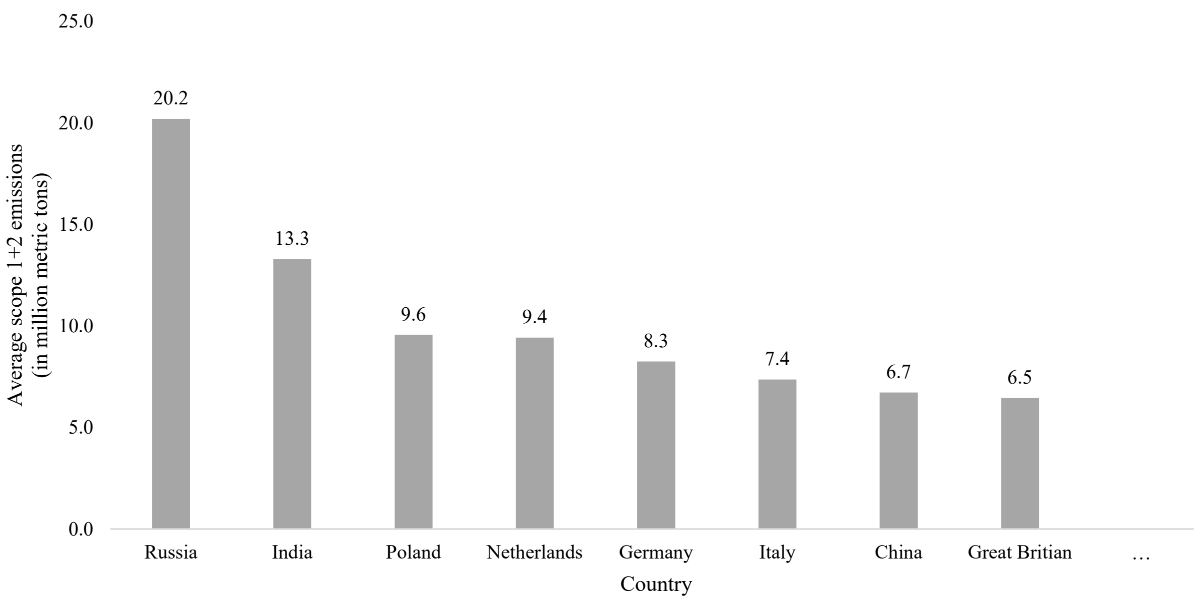
Variables used: Total revenue, gross PP&E, and employee count.

Appendix I: Breakdown of Carbon Emissions by Industry and Country

Carbon Emissions by Industry



Carbon Emissions by Country



Source: Research Affiliates, LLC, and University of Augsburg based on anonymized data from GHG emissions data providers. Note: These figures show cross-sectional averages of scope 1 and 2 emissions by industry (Panel A) and by country (Panel B) over the period 2010–2016. Industries are classified using Thomson Reuters Sector Code. Countries are classified according to their ISO 3166 standard. For both panels, a minimum of 10 observations is required to calculate the average.

Appendix J: Overview of Business Metrics Used in Regression Analyses

No.	Variable	Variable Code	Description (of nominator)	Captures	Rationale
1	Log (Net Sales)	WC01001	Net Sales represents gross sales and other operating revenue less discounts, returns, and allowances.	Company size and production levels	Larger companies are more likely to be carbon intensive.
2	Log (Employees / Net Sales)	WC07011; WC01001	Employees represent the number of both full and part time employees of the company.	Employee intensity of production	Companies with higher employee intensity are likely to be carbon intensive (e.g., need more office space).
3	Log (Market Capitalization / Net Sales)	WC08001; WC01001	Market Price-Year End * Common Shares Outstanding	Market valuation of production	Companies with lower market valuations are more likely to be carbon intensive because emissions are negatively associated with market values on average (Matsumura, Prakash, and Vera-Munoz (2017).
4	Log (Earnings before taxes / Net Sales)	WC08321; WC01001	Pre-tax Income / Net Sales or Revenues * 100	Profitability	Companies with higher profitability have more capacity to invest in carbon-efficient technologies.
5	Log (Funds from operations / Net Sales)	WC04201; WC01001	Funds from operations represents the sum of net income and all noncash charges or credits. It is the cash flow of the company.	Liquidity	Companies with higher liquidity have more cash available to invest in carbon-efficient technologies.
6	Log (Operating income / Net Sales)	WC01250; WC01001	Operating income represents the difference between sales and total operating expenses.	Profitability	Companies with higher profitability have more capacity to invest in carbon-efficient technologies.
7	Log (Property, plant, and equipment / Net Sales)	WC02501; WC01001	Net property, plant, and equipment represents gross property, plant, and equipment less accumulated reserves for depreciation, depletion, and amortization.	Tangible assets	Companies with a higher fraction of tangible assets are more likely to be carbon intensive (e.g., energy to heat buildings, energy for production machines). In contrast, intangibles are less likely to be carbon intensive (e.g., patents or brand values).
8	Log (Cost of goods sold / Net Sales)	WC01051; WC01001	For manufacturing companies, cost of goods sold represents specific or direct manufacturing cost of material and labor used in the production of finished goods. For merchandise companies, cost of goods sold represents the purchase price of items sold, as well as indirect overhead such as freight, inspection, and warehouse costs.	Production levels	Companies with higher production levels are more likely to be carbon intensive.