Modeling facility-level emissions data in corporate finance: A cautionary tale

Danilo V. Mascia* Enrico Onali†

Abstract

We uncover several empirical regularities in greenhouse gas (GHG) emissions at the facility level, in the United States, and study their potential impact on statistical inferences using Monte Carlo simulations. Facility-level emissions tend to be highly persistent and emissions tend to cluster more at the regional level than by firm. Scaling emissions by a proxy for firm size leads to highly leptokurtic distributions, and can affect the relation between emissions and Tobin's Q. Our Monte Carlo simulations highlight the impact of neglecting these empirical regularities on the explanatory power of the regressions and the estimated coefficients and t-statistics. Depending on the true underlying relation, scaling firm value and emissions can lead to large Type II errors and statistically significant coefficients with the wrong sign. Scaling the intercept does not solve the problem. Failing to account for these features can lead to incorrect inferences regarding the impact of climate policies: for example, the estimated impact of the California Cap-and-Trade Program on emissions depends on whether an autoregressive component is included in the regressions. We discuss alternative solutions under different scenarios.

JEL classification : C52; G1; Q51; M41

Keywords : Firm value; CO_2 emissions; GHG emissions; Climate policy

Acknowledgments. We would like to thank Samuel Engle, Alain Naef, Chrysovalantis Vasilakis, Pengguo Wang, Tiancheng Yu, and participants at the International Conference in Banking and Financial Studies (Catania, Italy, September 2024), and at the International Finance and Banking Society (Shanghai, China, December 2024) for very helpful comments.

*International Banking Institute, Leeds University Business School, University of Leeds, Maurice Keyworth Building, Leeds, LS2 9JT, United Kingdom. Email: D.V.Mascia@leeds.ac.uk.

†Corresponding author at University of Bristol Business School, United Kingdom. Email: E.Onali@bristol.ac.uk.

1 Introduction

It is widely acknowledged that climate change is a consequence of global warming induced by Greenhouse Gas (GHG) emissions (Stern 2008). Addressing GHG emissions has thus become a pivotal goal, especially since the past decade. Investors take environmental performance – as well as other non-financial and ethical issues – into consideration when making investment decisions, and therefore GHG emissions can affect firm value (Hartzmark & Sussman 2019, Garavaglia et al. 2023). In this paper, we address the following research questions: How should we model facility-level emissions data in corporate finance? How do modeling choices affect the relation between facility-level emissions and firm value?

We focus on data provided by the Greenhouse Gas Reporting Program (GHGRP), which is compiled by the United States Environmental Protection Agency (EPA). In addition to being publicly available, the GHGRP has two features that differentiate it from any other dataset on GHG emissions. First, due to a continuous verification process, the GHGRP is considered to be the most reliable public source of GHG emissions data (Kahn et al. 2023). Second, this data set provides facility-level data on emissions that can be matched with parent-level characteristics. This unique feature has enabled researchers to address research questions that cannot be answered using firm-level emissions data. For example, Mascia & Onali (2024) use the GHGRP to estimate the impact of climate regulations on emissions at the county-level, while other papers have focused on the role of firm-level characteristics in explaining heterogeneity in emissions, such as financial constraints (Bartram et al. 2022) or ownership structure (Shive & Forster 2020).

In this paper, we investigate potential methodological issues that might have led to inconsistencies in the findings provided by the literature. We exploit the most micro unit of analysis by focusing on facility-level emissions data.[1]

In fact, one of the main inconsistencies across studies using GHGRP data at the facility level is the inclusion of different types of fixed effects (for brevity, FE). Facility FE are ubiquitous in the literature on GHG emissions (among others, Mascia & Onali (2024), Tomar (2023), Bartram et al. (2022)), but it is unclear what this type of effects actually capture. Moreover, there is heterogeneity in the other types of FE used (e.g., firm, industry, county). As highlighted by Breuer & deHaan (2024), while including FE at a given level can reduce omitted variable bias and improve statistical test power, researchers should discuss what variation they want to capture and how including different types of FE can help isolate the variation they intend to zoom in to test a particular theory.

Thus, our analyses start by investigating the extent to which different types of FE can explain the variability in emissions data. Moreover, we attempt to understand whether facility-level emissions can be better modeled using econometric methods that account specifically for autoregressive components in emissions. In fact, facility-level emissions are likely to be directly linked to the capital invested in each facility, which is typically modeled as an AR(1) process due to a linear rate of depreciation (Warusawitharana 2008).

A second important source of inconsistencies is the proxy for environmental impact itself. To improve comparability across industries of such a measure, and to improve robustness in their results, some studies scale emissions by proxies for firm size (Shive & Forster 2020). However, scaling emissions by a proxy for firm size (e.g., sales), a variable commonly known as "emission intensity", leads to difficulties in the interpretation of the results: when emission intensity changes, it is unclear whether this change is driven

[1]This enables us to use the dataset in its entirety, thereby avoiding possible selection biases we would encounter if we were to match facility with company-level data in the initial step of our analyses. This is not inconsequential, given that our sample would shrink by about 70% if we were to immediately match accounting and/or stock market data related to facilities' owners.

by variations in the numerator (emissions) or the denominator (firm size proxy). More importantly, in this paper we try to understand the impact of scaling emissions from a statistical perspective.

Third, we analyse how different methodologies can affect the estimated impact of GHG emissions on Tobin's Q, one of the most widely used variables in corporate finance (Erickson & Whited 2012) as a proxy for firm value. In this analysis, we do not intend to provide evidence on the causal relationship between emissions and firm value. Our objective is to highlight how changes in model specification, such as scaling the proxy for GHG emissions for different firm-size proxies, might lead to different results.

Our main findings are as follows. We observe that facility FE alone explain 94% of the variability in emissions. This suggests that emissions are persistent, over time, at the facility level. Moreover, scaling emissions by a proxy for size results in substantial reductions in the explanatory power of all models, including those with facility FE. Regarding the association between emissions and firm value, we find that results can be negative and statistically significant or statistically insignificant, although we never observe a positive and significant correlation with Tobin's Q. These differences depend on whether we use emissions or emission intensity as a proxy for environmental impact.

Taking stock of these empirical results, similarly to Flannery & Hankins (2013) – who simulate data resembling 'real' corporate finance data – we use Monte Carlo simulations to explore these issues in greater depth. We generate 1,000 simulated samples of 10,000 fictitious facilities allocated to 1,000 firms and 1,000 counties for 10 years, leading to 100,000 observations for each simulated sample. Our Monte Carlo simulations focus on two levels of analysis: first, using facility-level data, we investigate how different types of Data Generating Processes (DGP) can affect the explanatory power of different types of FE in regressions where facility-level emissions are the dependent variable; second, we aggregate the facility-level data for emissions and capital stock at the firm level and examine how different DGPs affect inferences related to the impact of firm emissions (in levels) on firm value.

Since in our empirical analysis facility FE explain over 90% of the variation in emissions, we investigate whether facility FE capture actual time-invariant omitted components at the facility level, or if there is persistence in the form of an AR(1) component. Thus, we focus on two types of DGP: one where there is high persistence in facility-level emissions (with an autoregressive coefficient close to 0.9), but no facility-specific time-invariant component and negligible time-, firm- and county-specific component; and one where there is no persistence in facility-level emissions, no facility-specific time invariant component, but a large firm-specific component.

Our main findings for this part of the paper are as follows. High persistence alone can generate adjusted R-squared (*R2-adj*) values over 65% in regressions with facility and year FE, despite the absence of time-invariant components at the facility level. If we allocate randomly facilities to firm and counties, the negligible time-invariant components result in low *R2-adj* for this types of FE. Moreover, scaling emissions by a proxy for capital stock (or other proxies for size) results in substantial reductions in the explanatory power of all models, including those with facility FE.[2] We also show that including the first lag of emissions in these regressions increases the *R2-adj* to over 80% even in regressions without facility FE.

When we consider models with no persistence in emissions and large firm FE, facility FE do not capture such components, and scaling emissions by capital stock *increases* the

[2]If we reduce the AR(1) component to zero, the explanatory power of all the models shrinks to zero, confirming that it is the AR(1) component that is being captured by the different types of FE.

R2-adj. These discrepancies between the results of the DGP suggest that researchers can easily identify whether it is high persistence of actual time-invariant components that are driving the explanatory power of their models.

Distinguishing between time-invariant omitted variables at the facility level and dynamic effects due to an AR(1) component is important because, as highlighted by Breuer & deHaan (2024), including FE in the presence of dynamic effects can lead to bias due to correlation with past unobservable factors. This is likely to affect regressions where emissions are regressed against other variables which might also present an AR(1) component. In particular, since the AR(1) component at the facility level is not eliminated when aggregating emissions at the firm level, and firm-level variables such as firm value are also likely to exhibit AR(1)-type behavior, modelling the AR(1) component by using dynamic panel data models (e.g., Arellano & Bover (1995)) is likely to lead to different results relative to using models with firm FE. For this reason, in our firm-level analysis we also examine the impact of an AR(1) component in firm value.

We now describe the main results for our firm-level analysis. We run regressions where firm value is the dependent variable and firm profit and firm emissions are the independent variables. Even in this case, we focus on two cases: one where firm emissions have a negative impact on firm value and firm value has no AR(1) component; and one where firm emissions have a negative impact on firm value and there is moderate persistence (0.68). Moreover, we run the regressions both on the levels of firm emissions and firm emissions scaled by firm capital.[3] We show that, in the case without persistence in firm value, the regressions where firm-level emissions are in levels are consistent with the true simulated DGP: the estimated coefficient on firm-level emissions is unbiased and the

[3]For brevity, we use "capital" instead of "capital stock" in the subsequent discussion, unless otherwise stated.

statistical power of the tests are high, even for relatively small true (simulated) values of the slope coefficient on firm emissions. In fact, despite the fact that the our DGP implies leptokurtosis (to better simulate the behavior of our dataset) the distribution of the t-statistics is close to a Normal. We obtain very different results in the regressions where firm emissions are scaled by firm capital as an independent variable. These tests have low statistical power: the number of t-statistics lower than -1.96 (the threshold that would lead to rejection according to a two-tailed test assuming Normally-distributed t-statistics) drops substantially, leading to the possibility that researchers would incorrectly fail to reject the null hypothesis. Thus, scaling would increase the probability of Type II error in detecting a statistically significant impact of emissions on firm value.

When we scale both firm emissions and firm value by firm capital, for small simulated values of the (negative) coefficient on firm emissions the majority of the t-statistics are *positive*, rather than negative. This finding suggests that researchers might wrongly conclude that firm emissions have a positive (rather than negative) impact on firm value when they scale both firm value and firm emissions by a proxy for firm size. For larger values of the coefficient on firm emissions, the likelihood of failing to reject the null hypothesis decreases, since the proportion of negative t-statistics for the coefficient on emission intensity falls, and the number of positive t-statistics increases. However, researchers might still conclude that the relation between firm emissions and firm value is statistically insignificant, unless the value of the coefficient is very large. Moreover, we find that the distribution of the t-statistics does not converge to a Normal distribution when the magnitude of the negative simulated impact of emissions on firm value increases. Adding a scaled intercept in this regression, as is often done in the literature (e.g., Cohen & Zarowin (2010)), mitigates but does not solve the problem. In particular, while the bias decreases, the statistical power of the tests is low.

When we assume an AR(1) process in firm value, the results become biased even for regressions where firm value is regressed on firm emissions in levels, if the lag of firm value is excluded. Moreover, the statistical power of the tests is low. Adding the lag of firm value makes the coefficient on firm emissions consistent, although the statistical power of the test is not very high, even if a dynamic panel data model is employed.[4]

To investigate whether our main results are relevant for empirical applications, we evaluate the impact of the Californian Cap-and-Trade Program (CATP), which was implemented from 2013. Similar to previous literature (e.g., Bartram et al. (2022)), we consider facilities in California as the treated sample, and facilities out of California as a control. We show that regressions including the lag of facility-level emissions as an independent variable tend to have a negative coefficient, unlike those without it. This happens for regressions using estimators that allow for endogeneity of the lagged dependent variable (such as those by Arellano & Bover (1995) and Blundell & Bond (1998)) as well as Ordinary Least Squares (OLS) regressions with or without facility FE.

Our findings help to shed light on, at least, two strands of literature. First, we contribute to the literature on the relation between environmental performance and economic performance (Al-Tuwaijri et al. 2004). In particular, our findings complement recent studies on the impact of emission disclosure on firm value, or "value relevance" (Matsumura et al. 2014, Griffin et al. 2017). In this respect, we provide new insights regarding the potential impact of different types of FE on the explanatory power of the model, a key measure of association between accounting and market data (Brown et al. 1999).

[4]In particular, for this exercise we employ the approach developed by Arellano & Bover (1995), Blundell & Bond (1998).

Second, we contribute to the literature on the potential consequences of mis-specifying econometric models where different types of FE are included in the regressions without a theoretical justification (e.g., Breuer & deHaan (2024) and Plümper & Troeger (2019)). Unlike these studies, we narrow down the scope of the analysis to data on GHG emissions and we examine the extent to which their DGP is consistent with the existence of facility-level time-invariant components in emissions rather than dynamic effects. Moreover, to the best of our knowledge, we are the first to employ Monte Carlo simulations to mimic the distributional properties of both facility-level and firm-level data to better understand the potential impact of these mis-specifications on the results of empirical tests regarding firm value and firm emissions.

2 Main features of GHG emissions data and its relation with firm value

2.1 Brief description of the facility-level data

In this section, we explore some empirical regularities in emissions data at the facility level and describe briefly the main features of the databases employed. For our analysis, we use data from the Greenhouse Gas Reporting Program (GHGRP). As we explain in the Appendix, Section A, this database presents several advantages relative to other databases on GHG emissions.

We begin our empirical investigation from the most granular unit of analysis, which is the facility-level data. Total direct emissions provided by the GHGRP consist of a panel of 87,562 facility-level observations, spanning from 2010 to 2022. We then merge the information regarding the parent company structure of each facility, which is separately provided by the EPA. This results in a duplication of facility-year emissions data for facilities that are controlled by more than an ultimate owner. The structure of this dataset is facility-owner-year, rather than facility-year.

Due to the possibility of multiple observations for each facility-year combination, the sample size rises to 101,035 observations for 7,013 parents after the merge. Of the 101,035 facility-owner-year observations, 81,370 refer to observations where the facility has a unique ultimate owner, whereas the remaining 19,665 observations pertain to cases where the facility has more than one ultimate owner. Around 93% of the facilities have one owner only. To mitigate concerns related to double-counting of emissions, we run our regressions on the full sample of 101,035 observations either considering unadjusted levels of emissions or considering the share of emissions effectively "owned" by a specific parent (as per their corresponding ownership share of that specific facility in that specific year). In further analyses, we disregard parent-level information and look at the original sample of 87,562 facility-level observations.

For the analysis of the association between emissions and firm value for listed companies, we match the GHGRP dataset with financial data from Compustat (similarly to Li et al. (2024)) using a fuzzy-matching algorithm. To verify the accuracy of the fuzzy match, we manually inspect the dataset to rule out possible mistakes (in a similar fashion to Ivanov et al. (2024)). Overall, we match our dataset to 487 firms from Compustat. Table 1 reports the summary statistics for our main variables.

[insert Table 1]

2.2 Which types of FE matter?

The first step of our investigation consists of assessing whether using different types of FE has a differential impact on the explanatory power of regressions on emissions, whether unscaled or scaled by a proxy for firm size. Understanding whether different types of FE bear an impact on the coefficient of determination of a regression is important because it allows us to gauge the extent to which other variables, especially those that might be used to test a certain hypothesis, help explain the drivers of emissions. For example, in value relevance studies, *R2-adj* is a measure of the association between accounting and market data (Strong & Walker 1993, Brown et al. 1999, Barth et al. 2023).

We borrow a methodology that has previously been used in the governance literature (see, for instance Bertrand & Schoar (2003) and Schoar et al. (2024)). Specifically, we exploit the facility-year structure of our dataset to estimate eight linear regressions where the dependent variable is a proxy for GHG emissions (*Total Reported Emissions* in the GHGRP dataset), both unlogged (as in Seltzer et al. (2022)) and log-transformed,[5] and the independent variables are different combinations of FE types:

1. State FE and year FE
2. County FE and year FE
3. NAICS FE and year FE
4. County FE, NAICS FE and year FE
5. GVKEY FE and year FE (only for firms available in Compustat)

[5]Log transformations are frequently utilized and have also found application in other studies involving emissions data – see for instance Bartram et al. (2022). For consistency with the literature, we use both the natural logarithm of *Total Reported Emissions* and the natural logarithm of (*Total Reported Emissions*+1). The latter transformation allows to include cases for which *Total Reported Emissions* are equal to zero, but such practice has recently faced criticism from Chen & Roth (2024).

6. Firm FE and year FE

7. Facility FE and year FE

8. Facility FE, firm FE and year FE

We then examine how the *R2-adj* of the regression changes according to the type of FEs included. To examine the impact of possible duplicates due to multiple owners for a given facility, we run the regressions both on unadjusted emissions ("Total reported direct emissions" in the GHGRP database) and emissions adjusted by the ownership percentage for each recorded owner. Moreover, we also repeat the same analysis after collapsing the data at the facility-year level, rather than facility-firm-year level. In this case, however, we cannot examine the impact of firm FE.

In Table 2 we report the *R2-adj* obtained using three different versions of the dependent variable. Panel A reports regressions results of models that employ Total Direct Emissions as a dependent variable, while Panel B and C use $ln(Emissions + 1)$ and $ln(Emissions)$ respectively. These *R2-adj* are obtained from eight different specifications that include, by default, year FE. The columns in the table vary based on the type of additional fixed effect under consideration. More specifically, in each Panel, Column (1) reports the *R2-adj* obtained by simply adding state FE. Column (2) adds county FE, while Column (3) considers NAICS FE. Column (4) adds both County and NAICS FE in the same regression. Column (5) adds Gvkey FE, while Column (6) and (7) consider firm and facility FE respectively. Regressions in Column (8) add both firm and facility FE.

If we focus on Panel A, we observe that most of the variance is captured by facility FE alone, provided that their *R2-adj* is nearly 94%. The inclusion of firm FE, in addition to facility FE, improves only marginally the *R2-adj* which raises to just above 94%. Interestingly, the inclusion of firm FE alone does not seem to explain much of the variance;

if we look at Column (6), the related *R2-adj* is below 40%. Another couple of noteworthy results are those related to the inclusion of county FE (Column (2)) as well as County and NAICS FE (Column (4)); the *R2-adj* here reveal that such FE explain 69% and 70% of the variance, respectively.

[insert Tables 2]

If we adjust emissions to account for the share of them effectively owned by the corresponding reported parent companies, we observe that facility FE are still crucial drivers of the *R2-adj*. Looking at Column (7) of Table S1 in the Online Appendix, we find that the inclusion of facility FE alone explain more than 79% of the *R2-adj* across the three panels. Furthermore, the joint inclusion of firm and facility FE in Column (8) gives a boost to the *R2-adj* which jumps above 90% in both Panels A and C.

Finally, in Table S2 – reported in the Online Appendix – we disregard the ownership structure and simply look at the facility-year level data. In other words, at this stage of the analysis, we are looking at the original GHGRP dataset provided by the EPA, before the merge with the parent-companies information. Here we find that facility FE are still crucial in explaining the model's variance. If we zoom in on Column (5), Panel A, we find that the related *R2-adj* reaches 93%, suggestive of facility FE alone being able to explain most of the variance in our model when "Total reported direct emissions" is the dependent variable. If we focus on our log-transformed versions of emissions in Panels B and C, we observe that the *R2-adj* are about 77% and 85% respectively.

Overall, the findings from this section reveal that, irrespective of the way emissions are defined, facility FE as well as the combination of firm and facility FE are crucial determinants of a model's variance. This implies that failing to account for unobserved facility-level characteristics can undermine the reliability of results when appraising the impact of emissions in quantitative studies.

2.3 Scaling emissions by a firm-size proxy

In this section of the paper, we examine the impact of scaling firm-level emissions by a proxy for firm size. Overall, the aim of this section is to highlight how emissions – depending on how they are defined – can differently influence firm valuation. Since scaling is usually done using financial data (Thomas et al. 2022, Hsu et al. 2023), we match emissions data with financial data retrieved from Compustat.

Scaling variables by a proxy for size is common in the finance and accounting literature, despite the methodological problems this practice entails, known since the 70s (Lev & Sunder 1979). We explore the potential issues that might arise when scaling emissions with a proxy for firm size, such as total assets, sales, a proxy for capital stock (PPENT in Compustat), or market value of equity. As a dependent variable, we choose Tobin's Q.

In a similar vein to the analysis in Section 2.2, we start from facility-level data. Specifically, tables 3–4 report the results of regressions where a firm's Tobin's Q is the dependent variable and proxies for emissions are the main independent variable.[6] In Table S3 in the Online Appendix, we report the results of regressions after aggregating facility-level data at the firm level. Therefore, our dependent variable is the natural logarithm of Tobin's Q – defined as the market value of assets over the book value of assets (Byun et al. 2021, Upadhyay & Öztekin 2021) – while our key regressors are either the natural logarithm of emissions (columns (1)–(2)), unscaled emissions (columns (3)–(4)),

[6] In this dataset, each observation reports the actual facility-level emissions a parent is responsible for in a given year and for a given facility. For instance if, in a given year, a facility is wholly owned by one parent only, our dataset will associate the entirety of that facility's emissions to the sole owner. Similarly, if a facility is owned by four parent companies – each holding a quota of 91.8%, 4.1%, 3.4%, and 0.7%, respectively – the facility's emissions will be split according to these quotas and associated with the corresponding parent firm's financial data.

emissions scaled by total assets (columns (5)–(6)), emissions scaled by total sales (columns (7)–(8)), emissions scaled by Total Property, Plant and Equipment (PPENT) (columns (9)–(10)), and emissions scaled by the Market value of Equity (ME) (columns (11)–(12)). Standard firm-level controls are added to specifications reported in even columns, and include Size (the natural logarithm of total assets), Total debt ratio (total debt over total assets), ROA (net income over total assets), Sales ratio (sales over total assets), Cash ratio (cash and short term investments over total assets), Firm age (the natural logarithm of the difference between the current year and the first year a firm was recorded in Compustat, similarly to Adams et al. (2021), Dagostino et al. (2023), and Yim (2013)), and Multi-owners (a dummy equal to one when facilities are owned by more than one owner).

The results reveal that, depending on the proxy used, emissions differently correlate with firm valuation. More specifically, emissions scaled by total sales are negatively associated with firm valuation in Table 3; however this effect disappears once we include firm FE in the next Table 4. Such a finding once again supports the view that some of the variance in our models could be explained by unobserved characteristics that can only be captured by some FE, whose exclusion would lead to wrong inferences. Despite such criticalities, it is worth noting that, when emissions are scaled either by total assets (Column (6)) or Market value of equity (ME) (Columns (11)–(12)), their coefficients consistently enter negative and significant in Tables 3–4, as well as in Table S3 reported in the Online Appendix.

[insert Tables 3 and 4]

2.4 Discussion of the results

The results in this section suggest that facility FE explain a very large portion of variation in facility-level emissions. This finding suggests that emissions tend to be very stable over time at the facility level. In turn, this begs the question of whether an AR(1) component should be allowed for in regressions on facility-level emissions. To understand whether facility FE are capturing high persistence in facility emissions, in Table S4 in the Online Appendix we report the results of regressions where the dependent variable is facility-level emissions. In this table, we compare specifications with facility FE with specifications without facility FE, and we also consider the impact of adding the first lag of emissions.[7] The results in this table suggest that adding the first lag of emissions explains almost all the explanatory power of the model with both facility FE and year FE.[8]

Overall, the results in Section 2.3 lend support to the idea that the way our proxies for emissions are constructed is not inconsequential. Furthermore, considering the results in Section 2.3 in conjunction with those in Section 2.2, we note that county-level FE (or other types of regional characteristics) can be an important driver of firm value. Discarding information on facility-level NAICS might also lead to severe loss of information which might impair the validity of inferences regarding the relationship between emissions and firm value. Finally, due to the drop in observations in regressions using different firm-level controls, it is unclear whether some of the insignificant results are due to sample-selection

[7]We lose the observation pertaining to the first period for each facility once we impose that the first lag of emissions be available.

[8]Specifically, we start with a regression including facility FE and year FE (column (1)). The *R2-adj* for this regression is 0.9214. Then, to make the regressions with and without the first lag of emissions comparable, we run the same regression after excluding observations for which it is missing. In columns (2) and (3) we compare what happens to the explanatory power of the model when we exclude (column (2)) and include (column (3)) the first lag of emissions in regressions with facility FE and year FE. The *R2-adj* is 0.9283 and 0.9647, respectively. In column (4) we report the results for regressions with the first lag of emissions but without facility FE: the *R2-adj* is 0.9604. Finally, in column (5), we report the results using a dynamic panel data model. While this type of models do not allow us to report the *R2-adj*, the results in column (5) allow us to estimate the coefficient on the first lag of emissions more precisely: the coefficient is 0.956, very similar to that of column (4).

bias due to data availability for certain Compustat items, rather than a lack of correlation between emissions and firm value.

3 Monte Carlo simulations

In this section, we examine, using Monte Carlo simulations, several empirical properties uncovered in the datasets employed in Section 2. In particular, we study: *i*) how facility-level autocorrelation in emissions affects the contribution of different types of FE on the R-squared (adjusted) of regressions where emissions are the dependent variable (Section 3.1); *ii*) how, in firm-level regressions on firm value, scaling emissions using a firm-size proxy affects the estimated slope coefficient and t-statistic for the emissions proxy (Section 3.2). We describe the details of the Data Generating Process (DGP) of our Monte Carlo simulations in the Appendix (Section B).

The key firm-level equation in our simulations is that for Firm Value of firm f in period t, V_{ft}. This equation assumes autocorrelation in Firm Value, a positive correlation between Firm Value and Firm Profit (Π_{ft}), and a negative correlation between Firm Value and Firm Emissions (G_{ft}), For convenience, we report this equation here (the same as equation (A4) in the Appendix):

$$V_{ft} = \varepsilon_{ft} + \rho_V V_{i(t-1)} + 20\Pi_{ft} + \psi_G G_{ft} + \mu_t + \mu_f \tag{1}$$

where μ_t simulates the effect of time-varying unobserved factors that might affect all firms and μ_f simulates the impact of time-invariant firm-specific unobserved factors. After simulating the data, using 1,000 replications, we then run regressions with firm (α_f) and year (α_t) FE, based on equation (1). We start from regressions where none of the variables is scaled by Firm Capital:

$$V_{ft} = \beta_0 + \beta_1 \Pi_{ft} + \beta_2 G_{ft} + \alpha_t + \alpha_f + \mu_{ft} \tag{2}$$

where α_t are time FE and α_f are firm FE. Note that we are controlling for Firm Profit, and thus for any size effect, since profit is positively correlated with Firm Capital (see equation (A3)). Because Firm Capital does not directly affect Firm Value, our estimates should not suffer from omitted variable bias, as long as we include in our regression Firm Profit, Firm Emissions, and there is no autocorrelation in Firm Value (i.e. $\rho_V = 0$). If $\rho_V \neq 0$, then omitted variable bias will be present.

We then run a regression where we scale Firm Emissions by Firm Capital, while Firm Value and Firm Profit remain unscaled:

$$V_{ft} = \beta_0 + \beta_1 \Pi_{ft} + \beta_2 (G_{ft}/K_{ft}) + \alpha_t + \alpha_f + \mu_{ft} \tag{3}$$

Third, we scale all three variables by Firm Capital:

$$(V_{ft}/K_{ft}) = \beta_0 + \beta_1 (\Pi_{ft}/K_{ft}) + \beta_2 (G_{ft}/K_{ft}) + \alpha_t + \alpha_f + \mu_{ft} \tag{4}$$

Finally, as it is sometimes done in the accounting literature (Cohen & Zarowin 2010, Louis 2003), we also scale the constant by Firm Capital

$$(V_{ft}/K_{ft}) = \frac{\beta_0}{K_{ft}} + \beta_1 (\Pi_{ft}/K_{ft}) + \beta_2 (G_{ft}/K_{ft}) + \alpha_t + \alpha_f + \mu_{ft} \tag{5}$$

We cluster the standard errors of all regressions by firm.

3.1 Explanatory power of different types of FE on unscaled and scaled emissions

3.1.1 Case 1: High persistence in emissions and negligible FE

Table 5 reports the results of our simulations investigating the impact of different types of FE on the variance of emissions, both unscaled (facility-level Emissions) and scaled by capital stock (facility-level Emissions/Capital). In addition to year FE (included in all regressions), the first specification adds facility FE, the second one firm FE and county FE, the third one firm FE, and the fourth one county FE. Importantly, we have allocated a very small weight to the time-, firm- and county-specific time invariant components of Capital and Emissions (with a mean of 0.000005).

In this exercise we focus on facility-level regressions. Therefore, the relevant parameter values for the simulations in this section are: $\rho_g = \rho_k = 0.92$, $\beta_k = 0.01$, $k_{\Gamma 1} = 0.1$, $k_{\Gamma 2} = 1500$, $g_{\Gamma 1} = 0.05$, and $g_{\Gamma 2} = 150$.

For the results in Panel A and B, we impose correlation between allocation of firms and facility-level capital, but not for counties. We do this by sorting facilities by average k_{ifct} before generating the fictitious firms. In Panel C, instead, we allocate the (hypothetical) facilities randomly to both firms and counties and compare our results for regressions with and without the first lag of Emissions ($g_{ifc(t-1)}$).

As reported in the beginning of Panel A, the estimated AR(1) coefficient for facility-level emissions ($\hat{\rho_g}$=0.9197) is very close to the true value (ρ_g=0.92). In Panel A, the results for unscaled facility-level Emissions confirm that persistence in facility-level emissions leads to a high *R2-adj* for specifications with facility FE. The average *R2-adj* is almost 0.68, although we have not specifically inserted facility FE in equation A2. The specifications with other types of FE (firm FE, county FE) have a much lower *R2-adj*.

The *R2-adj* for the specification using firm FE and year FE is around 0.33, almost twice as much as those using county FE and year FE (*R2-adj*=0.189). In fact, the explanatory power in the specification with county FE and year FE comes from the year FE, due to the persistence in capital stock (as said above, $\rho_k = 0.92$).[9]

Since, as said above, in our DGP we sorted facilities by average k_{ifct} before allocating them to firms, and the county-specific time-invariant component is by construction orthogonal to both the firm component and capital stock, firm FE explain a much larger portion of variation in facility-level emissions. However, this is not because of an actual time-invariant firm-specific effect: as already mentioned, the size of these components is negligible. The reason for such a high *R2-adj* for the specifications with firm FE stems from the AR(1) component in capital and the non-random allocation of facilities to firms. This might be considered as a potential "channel" through which omitted variable bias affects regressions without the lag of Emissions.

The results in Panel B suggest that using Emissions/Capital results in a much lower level of *R2-adj*, even when facility-level emissions are included: the *R2-adj* for the specification including facility FE and year FE drops to around 0.06 in Panel B, and for the other specifications they drop to less than 0.02 (Panel B). Therefore, the impact of autocorrelation in Emissions on *R2-adj* is eliminated by the scaling.

In Panel C, we investigate the impact of including the lag of facility-level emissions in the regressions. We also randomize the allocation of facilities to both counties and firms, to show how this affects the explanatory power of county FE and firm FE. The results in Panel C confirm that including the lag of Emissions raises the *R2-adj* to a level

[9] In fact, eliminating the AR(1) component would result in a reduction of the explanatory power of the model to virtually zero. To show what happens as autocorrelation in emissions drops, we run simulations even for $\rho_g = 0$ and $\rho_g = 0.5$. We report the results in Table S6.

higher than that using only facility FE. Moreover, the *R2-adj* is over 0.80 even for the regressions without facility FE. In particular, for both the regressions using firm FE and county FE, the *R2-adj* for the regressions without the lag have are around 0.18. Thus, adding $g_{ifc(t-1)}$ increases the explanatory power of the regressions by over 60%.

[insert Table 5]

The results in this section confirm that an AR(1) component in emissions might lead to high values in *R2-adj* of regressions with facility FE. The results for firm FE and county FE, in conjunction with those in Tables 2 and S1 (provided in the Online Appendix), suggest that the large explanatory power of county FE in our sample, relative to firm FE, might be due to larger facilities clustering around certain regions. In fact, in our Monte Carlo simulations the random allocation of facilities to counties and the non-random allocation to firms in Panels A and B leads to a much lower explanatory power for county FE than for firm FE. Finally, the results in Panel C suggest that including facility FE might not capture all the variation in the Emissions due to autocorrelation.

3.1.2 Case 2: No persistence in emissions and large firm FE

This section studies the impact of large time-invariant firm-specific components on the explanatory power of regressions on Emissions. Similar to Panel C of Table 5, we randomize the allocation of facilities to both firms and counties. However, rather than a very small weight to all the time-invariant components of Emissions, we simulate a much larger weight to firm-specific time invariant components. To be more precise, we impose a weight in the firm-specific time invariant component of k_{ifct} and g_{ifct} such that it should explain around 50% of the variation in Emissions.[10] Moreover, unlike Table 5, we consider $\rho_k = \rho_g = 0$.

[10]To do this, we implemented a procedure similar to that of Petersen (2008).

We report the results of this exercise in Table 6. For each of the 1,000 replications, the dependent variable is either Emissions (Panels A and B) or Emissions/Capital (Panel C). Neither firm FE nor county FE are correlated with capital, but the large firm-specific time-invariant component results in an *R2-adj* of 0.49 for the specification with firm FE in Panel A, where we include the lag of Emissions (g_{ifct}) in the regressions. In Panel B, where the lag is excluded, the *R2-adj* remains very close to 0.49 (0.4899). For the other types of FE, the explanatory power is very low, especially for the specifications with county and year FE, both in Panel A and B. In Panel B, even the explanatory power for the specification with facility FE and year FE is zero.

In Panel C we report the results for Emissions/Capital, without the lag of Emissions. Similar to Panel B, the specifications without firm FE have an *R2-adj* of zero, while the others have an explanatory power even larger than in Panels A and B (0.6664).

[insert Table 6]

3.2 Scaling emissions in regressions on firm value

In this section, we report the results of our Monte Carlo simulations concerning the impact of different scaling assumptions on inferences regarding the impact of emissions on firm value.

3.2.1 Case 1: No autocorrelation in firm value

In Table 7 we report the simulation results where we regress firm value on emissions according to equations (2)–(4). We start our discussion assuming a DGP for Firm Value without autocorrelation ($rho_V = 0$). Consistent with our empirical results and studies that suggest investors penalize firms with high carbon emissions, we choose a negative coefficient for our simulations: $\psi_G = -0.0004$. Thus, the parameter values for the simulations in Table 7 are: $\rho_g = \rho_k = 0.92$, $\beta_k = 0.01$, $k_{\Gamma 1} = 0.1$, $k_{\Gamma 2} = 1500$, $g_{\Gamma 1} = 0.05$, $g_{\Gamma 2} = 150$, $\Pi_{\Gamma 1} = 0.5$, $\Pi_{\Gamma 2} = 1200$, $\beta_K = 0.0001$, $\rho_V = 0$, and $\phi_G = -0.0004$.

Panel A reports the average simulated median and kurtosis of the main variables used in the simulations and their corresponding real-data counterparts (from Table 1): G_{ft}, K_{ft}, G_{ft}/K_{ft} and V_{ft}.

While we do not aim to match the moments of our simulated data to those of the actual sample — since our objective is to understand how scaling might affect inferences regarding t-statistics — the average median value of our simulated data tends to be comparable to the actual data in our sample. For example, the average median value for G_{ft} (Firm Emissions) is around 525 and the average median value for K_{ft} (Firm Capital) is around 3500, leading to a ratio of Firm Emissions to Firm Capital with an average median value of around 0.12 (in our sample, it is 0.15). The average kurtosis for G_{ft} and K_{ft} is around 8.5 and 11, respectively. This indicates that these variables are leptokurtic, but the degree of leptokurtosis is lower than for our sample (around 34 for G_{ft} and around 43 for K_{ft}). Despite this, the kurtosis for the ratio G_{ft}/K_{ft} is above 250 (795 in our sample).[11] This result is important because it suggests that scaling may exacerbate leptokurtosis and researchers might be unaware of this problem, since kurtosis is often unreported in descriptive statistics tables. Finally, the average median value for V_{ft} is comparable to that in our sample (around 5500 and 5100, respectively), and its kurtosis is around 15 (around 250 in our sample).

Panel B reports the results for the estimated coefficient for G_{ft} (β_2) and its t-statistic in regressions run according to equation (2). We report the following descriptive statistics

[11] The distribution of the ratio of two Normal variables has been subject of numerous studies for decades (Geary 1930, Fieller 1932, Hinkley 1969). However, a closed-form of the distribution of the ratio of two Normal variables has not been provided (Nguyen et al. 2019).

for both the slope coefficient and its t-statistic: number of observations (1,000 for all cases), mean, standard deviation (Std. Dev.), skewness, kurtosis, first percentile (P1), median (P50), and 99^{th} percentile (P99). As expected, the average estimated coefficient is very close to the true one (-0.0004), and the t-statistics are all negative (the untabulated minimum value is -0.9776998). The t-statistics are lower than -1.96 in 967 cases out of 1,000, suggesting that researchers would most likely conclude that firm emissions decrease firm value if they assumed a Normal distribution for the t-statistics.

As reported in Panel C, when we run the regressions according to equation (3), where G_{ft} is replaced by G_{ft}/K_{ft}, the average t-statistic for β_2 is around -0.99. The number of cases for which the t-statistic for G_{ft}/K_{ft} is < -1.96 is 214 (untabulated). Thus, scaling emissions by a size proxy does not necessarily lead to the same conclusions as without scaling. Researchers might conclude that there is no statistically significant relation between firm emissions and firm value, even when the true relationship is negative and statistically significant.

The problem becomes even worse when both the dependent and the explanatory variable are scaled by a size proxy. As shown in Panel D of the table, scaling all variables by K_{ft} leads to a *positive* average t-statistic: 2.9108, and in 803 cases the t-statistic for G_{ft}/K_{ft} is > 1.96. Thus, the probability of wrongly inferring that emission intensity positively relates to firm value, while the actual relationship is negative, is very high when using equation (4). In fact, only in 22 cases out of 1000 do we have negative slope coefficients and t-statistics.[12]

Finally, in Panel E, although the bias of the coefficient is lower than for Panel D, thanks to the inclusion of a constant term scaled by Firm Capital, the rejection rates are

[12]The number of negative slope coefficients and t-statistics increases as the magnitude of the true value of ψ_G increases. For example, for $\psi_G = -0.0008$, we obtain 35 cases out of 1,000. See Figure 1 for a more exhaustive investigation of this issue.

still quite low: the average t-statistic on β_2 is 0.3563. The number of t-statistics for β_2 that are < -1.96 is 102 (untabulated). Thus, the rejection rate is still much lower than for equation (2).[13]

[insert Table 7]

In Figure 1 we run the simulations considering different values for ϕ_G and we run regressions according to equation (4) to investigate how the probability of rejecting the null hypothesis with the right sign of the coefficient on G_{ft}/K_{ft} is affected by the true value of ϕ_G. We simulate our regressions using $\phi_G = -0.004, -0.0036, ..., -0.0004, 0.0000$. Then, we count the number of t-statistics for the coefficient on G_{ft}/K_{ft} that are negative (top panel) and the number of t-statistics that are lower than -1.96 (center panel). Moreover, we also report the value of the skewness of the distribution for each case. As expected, the number of negative t-statistics increases as the true value of ϕ_G becomes more negative, and so does the number of t-statistics lower than -1.96. Thus, as the magnitude of the negative impact of G_{ft} on V_{ft} increases, the probability of obtaining a positive coefficient on G_{ft}/K_{ft} in regressions where Firm Emissions and Firm Value are scaled by Firm Capital (such as in equation (4)) decreases. However, as shown in the center panel of Figure 1, even for $\phi_G = -0.004$ the null hypothesis of $\beta_2 = 0$ is rejected in less than 200 cases out of 1,000. Finally, the bottom graph suggests that, as ϕ_G becomes more negative, the skewness of the distribution of the t-statistics becomes negative, instead of positive. Such relation between skewness and ϕ_G is monotonic and suggests that the distribution of the t-statistics does not tend to a Normal as the true value of ϕ_G becomes more negative.

[13] In Figure S1, we report the kernel densities associated with the slope coefficients and t-statistics of G_{ft} and G_{ft}/K_{ft} in Panels B, C, D, and E of Table 7. The inferences from the graphs are consistent with the descriptive statistics in the table. In particular, while the t-statistics for Panel C (equation(3)) are relatively close to a Normal distribution, the ones for Panel D are positively skewed and leptokurtic. For Panel E, which corresponds to the regression run according to equation (5) the probability of obtaining a positive coefficient is still relatively high.

These findings suggest that in studies that find a negative impact of emissions on firm value for regressions in levels but not after scaling both firm value and emissions by a proxy for firm size, the magnitude of the true impact of emissions on firm value might be small. Moreover, as the magnitude of the coefficients increases, the distribution of the t-statistics does not converge to a Normal distribution. Therefore, obtaining critical values via bootstrapping exercises might be required.

[insert Figure 1]

3.2.2 Case 2: Positive autocorrelation in firm value

What happens if we assume the same value ϕ_G, but we change the assumptions for the autocorrelation parameter in Firm Value?

Looking at our DGP, it is clear that we are assuming a positive correlation between Firm Capital and Firm Emissions ($\beta_k = 0.01$ in equation (A2)). Moreover, both Firm Capital and Firm Emissions are highly persistent ($\rho_g = \rho_k = 0.92$), and Firm Emissions directly affect Firm Value ($\phi_G = -0.0004$). This means that, if there is autocorrelation in Firm Value, there will be an omitted variable bias problem even in regressions run according to equation (2), because $cov(G_{ft}, V_{f(t-1)}) \neq 0$.[14] Deriving the actual extent of the bias is not straightforward *a priori* because ρ_g in our simulations is for facility-level emissions (g_{ifct}), instead of Firm Emissions (G_{ft}).

In Table S5 we repeat our analysis maintaining $\phi_G = -0.0004$, but we assume $\rho_V = 0.68$. Moreover, to keep the other features of the distribution similar to the ones for Table 7, we reduce the scale parameter for Firm Profit in equation (A3) from 0.5 to 0.17. We keep the other parameter values unaltered. Thus, the parameter values for the simulations in

[14]This results from $cov(G_{f(t-1)}, V_{f(t-1)}) < 0$ and $cov(G_{ft}, G_{f(t-1)}) > 0$.

Table S5 are: $\rho_g = \rho_k = 0.92$, $\beta_k = 0.01$, $k_{\Gamma 1} = 0.1$, $k_{\Gamma 2} = 1500$, $g_{\Gamma 1} = 0.05$, $g_{\Gamma 2} = 150$, $\Pi_{\Gamma 1} = 0.17$, $\Pi_{\Gamma 2} = 1200$, $\beta_K = 0.0001$, $\rho_V = 0.68$, and $\phi_G = -0.0004$.

The results in Panel B suggest that neglecting autocorrelation in Firm Value leads to a biased coefficient even when Firm Emissions are not scaled. Interestingly, the autocorrelation in Firm Value also leads to a negative skewness for the t-statistics, despite a positive (and near zero) skewness for the coefficient. Overall, the results in this panel suggest that researchers would be likely to conclude that there is no correlation between Firm Emissions and Firm Value.

Panels (C–E) are similar, in that the mean t-statistic tends to be positive and lower than 1.96, although for Panel E the median t-statistic is 1.625.

In addition to the equations (2-5), we report the results for a fifth equation where equation (2) is augmented with a term allowing for the positive autocorrelation coefficient in Firm Value:

$$V_{ft} = \beta_0 + \beta_1 \Pi_{ft} + \beta_2 G_{ft} + \beta_3 V_{f(t-1)} + \alpha_t + \alpha_f + \mu_{ft} \tag{6}$$

We estimate regression equation (6) using a dynamic panel data model often used by researchers: the so-called GMM-in-System by Arellano & Bover (1995), Blundell & Bond (1998). In particular, similar to Flannery & Hankins (2013) and Zhou et al. (2014), we use the Stata command “xtdpdsys”, with robust standard errors.[15] As expected, once we include the first lag of Firm Value as an independent variable, the estimated β_2 converges towards its true value (-0.0004). Nevertheless, the power of the test is not very high: the rejection rate considering as a threshold a t-statistic of -1.96 is 40.2% (402 cases out of 1,000).

[15]This command does not allow the an option with clustered standard errors at the firm level.

3.2.3 Discussion of the results

The results in Section 3.1.2, as compared with those in Section 3.1.1, indicate that high persistence in facility-level Emissions should lead to high *R2-adj* in regressions with facility FE even when facility-specific time-invariant omitted variables do not exist. Scaling Emissions by Capital leads to a large drop in the *R2-adj*. On the other hand, large firm-specific time-invariant components in Emissions does not lead to large *R2-adj* in regressions with facility FE, as long as firm FE are not included. Scaling Emissions by Capital leads to an *increase* in the *R2-adj* for regressions with firm FE.

The analysis of the impact of scaling emissions in Sections 3.2.1 and 3.2.2 suggests that using emissions intensity leads to two problems: from a conceptual perspective, it is hard to disentangle the impact of the numerator from that of the denominator; from a statistical perspective, scaling leads to a loss in statistical power when the null hypothesis of a slope coefficient statistically indistinguishable from zero should be rejected. Moreover, an investigation of the statistical properties of actual data on firm value suggests that the DGP might contain an AR(1) component. Since the first lag of firm value is likely to be correlated with other independent variables in the regression, regressions without the lag of firm value might suffer from omitted variable bias. In these sections, we have not simulated the potential effect of taking the logarithm of Firm Value and Firm Emissions, as is sometimes done in the literature. However, in addition to changing the interpretation of the results for the slope coefficients,[16] the log-transformation can result in a substantial drop in the statistical power of the model (see Table S7 in the Online Appendix).

[16]Coefficients in a log-log model can be interpreted as elasticities.

4 Application: Role of autocorrelation in facility emissions

In this section we investigate the impact of omitting the lagged dependent variable in facility-level regressions on emissions. This empirical application considers the Cap-and-Trade Program (CATP), implemented in 2013 in California, as a regulation that might have impacted emissions in treated facilities. While we do not attempt to replicate a particular paper, this regulation has been used in several studies, such as Bartram et al. (2022), Ivanov et al. (2024), Griffin et al. (2020) and Mascia & Onali (2024).

We start with specifications considering a simple Difference-in-Differences (DiD) approach, where facilities in California are treated (*Treated* is equal to one) and facilities in other states are not (*Treated* is equal to zero). The variable *Post* is zero for 2010–2012, and one for 2013–2022. Our baseline equation is as follows:

$$g_{it} = \alpha_0 + \alpha_1 Treated_i + \alpha_2 Post_t + \alpha_3 Treated_i \times Post_t + \rho gi(t-1) + \gamma_i + \gamma_t + \varepsilon_{ft} \quad (7)$$

To use the dynamic panel data model, we need to collapse the observations at the facility-year level. We use four different specifications for equation (7): a dynamic panel data model, estimated using Arellano & Bover (1995) and Blundell & Bond (1998); an OLS model with facility FE and without the lag of the dependent variable; an OLS model with facility FE and the lag of the dependent variable; an OLS model without facility FE but with the lag of the dependent variable. As can be seen from equation (7), all the specifications include year FE. Since adding the lag of the dependent variable reduces the available number of periods for each facility, we use $t-1$ periods for all regressions, even those without the lag as a regressor. In the regressions with facility FE, we do not include the time-invariant dummy "Treated" because it is subsumed by the facility FE.

As we said earlier, one of the reasons for persistence in facility-level emissions is that owners of facilities change their overall firm-level emissions by buying or selling facilities, instead of changing the levels of emissions in each facility they own. We also said that around 11% of the observations in our sample pertain to cases where the main owner of a facility changes. To investigate whether changes in the ownership structure of a facility affect emissions and how such changes interact with the CATP, we also run the following triple-DiD model:

$$\begin{aligned} g_{it} =& \alpha_0 + \alpha_1 Treated_i + \alpha_2 Post_t + \alpha_3 Treated_i \times Post_t + \rho g_{i(t-1)} + \alpha_4 Change\ in\ Parent\ Firm_{it} \\ &+ \alpha_5 Change\ in\ Parent\ Firm_{it} \times Post_t + \alpha_6 Change\ in\ Parent\ Firm_{it} \times Treated_i \\ &+ \alpha_7 Change\ in\ Parent\ Firm_{it} \times Treated_i \times Post_t + \gamma_i + \gamma_t + \varepsilon_{ft} \end{aligned} \tag{8}$$

Table 8 reports the results of regressions on facility-level emissions run using: the System-GMM (columns (1) and (5)); the model with facility FE but without $g_{i(t-1)}$ (columns (2) and (6)); an OLS model with facility FE and $g_{i(t-1)}$ (columns (3) and (7)); an OLS model without facility FE but with $g_{i(t-1)}$ (columns (4) and (8)).

The results show that, regardless of whether we use the System-GMM estimator or an OLS with facility FE, adding the first lag of the dependent variable leads to a negative value for $alpha_3$, the coefficient on the interaction term $Treated \times Post$. When we do not include $g_{i(t-1)}$, instead, the coefficient is positive and, in columns (2) and (6), $alpha_3$ is even statistically significant. Therefore, these results suggest that the choice of whether adding the first lag of the dependent variable or not does have an impact on inferences regarding the effect of the CATP.

The coefficient ρ is always strongly significant and lies between 0.94 and 0.967 when we do not include facility FE. Adding facility FE decreases the magnitude of the coefficient by around 0.241.

Finally, the results for columns (5)–(8) suggest that there may be heterogeneous responses to the CATP related to changes in facility ownership. The coefficient α_7 is positive and statistically significant in columns (5) and (8). This indicates that for Californian facilities that change owner, the CATP has a weaker negative effects than for Californian facilities that do not experience changes in ownership.

[insert Table 8]

5 Conclusions

The literature on GHG emissions is vast and the econometric methods used are highly heterogeneous. In this paper, we provide some insights regarding issues that might arise while examining facility-level data on GHG emissions. First, we show that a very large portion of variation in facility-level emissions can be explained by facility-level FE. Thus, emissions tend to be stable over time at the facility-level. Moreover, we observe that firm FE appear less important than county FE. This finding is not inconsequential because, by collapsing observations at the firm level, researchers might discard important factors underlying such variation, such as local economic conditions. Moreover, such a high degree of stability in facility-level emissions might indicate that firms adopt a portfolio-level approach: rather than reducing/increasing the average level of emissions in facilities, they might sell/buy ownership stakes in facilities. Since the fraction of facilities with multiple owners in the GHGRP database is over 10%, this finding suggest that using other databases, without granular information on the owners of a facility, might lead to potentially wrong inferences regarding the impact of firm strategies on emissions.

We then study whether facility FE might actually capture, at least in part, the effect of an autoregressive component in facility emissions. We thus examine, using Monte Carlo simulations, the impact of including the lagged dependent variable in regressions on facility-level emissions, as opposed to different types of FE. We find that the explanatory power of econometric models using facility FE might be a result of an autoregressive component in facility-level emissions: although we do not include any time-invariant facility-specific component in our simulations, regressions with facility FE have high explanatory power (*R2-adj*>0.65) if the AR(1) coefficient is around 0.9. Neglecting such an AR(1) component bears important implications for researchers and policymakers alike. In an empirical application aimed at evaluating the impact of the Californian Cap-and-Trade Program on emissions, we find that including the lag of the dependent variable results in coefficients with an opposite sign relative to those without the lag.

Third, following a recent debate in the finance literature on the use of emission intensity (emissions scaled by a proxy for firm size) instead of the levels of emissions, we show that such scaling leads to highly leptokurtic distributions even when firm-level emissions and the scaling variable are only mildly leptokurtic. This finding is also very important once we consider regressions where firm value (or a proxy such as Tobin's Q) is regressed on firm emissions. Using Monte Carlo simulations, we show that using emission intensity, rather than firm emissions, can lead to a severe drop in statistical power. To the best of our knowledge, we are also the first to show that, when the true coefficient on unscaled emissions is small, scaling both firm value and emissions can lead to large Type II errors and even statistically significant coefficients with the wrong sign. This finding is important even for the literature on "value relevance", which examines the relation between market data and the corresponding accounting variables. While in this literature it is known that scaling can be inappropriate, this practice is still common, although often it is also done in conjunction with scaling of the intercept. However, our Monte Carlo simulations show that even this practice can be problematic in the presence of highly leptokurtic variables and ratios, as is the case with emission intensity. In Table 9, we provide a summary of the main econometric issues researchers may face in empirical studies and offer some possible solutions.

[insert Table 9]

References

Adams, R. B., Ragunathan, V. & Tumarkin, R. (2021), 'Death by committee? An analysis of corporate board (sub-) committees', *Journal of Financial Economics* **141**(3), 1119–1146.

Al-Tuwaijri, S. A., Christensen, T. E. & Hughes Ii, K. (2004), 'The relations among environmental disclosure, environmental performance, and economic performance: a simultaneous equations approach', *Accounting, Organizations and Society* **29**(5-6), 447–471.

Arellano, M. & Bover, O. (1995), 'Another look at the instrumental variable estimation of error-components models', *Journal of Econometrics* **68**(1), 29–51.

Barth, M. E., Li, K. & McClure, C. G. (2023), 'Evolution in value relevance of accounting information', *The Accounting Review* **98**(1), 1–28.

Bartram, S. M., Hou, K. & Kim, S. (2022), 'Real effects of climate policy: Financial constraints and spillovers', *Journal of Financial Economics* **143**(2), 668–696.

Bertrand, M. & Schoar, A. (2003), 'Managing with style: The effect of managers on firm policies', *The Quarterly Journal of Economics* **118**(4), 1169–1208.

Blundell, R. & Bond, S. (1998), 'Initial conditions and moment restrictions in dynamic panel data models', *Journal of Econometrics* **87**(1), 115–143.

Breuer, M. & deHaan, E. (2024), 'Using and interpreting fixed effects models', *Journal of Accounting Research* **62**(4), 1183–1226.

Brown, S., Lo, K. & Lys, T. (1999), 'Use of R2 in accounting research: Measuring changes in value relevance over the last four decades', *Journal of Accounting and Economics* **28**(2), 83–115.

Byun, S. K., Fuller, K. & Lin, Z. (2021), 'The costs and benefits associated with inventor CEOs', *Journal of Corporate Finance* **71**, 102094.

Chen, J. & Roth, J. (2024), 'Logs with zeros? Some problems and solutions', *The Quarterly Journal of Economics* **139**(2), 891–936.

Cogley, T. & Nason, J. M. (1995), 'Output dynamics in real-business-cycle models', *The American Economic Review* **85**(3), 492–511.

Cohen, D. A. & Zarowin, P. (2010), 'Accrual-based and real earnings management activities around seasoned equity offerings', *Journal of Accounting and Economics* **50**(1), 2–19.

Dagostino, R., Gao, J. & Ma, P. (2023), 'Partisanship in loan pricing', *Journal of Financial Economics* **150**(3), 103717.

Erickson, T. & Whited, T. M. (2012), 'Treating measurement error in Tobin's q', *The Review of Financial Studies* **25**(4), 1286–1329.

Fieller, E. C. (1932), 'The distribution of the index in a normal bivariate population', *Biometrika* **24**(3/4), 428–440.

Flannery, M. J. & Hankins, K. W. (2013), 'Estimating dynamic panel models in corporate finance', *Journal of Corporate Finance* **19**, 1–19.

Garavaglia, S., Van Landuyt, B. W., White, B. J. & Irwin, J. (2023), 'The ESG stopping effect: Do investor reactions differ across the lifespan of ESG initiatives?', *Accounting, Organizations and Society* p. 101441.

Geary, R. C. (1930), 'The frequency distribution of the quotient of two normal variates', *Journal of the Royal Statistical Society* **93**(3), 442–446.

Griffin, P. A., Lont, D. H. & Sun, E. Y. (2017), 'The relevance to investors of greenhouse gas emission disclosures', *Contemporary Accounting Research* **34**(2), 1265–1297.

Griffin, P. A., Neururer, T. & Sun, E. Y. (2020), 'Environmental performance and analyst information processing costs', *Journal of Corporate Finance* **61**, 101397.

Hartzmark, S. M. & Sussman, A. B. (2019), 'Do investors value sustainability? A natural experiment examining ranking and fund flows', *The Journal of Finance* **74**(6), 2789–2837.

Hinkley, D. V. (1969), 'On the ratio of two correlated normal random variables', *Biometrika* **56**(3), 635–639.

Hsu, P.-H., Li, K. & Tsou, C.-Y. (2023), 'The pollution premium', *The Journal of Finance* **78**(3), 1343–1392.

Ivanov, I. T., Kruttli, M. S. & Watugala, S. W. (2024), 'Banking on carbon: Corporate lending and cap-and-trade policy', *The Review of Financial Studies* **37**(5), 1640–1684.

Kahn, M. E., Matsusaka, J. & Shu, C. (2023), 'Divestment and engagement: The effect of green investors on corporate carbon emissions', *National Bureau of Economic Research* (Working Paper 31791).

Lev, B. & Sunder, S. (1979), 'Methodological issues in the use of financial ratios', *Journal of Accounting and Economics* **1**(3), 187–210.

Li, Q., Shan, H., Tang, Y. & Yao, V. (2024), 'Corporate climate risk: Measurements and responses', *The Review of Financial Studies* **37**(6), 1778–1830.

Louis, H. (2003), 'The value relevance of the foreign translation adjustment', *The Accounting Review* **78**(4), 1027–1047.

Mascia, D. V. & Onali, E. (2024), 'Keep calm and carry on emitting: Cap-and-trade rules, local emissions and growth', *Regional Studies* **58**(1), 220–237.

Matsumura, E. M., Prakash, R. & Vera-Munoz, S. C. (2014), 'Firm-value effects of carbon emissions and carbon disclosures', *The Accounting Review* **89**(2), 695–724.

Nguyen, H. D., Tran, K. P. & Heuchenne, C. (2019), 'Monitoring the ratio of two normal variables using variable sampling interval exponentially weighted moving average control charts', *Quality and Reliability Engineering International* **35**(1), 439–460.

Petersen, M. A. (2008), 'Estimating standard errors in finance panel data sets: Comparing approaches', *The Review of Financial Studies* **22**(1), 435–480.

Plümper, T. & Troeger, V. E. (2019), 'Not so harmless after all: The fixed-effects model', *Political Analysis* **27**(1), 21–45.

Schoar, A., Yeung, K. & Zuo, L. (2024), 'The effect of managers on systematic risk', *Management Science* **70**(2), 815–833.

Seltzer, L. H., Starks, L. & Zhu, Q. (2022), 'Climate regulatory risk and corporate bonds', *National Bureau of Economic Research* (Working Paper 29994).

Shive, S. A. & Forster, M. M. (2020), 'Corporate governance and pollution externalities of public and private firms', *The Review of Financial Studies* **33**(3), 1296–1330.

Stern, N. (2008), 'The economics of climate change', *American Economic Review* **98**(2), 1–37.

Strong, N. & Walker, M. (1993), 'The explanatory power of earnings for stock returns', *The Accounting Review* **68**(2), 385–399.

Thomas, J., Yao, W., Zhang, F. & Zhu, W. (2022), 'Meet, beat, and pollute', *Review of Accounting Studies* **27**(3), 1038–1078.

Tomar, S. (2023), 'Greenhouse gas disclosure and emissions benchmarking', *Journal of Accounting Research* **61**(2), 451–492.

Upadhyay, A. & Öztekin, Ö. (2021), 'What matters more in board independence? Form or substance: Evidence from influential CEO-directors', *Journal of Corporate Finance* **71**, 102099.

Warusawitharana, M. (2008), 'Corporate asset purchases and sales: Theory and evidence', *Journal of Financial Economics* **87**(2), 471–497.

Yang, L., Muller, N. Z. & Liang, P. J. (2021), 'The real effects of mandatory CSR disclosure on emissions: Evidence from the greenhouse gas reporting program', *National Bureau of Economic Research* (Working Paper 28984).

Yim, S. (2013), 'The acquisitiveness of youth: CEO age and acquisition behavior', *Journal of Financial Economics* **108**(1), 250–273.

Zhou, Q., Faff, R. & Alpert, K. (2014), 'Bias correction in the estimation of dynamic panel models in corporate finance', *Journal of Corporate Finance* **25**, 494–513.

Table 1: Summary statistics of the main variables.

Panel A: Facility-firm-year level	Obs	Mean	Std. Dev.	Skewness	Kurtosis	P1	P50	P99
Emissions	101,035	552236.8	1724447	5.6759	41.8946	247.298	67814.35	9988362
ln(Emissions+1)	101,034	11.3745	1.9446	-.8499	9.5503	5.516	11.1246	16.1169
ln(Emissions)	100,523	11.4316	1.776	.0744	5.6024	6.9003	11.1324	16.1191
Emissions-adj	100,049	362259.3	1133673	7.0729	70.4759	.0548	56852.23	5893907
ln(Emissions-adj+1)	100,048	10.8559	2.5061	-1.7987	8.9909	.055	10.9482	15.5894
ln(Emissions-adj)	99,405	10.8955	2.5108	-2.2051	12.7803	-.3527	10.9582	15.5934
Panel B: Facility-year level	**Obs**	**Mean**	**Std. Dev.**	**Skewness**	**Kurtosis**	**P1**	**P50**	**P99**
Emissions	87,562	421278.9	1329929	6.7057	60.8588	179.496	64288.41	7287881
ln(Emissions+1)	87,561	11.2694	1.8865	-1.1097	10.6777	5.1963	11.0712	15.8017
ln(Emissions)	87,082	11.3307	1.7005	-.0831	6.13	6.7316	11.0793	15.8038
Panel C: Firm-year level	**Obs**	**Mean**	**Std. Dev.**	**Skewness**	**Kurtosis**	**P1**	**P50**	**P99**
Firm Emissions(/000)	2,962	4605.068	13603.79	5.1157	33.7462	2.836	327.745	81967.52
PPENT	2,956	11275.93	24307.07	5.4659	42.7675	51.479	3060.5	112898
Total Assets	2,962	25865.97	53078.61	4.9728	36.4377	121.471	7354.1	277787
Sales	2,961	13902.9	32886.57	6.1417	53.839	119.085	4173	166089
Common Equity (CE)	2,962	8925.456	21965.98	6.0884	48.4844	-916	2401.379	144213
Net Income	2,961	1161.533	4035.155	7.0942	91.9227	-3850	214	18680
Market Value of Equity (ME)	2,962	24226.04	70878.57	12.4682	249.7499	44.1628	5105.557	256756.5
Price/Earnings Ratio	2,961	24.8033	328.8473	13.4243	303.4703	-247.8931	15.3843	225.8612
ME/CE	2,962	1.9221	39.0807	-21.3838	877.552	-10.0039	1.9391	25.0923
Tobin's Q	2,962	1.5138	.8074	5.8711	85.7951	.6911	1.29	4.197
Firm Emissions(/000)/Total Assets	2,962	.3179	.6366	5.44	56.405	.0001	.0614	2.6838
Firm Emissions(/000)/Sales	2,952	.761	1.8577	6.2808	70.1947	.0002	.1155	7.908
Firm Emissions(/000)/PPENT	2,953	5.8026	117.2177	26.897	795.8539	.0003	.1483	7.3348
Firm Emissions(/000)/ME	2,962	269.5438	14352.63	54.3864	2959.248	.0001	.0879	15.4695

This table reports the main descriptive statistics for the main variables used in the subsequent analysis. Panel A reports the number of observations (Obs), average value (Mean), standard deviation (Std. Dev.), first percentile (P1), median (P50), $99^t h$ percentile (P99) and kurtosis (Kurtosis) for facility-level emissions, both logged and unlogged (Panel A). The data structure is based on facility-firm-year structure. Facilities with multiple owners have repeated values for each facility-year combination. Panel B reports the same data after collapsing observations at the facility-year level (without duplicate observations for each facility-year combination). Panel C reports the same statistics after the data has been collapsed at the firm-year level, including for firm-level emissions (divided by 1,000).

Table 2: Regressions on facility-firm level emissions (unadjusted): R-squared contribution of different types of FE.

	(1) State FE	(2) county FE	(3) NAICS FE	(4) County & NAICS FE	(5) Gvkey FE	(6) Firm FE	(7) Facility FE	(8) Firm & Facility FE
	Emissions	Emissions	Emissions	Emissions	Emissions	Emissions	Emissions	Emissions
Observations	101,035	97,417	100,900	97,292	31,798	99,546	100,804	99,359
R2-adj	0.0317	0.6872	0.1888	0.7035	0.3007	0.3986	0.9372	0.9435
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	No	No	No	No	No	No	No
County FE	No	Yes	No	Yes	No	No	No	No
NAICS FE	No	No	Yes	Yes	No	No	No	No
Gvkey FE	No	No	No	No	Yes	No	No	No
Firm FE	No	No	No	No	No	Yes	No	Yes
Facility FE	No	No	No	No	No	No	Yes	Yes
	ln(Emissions+1)	ln(Emissions+1)	ln(Emissions+1)	ln(Emissions+1)	ln(Emissions+1)	ln(Emissions+1)	ln(Emissions+1)	ln(Emissions+1)
Observations	101,034	97,416	100,899	97,291	31,798	99,545	100,803	99,358
R2-adj	0.0250	0.3807	0.2184	0.4573	0.2705	0.3859	0.7921	0.8052
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	No	No	No	No	No	No	No
County FE	No	Yes	No	Yes	No	No	No	No
NAICS FE	No	No	Yes	Yes	No	No	No	No
Gvkey FE	No	No	No	No	Yes	No	No	No
Firm FE	No	No	No	No	No	Yes	No	Yes
Facility FE	No	No	No	No	No	No	Yes	Yes
	ln(Emissions)	ln(Emissions)	ln(Emissions)	ln(Emissions)	ln(Emissions)	ln(Emissions)	ln(Emissions)	ln(Emissions)
Observations	100,523	96,906	100,390	96,783	31,592	99,039	100,280	98,840
R2-adj	0.0299	0.4435	0.2794	0.5369	0.3311	0.4505	0.8691	0.8796
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	No	No	No	No	No	No	No
County FE	No	Yes	No	Yes	No	No	No	No
NAICS FE	No	No	Yes	Yes	No	No	No	No
Gvkey FE	No	No	No	No	Yes	No	No	No
Firm FE	No	No	No	No	No	Yes	No	Yes
Facility FE	No	No	No	No	No	No	Yes	Yes

This table reports the results of FE panel regressions on facility-level emissions. Panel A reports the *R2-adj* resulting from regressions of models that employ Total Direct Emissions as a dependent variable, while Panel B and C use $ln(Emissions + 1)$ and $ln(Emissions)$ respectively. For each panel, the FE included are: State FE (Column 1); County FE (Column 2); NAICS FE (Column 3); County and NAICS FE (Column 4); Gvkey FE (Column 5); Firm FE (Column 6); Facility FE (Column 7); Firm and Facility FE (Column 8). All specifications include Year FE, by default.

Table 3: Regressions on Tobin's Q allowing for county and industry FE.

	(1) Tobin	(2) Tobin	(3) Tobin	(4) Tobin	(5) Tobin	(6) Tobin	(7) Tobin	(8) Tobin	(9) Tobin	(10) Tobin	(11) Tobin	(12) Tobin
Emissions (ln)	-0.0198***	-0.0231**										
	(-2.8707)	(-2.4512)										
Emissions			-0.0000***	-0.0000								
			(-2.8461)	(-1.1784)								
Emissions/Assets					-0.0369	-0.1048***						
					(-1.5072)	(-3.1061)						
Emissions/Sales							-0.0151***	-0.0210**				
							(-4.4361)	(-2.2715)				
Emissions/PPENT									-0.0140	-0.0480***		
									(-1.1847)	(-2.9814)		
Emissions/ME											-0.0132***	-0.0160**
											(-4.0414)	(-2.4589)
Size		-0.0018		-0.0146		-0.0278**		-0.0241**		-0.0279**		-0.0265**
		(-0.1072)		(-0.9109)		(-2.3488)		(-2.0530)		(-2.3314)		(-2.2590)
Total debt ratio		0.4302***		0.4157***		0.4505***		0.4414***		0.4226***		0.4587***
		(3.7425)		(3.6387)		(3.8635)		(3.7907)		(3.7704)		(3.8738)
ROA		1.6706***		1.6792***		1.6614***		1.6873***		1.6541***		1.5988***
		(7.6589)		(7.6959)		(7.9744)		(7.6205)		(7.8710)		(7.7706)
Sales ratio		0.0666**		0.0579*		0.0638**		0.0446		0.0643**		0.0637**
		(2.1530)		(1.8072)		(2.0015)		(1.4319)		(2.0356)		(2.0446)
Cash ratio		0.7893***		0.8000***		0.7885***		0.8229***		0.8283***		0.8139***
		(4.3208)		(4.3972)		(4.0993)		(4.4219)		(4.4463)		(4.2832)
Firm age (ln)		0.0640***		0.0652***		0.0656***		0.0656***		0.0641***		0.0711***
		(2.7967)		(2.8312)		(2.8620)		(2.8742)		(2.8462)		(3.0347)
Multi-owners		-0.0363		-0.0256		-0.0345		-0.0339		-0.0293		-0.0318
		(-1.3842)		(-0.9045)		(-1.3007)		(-1.2544)		(-1.1118)		(-1.2030)
Constant	0.4518***	0.0587	0.3046***	0.0207	0.3030***	0.1648	0.3051***	0.1150	0.2970***	0.1715	0.3021***	0.1046
	(7.7631)	(0.3452)	(44.3548)	(0.1033)	(24.3542)	(1.0722)	(69.9895)	(0.7464)	(27.6499)	(1.0982)	(70.6045)	(0.6443)
Observations	26,160	17,593	26,169	17,602	26,169	17,602	26,155	17,593	26,151	17,584	26,169	17,602
R2-adj	0.6516	0.7108	0.6497	0.7095	0.6494	0.7141	0.6501	0.7105	0.6492	0.7136	0.6550	0.7146
Counties	1526	1241	1526	1241	1526	1241	1525	1241	1526	1241	1526	1241
Industries	193	190	193	190	193	190	193	190	191	188	193	190
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	No	No	No	No	No	No	No	No	No	No
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
NAICS FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	CN	CN	CN	CN	CN	CN	CN	CN	CN	CN	CN	CN

This table reports the results of FE panel regressions of emissions on firm valuation. The dependent variable is the natural logarithm of Tobin's Q – defined as the market value of assets over the book value of assets. Key regressors are: the natural logarithm of emissions (columns (1)–(2)), unscaled emissions (columns (3)–(4)), emissions scaled by total assets (columns (5)–(6)), emissions scaled by total sales (columns (7)–(8)), emissions scaled by Total Property, Plant and Equipment (PPENT) (columns (9)–(10)), and emissions scaled by the Market value of Equity (ME) (columns (11)–(12)). Standard firm-level controls are added to specifications reported in even columns, and include Size (the natural logarithm of total assets), Total debt ratio (total debt over total assets), ROA (net income over total assets), Sales ratio (sales over total assets), Cash ratio (cash and short term investments over total assets), Firm age (the natural logarithm of the difference between the current year and the first year a firm was recorded in Compustat), and Multi-owners (a dummy equal to one when facilities are owned by more than one owner). All specifications include county, NAICS and year FE. *t*-statistics are reported in parentheses. Standard errors are clustered at the County-NAICS level. ***, **, * denote statistical significance at the 1%, 5%, and 10% level respectively.

Table 4: Regressions on Tobin's Q allowing for firm, county and industry FE.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Tobin	Tobin	Tobin	Tobin	Tobin	Tobin	Tobin	Tobin	Tobin	Tobin	Tobin	Tobin
Emissions (ln)	-0.0503***	-0.0307										
	(-2.8624)	(-1.3518)										
Emissions			-0.0000**	-0.0000								
			(-2.2124)	(-1.4381)								
Emissions/Assets					-0.0356	-0.2550**						
					(-0.4545)	(-2.2228)						
Emissions/Sales							-0.0159	-0.0238				
							(-0.9394)	(-0.5318)				
Emissions/PPENT									0.0032	-0.1008*		
									(0.0841)	(-1.9270)		
Emissions/ME											-0.0156**	-0.0168*
											(-2.5395)	(-1.7179)
Size		-0.1331***		-0.1367***		-0.1833***		-0.1542***		-0.1737***		-0.1632***
		(-3.1714)		(-3.2980)		(-7.2111)		(-4.8166)		(-6.4641)		(-5.1424)
Total debt ratio		0.1992		0.2256*		0.1953		0.2064		0.1793		0.2438*
		(1.5291)		(1.7404)		(1.5045)		(1.5823)		(1.3548)		(1.8042)
ROA		0.9271***		0.9586***		0.8946***		0.9483***		0.9168***		0.8798***
		(6.3203)		(6.2717)		(5.8930)		(6.2329)		(6.1740)		(6.5004)
Sales ratio		0.1534***		0.1439***		0.1539***		0.1391***		0.1565***		0.1674***
		(2.9846)		(3.1523)		(3.0934)		(3.4436)		(3.0908)		(3.7202)
Cash ratio		0.6416***		0.6164***		0.5864***		0.6560***		0.6736***		0.6355***
		(3.7476)		(3.4015)		(2.9181)		(3.7585)		(3.8345)		(3.6855)
Firm age (ln)		-0.0827		-0.0691		-0.0505		-0.0945*		-0.0636		-0.0851
		(-1.6092)		(-1.3625)		(-1.0509)		(-1.7024)		(-1.2805)		(-1.5069)
Multi-owners		-0.0132*		-0.0084		-0.0135*		-0.0128		-0.0127*		-0.0123
		(-1.7703)		(-1.2331)		(-1.8650)		(-1.6346)		(-1.7427)		(-1.6504)
Constant	0.7111***	1.9664***	0.3572***	1.7588***	0.3023***	2.2242***	0.3063***	1.9930***	0.2811***	2.1444***	0.3054***	2.0215***
	(4.7691)	(8.6635)	(10.8019)	(5.5773)	(7.4976)	(9.4025)	(12.9928)	(7.5247)	(7.9320)	(9.0238)	(36.2993)	(8.2454)
Observations	26,141	17,577	26,150	17,586	26,150	17,586	26,136	17,577	26,132	17,568	26,150	17,586
R2-adj	0.8343	0.8585	0.8345	0.8594	0.8307	0.8641	0.8314	0.8580	0.8304	0.8622	0.8349	0.8608
Counties	1525	1240	1525	1240	1525	1240	1524	1240	1525	1240	1525	1240
Industries	193	190	193	190	193	190	193	190	191	188	193	190
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
NAICS FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	CN	CN	CN	CN	CN	CN	CN	CN	CN	CN	CN	CN

This table reports the results of FE panel regressions of emissions on firm valuation. The dependent variable is the natural logarithm of Tobin's Q – defined as the market value of assets over the book value of assets. Key regressors are: the natural logarithm of emissions (columns (1)–(2)), unscaled emissions (columns (3)–(4)), emissions scaled by total assets (columns (5)–(6)), emissions scaled by total sales (columns (7)–(8)), emissions scaled by Total Property, Plant and Equipment (PPENT) (columns (9)–(10)), and emissions scaled by the Market value of Equity (ME) (columns (11)–(12)). Standard firm-level controls are added to specifications reported in even columns, and include Size (the natural logarithm of total assets), Total debt ratio (total debt over total assets), ROA (net income over total assets), Sales ratio (sales over total assets), Cash ratio (cash and short term investments over total assets), Firm age (the natural logarithm of the difference between the current year and the first year a firm was recorded in Compustat), and Multi-owners (a dummy equal to one when facilities are owned by more than one owner). All specifications include firm, county, NAICS and year FE. *t*-statistics are reported in parentheses. Standard errors are clustered at the County-NAICS level. ***, **, * denote statistical significance at the 1%, 5%, and 10% level respectively.

Table 5: Monte Carlo simulations at the facility level: impact of an AR(1) component in Emissions and Capital. True $\rho_k = \rho_g = 0.92$

Panel A: Regressions on facility-level Emissions	Obs	Mean	Std. Dev.	Skewness	Kurtosis	P1	P10	P50	P90	P99
$\hat{\rho_g}$	1,000	.9197	.0045	.282	3.1988	.9104	.9139	.9195	.9257	.9312
R2-adj Facility FE and Year FE	1,000	.6793	.0055	-.0478	3.034	.6658	.6721	.6794	.6865	.6917
R2-adj Firm FE, County FE and Year FE	1,000	.3345	.0079	.0786	3.0092	.3161	.3242	.3345	.3449	.3536
R2-adj Firm FE and Year	1,000	.3346	.0079	.0758	3.025	.3158	.3242	.3345	.345	.3537
R2-adj County FE and Year FE	1,000	.1809	.0041	-.1245	3.3425	.171	.1756	.181	.1859	.1898
Panel B: Regressions on facility-level Emissions/Capital	**Obs**	**Mean**	**Std. Dev.**	**Skewness**	**Kurtosis**	**P1**	**P10**	**P50**	**P90**	**P99**
R2-adj Facility FE and Year FE	1,000	.0644	.0189	-.0655	2.8938	.0213	.0389	.0645	.0875	.1104
R2-adj Firm FE, County FE and Year FE	1,000	.0189	.0054	-.1207	3.1514	.0062	.0118	.0192	.0256	.0313
R2-adj Firm FE and Year	1,000	.0189	.0054	-.1211	3.1283	.0062	.0118	.0192	.0257	.0312
R2-adj County FE and Year FE	1,000	.0121	.0034	-.2458	3.0953	.0037	.0076	.0123	.0162	.0193
Panel C: Regressions on facility-level Emissions	**Obs**	**Mean**	**Std. Dev.**	**Skewness**	**Kurtosis**	**P1**	**P10**	**P50**	**P90**	**P99**
R2-adj Facility FE and Year FE (with lag)	1,000	.8407	.0034	.1016	2.7877	.8334	.8361	.8406	.8452	.8485
R2-adj Facility FE and Year FE (without lag)	1,000	.6795	.0053	.1444	3.1941	.6678	.6728	.6795	.6859	.6923
R2-adj Firm FE, County FE and Year FE (with lag)	1,000	.8172	.0035	.0825	2.8139	.8096	.8125	.8171	.8219	.8254
R2-adj Firm FE, County FE and Year FE (without lag)	1,000	.1809	.0043	-.0342	2.9999	.1713	.1755	.181	.1865	.1907
R2-adj Firm FE and Year FE (with lag)	1,000	.8172	.0035	.0833	2.8105	.8096	.8126	.8171	.8218	.8254
R2-adj Firm FE and Year FE (without lag)	1,000	.1809	.0043	-.0426	3.0375	.1713	.1756	.1809	.1864	.1903
R2-adj County FE and Year FE (with lag)	1,000	.8172	.0035	.0874	2.8321	.8096	.8126	.8171	.8218	.8255
R2-adj County FE and Year FE (without lag)	1,000	.1809	.0043	-.0421	2.9567	.1711	.1755	.1809	.1864	.1908

This table reports the results of Monte Carlo simulations based on 1,000 simulated samples. For each of the 1,000 replications, we run an OLS regression where the dependent variable is either Emissions (Panel A) or Emissions/Capital (Panel B), and the independent variables are different types of FE. We then collect the adjusted R-squared (*R2-adj*) for each regression and compute the mean, standard deviation and percentiles of the distribution of *R2-adj* for the whole set of 1,000 replications. The Data Generating Process (DGP) is described in Section B and Section 3.1. The first line of Panel A reports the estimated autocorrelation coefficient in emissions, $\hat{\rho_g}$, according too the dynamic panel data estimator by Arellano & Bover (1995) and Blundell & Bond (1998). In both Panel A and B, firm FE are correlated with the average facility-level capital stock, k_{ifct}. In Panel C, neither firm FE nor county FE are correlated with the average facility-level capital stock, k_{ifct} – as explained in Section 3.1.1. Furthermore, in Panel C we report the results of regressions including–"(with lag)"–and escluding–"(without lag)"–the lag of g_{ifct}. The true value of ρ_g is 0.92.

Table 6: Monte Carlo simulations at the facility level: impact of a large time-invariant firm-specific component in Emissions and Capital. True $\rho_k = \rho_g = 0$

Panel A: Regressions on facility-level Emissions	Obs	Mean	Std. Dev.	Skewness	Kurtosis	P1	P10	P50	P90	P99
$\hat{\rho_g}$	1000	.0002	.004	.0112	2.833	-.0084	-.0049	.0004	.0053	.0094
R2-adj Facility FE and Year FE (with lag)	1000	.0123	.0018	.0023	2.8161	.0081	.01	.0122	.0147	.0164
R2-adj Firm FE, County FE and Year FE (with lag)	1000	.49	.0143	-.0506	2.8456	.4554	.4718	.4904	.5085	.522
R2-adj Firm FE and Year (with lag)	1000	.49	.0143	-.0468	2.8559	.4556	.472	.4904	.5085	.522
R2-adj County FE and Year FE (with lag)	1000	0	.0005	.0271	2.7955	-.0012	-.0007	0	.0006	.0011
Panel B: Regressions on facility-level Emissions	**Obs**	**Mean**	**Std. Dev.**	**Skewness**	**Kurtosis**	**P1**	**P10**	**P50**	**P90**	**P99**
R2-adj Facility FE and Year FE (without lag)	1000	0	.0015	-.0358	2.9518	-.0038	-.002	0	.0019	.0035
R2-adj Firm FE, County FE and Year FE (without lag)	1000	.4899	.014	-.0315	2.8634	.4564	.4718	.4901	.5079	.5209
R2-adj Firm FE and Year (without lag)	1000	.4899	.014	-.0277	2.871	.4565	.4718	.4901	.5079	.5211
R2-adj County FE and Year FE (without lag)	1000	0	.0004	-.0265	2.9067	-.001	-.0006	0	.0006	.001
Panel C: Regressions on facility-level Emissions/Capital	**Obs**	**Mean**	**Std. Dev.**	**Skewness**	**Kurtosis**	**P1**	**P10**	**P50**	**P90**	**P99**
R2-adj Facility FE and Year FE (without lag)	1000	0	.0015	-.01	3.1216	-.0037	-.0019	0	.0019	.0035
R2-adj Firm FE, County FE and Year FE (without lag)	1000	.6664	.0126	-.0505	3.1522	.6349	.65	.6665	.6819	.6965
R2-adj Firm FE and Year (without lag)	1000	.6664	.0126	-.051	3.1477	.6349	.65	.6666	.6819	.6967
R2-adj County FE and Year FE (without lag)	1000	0	.0004	-.0274	2.7612	-.001	-.0006	0	.0006	.001

This table reports the results of Monte Carlo simulations based on 1,000 simulated samples. For each of the 1,000 replications, we run an OLS regression where the dependent variable is either Emissions (Panels A and B) or Emissions/Capital (Panel C), and the independent variables are different types of FE. We then collect the adjusted R-squared (*R2-adj*) for each regression and compute the mean, standard deviation and percentiles of the distribution of *R2-adj* for the whole set of 1,000 replications. The Data Generating Process (DGP) is described in Section B and Section 3.1. The first line of Panel A reports the estimated autocorrelation coefficient in emissions, $\hat{\rho_g}$, according too the dynamic panel data estimator by Arellano & Bover (1995) and Blundell & Bond (1998). Neither firm FE nor county FE are correlated with the average facility-level capital stock, k_{ifct}. In Panel A, we report the results of regressions including the lag of g_{ifct} ("with lag"), while in Panels B and B we exclude it("without lag"). The true value of ρ_g is 0.

Table 7: Effect of scaling by a proxy for firm size: Monte Carlo simulations at the firm level assuming a negative correlation between Firm Emissions and Firm value (true value of $\psi_G = -0.0004$).

Panel A: Summary statistics	Obs	Monte Carlo Mean	Sample (Table 1)	Sample winsorized				
G_{ft} - Median	1,000	524.8139	327.745	327.745				
G_{ft} - Kurtosis	1,000	8.5615	33.7462	27.7048				
K_{ft} - Median	1,000	3512.009	3060.5	3060.5				
K_{ft} - Kurtosis	1,000	11.0309	42.7675	41.9282				
G_{ft}/K_{ft} - Median	1,000	.1198	.1483	.1483				
G_{ft}/K_{ft} - Kurtosis	1,000	258.6443	795.8539	20.2895				
V_{ft} - Median	1,000	5477.968	5105.557	5105.557				
V_{ft} - Kurtosis	1,000	14.9032	249.7499	25.4312				
Panel B: equation (2)	Obs	Mean	Std. Dev.	Skewness	Kurtosis	P1	P50	P99
Coefficient on G_{ft} (β_2)	1,000	-.0004	.0001	.0098	2.9919	-.0006	-.0004	-.0002
T-statistic for G_{ft}	1,000	-4.0144	1.1138	-.0031	3.0295	-6.5684	-4.028	-1.2885
Panel C: equation (3)	Obs	Mean	Std. Dev.	Skewness	Kurtosis	P1	P50	P99
Coefficient on G_{ft}/K_{ft} (β_2)	1,000	-.0296	.0415	.7411	6.4712	-.1206	-.0309	.0784
T-statistic for G_{ft}/K_{ft}	1,000	-.9889	1.1577	-.1631	2.9991	-3.7505	-.9641	1.4955
Panel D: equation (4)	Obs	Mean	Std. Dev.	Skewness	Kurtosis	P1	P50	P99
Coefficient on G_{ft}/K_{ft} (β_2)	1,000	.0032	.0021	1.9796	13.1111	-.0007	.0029	.0105
T-statistic for G_{ft}/K_{ft}	1,000	2.3417	1.1874	1.1791	11.5083	-.3736	2.2577	5.3511
Panel E: equation (5)	Obs	Mean	Std. Dev.	Skewness	Kurtosis	P1	P50	P99
Coefficient on G_{ft}/K_{ft} (β_2)	1,000	-.0002	.0022	1.3212	14.102	-.0051	-.0003	.0056
T-statistic for G_{ft}/K_{ft}	1,000	-.3563	1.2641	-.2084	3.1801	-3.4841	-.2885	2.4378

This table reports the results of OLS regressions on 1,000 simulated samples. In Panel A, we report the descriptive statistics for the average value in each of the 1,000 simulated samples for the following variables: Firm Emissions (G_{ft}), Firm Capital (K_{ft}), Firm Emissions scaled by Firm Capital (G_{ft}/K_{ft}), and Firm Value (V_{ft}). For the same variables, we report the corresponding statistic in our sample, without any adjustment, and with a winsorization at the 1^{st} and 99^{th} percentile. In Panel B, we report the distribution of the estimated coefficients and t-statistics for the coefficient on G_{ft} (β_2) in regressions run according to equation (2). In Panels C, D and E we report the distribution of the estimated coefficients and t-statistics for G_{ft}/K_{ft} (β_2) in regressions run according to equations (3), (4), and (5), respectively. The Data Generating Process (DGP) is based on the same 1,000 simulated samples used for Table 5, but the data is aggregated at the firm-year level, although we start from the datasets with facility-firm-county-year structure. Further details of the DGP are described in Section B. All regressions have firm FE and year FE and standard errors are clustered at the firm level.

Table 8: Regressions on facility level emissions: Assessing the impact of the CATP.

	(1) Dyn	(2) FE no lag	(3) FE lag	(4) Lag no facility FE	(5) Dyn	(6) FE no lag	(7) FE lag	(8) Lag no facility FE
$g_{i(t-1)}$	0.940***		0.699***	0.949***	0.967***		0.699***	0.949***
	(75.634)		(24.418)	(261.555)	(86.476)		(24.421)	(261.493)
Treated × Post	-20.739***	48.807***	4.885	-12.830***	-963.835***	46.158***	1.539	-17.516***
	(-3.755)	(4.920)	(1.051)	(-3.171)	(-3.306)	(4.394)	(0.297)	(-3.850)
Change in Parent Firm					6,321.289**	-5.372	-6.120	-5.685
					(2.494)	(-0.413)	(-0.902)	(-0.923)
Change in Parent Firm × Post					-6,112.616**	2.054	5.240	6.133
					(-2.415)	(0.146)	(0.690)	(0.854)
Change in Parent Firm × Treated					-7,710.496***	-14.850	-20.261	-27.037**
					(-3.351)	(-0.821)	(-1.574)	(-2.440)
Treated × Post × Change in Parent Firm					7,509.696***	20.793	25.834*	32.368**
					(3.281)	(1.014)	(1.676)	(2.386)
Treated	13.798**			8.931**	1,018.375***			12.996***
	(2.330)			(2.357)	(3.368)			(3.068)
Observations	77,128	77,128	77,128	77,128	77,128	77,128	77,128	77,128
R-squared		0.9359	0.9685	0.9604		0.9359	0.9685	0.9605
Number of Facilities	8,333	8,333	8,333	8,333	8,333	8,333	8,333	8,333
Facility FE	No	Yes	Yes	No	No	Yes	Yes	No
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	No	No	No	No	No
Cluster	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm

This table reports the results of regressions run according to the estimator by Arellano & Bover (1995) and Blundell & Bond (1998), System-GMM (columns 1 and 5), and OLS regressions (columns 2–4 and 6–8). The dependent variable is facility-level emissions. The independent variables include: the lag of the dependent variable (apart from columns (2) and (6)); Post (a dummy variable equal to zero for the period 2010–12, and one for the period 2013–22); Treated (a dummy variable identifying counties in California, and 0 otherwise). For columns (5)–(8) we also include: "Change in Parent Firm" a dummy variable equal to one if the majority owner of the facility changes, and its interactions with Treated, Post, and Treated × Post. Constant included but not reported. Columns 2–3 and 6–7 include Facility FE. All specifications include Year FE.

Table 9: Toolbox.

Econometric issue	**Approaches used in the literature**	**Potential issues**	**Potential solutions**
1. Size effects	Scaling by total assets, EBIT, sales, market value of equity, COGS	a. Conflating changes in the ratio due to changes in denominator instead of numerator. b. Large Type II error and/or bias. c. Potential omitted variable bias due to autocorrelation in both dependent and independent variables.	a. Avoid scaling without underpinning theory. b. Use bootstrap to allow for non-Normal t-statistics. c. Robustness checks including/excluding lag of dependent variable on slope coefficients and standard errors.
2. Nested error term structure	Firm FE, facility FE, industry FE, year FE, county FE	Omitting information in facility-level data. E.g., firm-level (regional-level) aggregation neglects regional (firm-level) factors.	a. Avoid aggregating facility-level data at the firm-level. b. Use both firm-level, industry-level and regional-level FE and controls.
3. Functional form	Logged variables	Lower statistical power	a. Justify choice of functional form from a theoretical perspective. b. For log(x+1), see recommendations from Chen & Roth (2024).

This table reports summary of the main econometric issues explored in this paper (column (1)), the approaches used in the literature to address these issues (column (2)), the potential problems inherent in these approaches (column (3)), and potential solutions based on our findings and other relevant studies (column (4)).

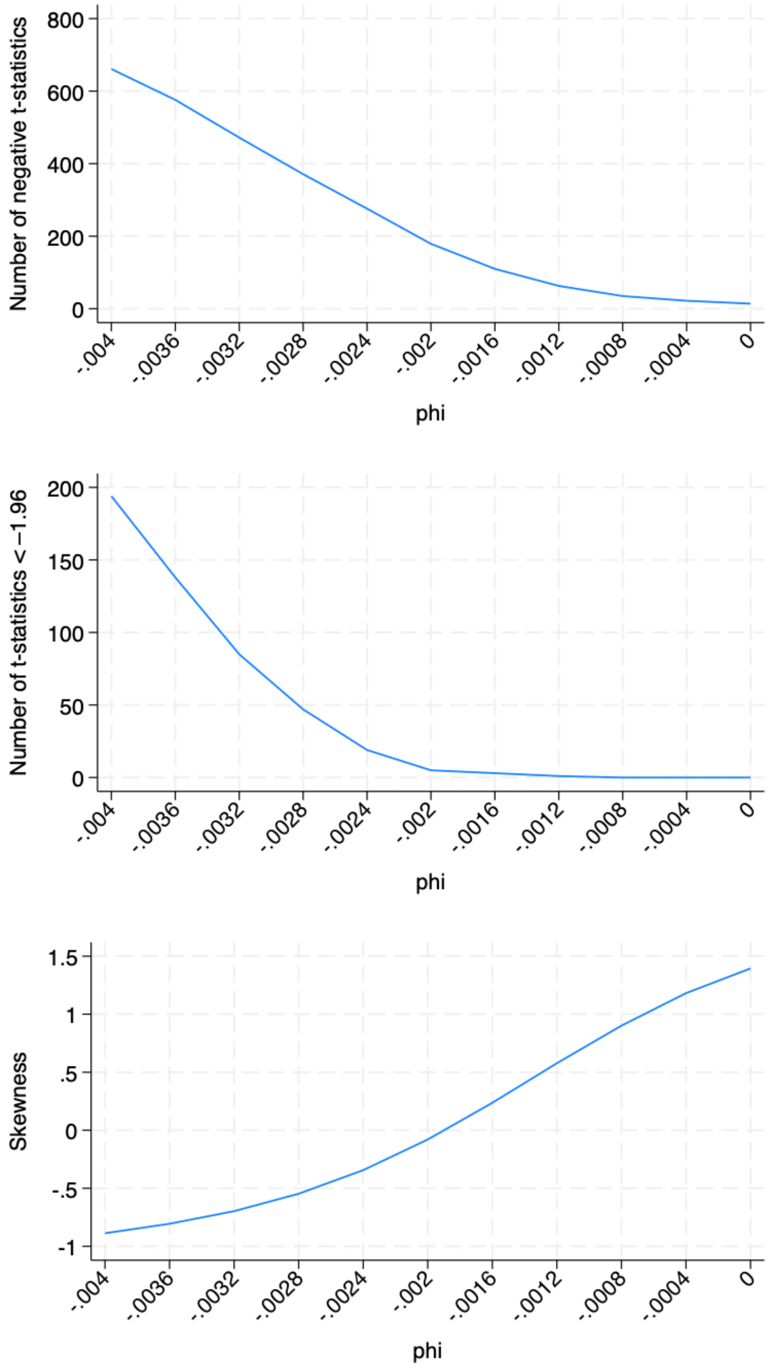

Figure 1: Effect of true value of ϕ_G in equation (1) (phi) on the distribution of the t-statistics for the coefficient on Emissions/Firm Capital (G_{ft}/K_{ft}) (β_2) in regressions run according to equation (4). Top panel: number of negative t-statistics (y axis) and phi (x axis). Center panel: number of t-statistics smaller than -1.96 (y axis) and phi (x axis). Bottom panel: skewness of the distribution of the t-statistics and (y axis) and phi (x axis).

APPENDIX

Modeling facility-level emissions data in corporate finance:

A cautionary tale

A Data

Our main database is the GreenHouse Gas Reporting Program (GHGRP),[1] which is administered by the Environmental Protection Agency (EPA).[2] For each facility subject to the GHGRP, we retrieve annual facility-level Scope 1 greenhouse gas emissions (i.e., GHG emissions directly produced by the facilities), as well as information regarding each facility's parent companies and related percent ownership.[3,4]

Facility-level ownership structure is essential to our study because it allows us to control for parent-level FE of all owners and not just those who control the facility. Since our study centers on the strategic considerations regarding emissions and, in particular, whether parent companies tend to change emissions at each facility or whether they buy/sell facilities, focusing on the controlling owner might disregard important portfolio-level dynamics.

The GHGRP is therefore superior to databases that report only the largest parent company. For instance, although its focus is not GHG emissions, the Toxics Release Inventory (TRI) Program – compiled by the same EPA – only provides the highest-level company with the largest ownership interest in the facility, when such facility is owned by more than one company.[5] Similarly, another very popular database whose unit of analysis

[1]The GHGRP reporting initiative excludes emissions stemming from *i*) agricultural activities; *ii*) direct emission sources generating less than 25,000 metric tons of CO_2 equivalent (CO_2e, the volume of CO_2 emissions that generate an equivalent global warming impact as a specified quantity of metric tons of another GHG) annually, unless obligated to report regardless of their yearly emissions; *iii*) sinks of GHGs; *iv*) data reporting on electricity procurement or indirect emissions from energy usage, categorized as Scope 2 emissions (https://www.epa.gov/ghgreporting/learn-about-greenhouse-gas-reporting-program-ghgrp). It is important to highlight that there are instances (about 16% of the direct emissions reported between 2010–2022) in which facilities reported less than 25,000 metric tons of CO_2e. This scenario is common for municipal solid waste landfills (https://ccdsupport.com/confluence/pages/viewpage.action?pageId=189038672). See Yang et al. (2021) for an extensive description of the database.

[2]https://www.epa.gov/ghgreporting/archive-ghg-reporting-program-data-sets.

[3]https://www.epa.gov/ghgreporting/data-sets.

[4]According to the EPA, total reported emissions from these facilities are about 3 billion metric tons CO_2e, which accounts for about 50 percent of total U.S. GHG emissions. https://www.epa.gov/ghgreporting/learn-about-greenhouse-gas-reporting-program-ghgrp.

[5]See Section 5.2, (3), at: https://www.epa.gov/toxics-release-inventory-tri-program.

is the facility/establishment – namely the National Establishment Time-Series (NETS) Database – only provides information on the ultimate parent company. Furthermore, due to the continuous verification process they go through, data from the GHGRP are considered the most reliable GHG emissions figures (Kahn et al. 2023).

B Data Generating Processes

We employ Monte Carlo simulations to investigate how facility-level persistence in emissions affects the explanatory power of models with different types of FE. Then, we run firm-level simulations to investigate how scaling firm emissions using a proxy for firm size might affect the estimated coefficient of regressions where firm value is the dependent variable and firm emissions are one of the explanatory variables.

Facility-level simulations. We simulate 100,000 observations, similar to our sample in Panel A of Table 1, for 10,000 fictitious facilities and 10 fictitious time units (e.g., years). Each facility is allocated to 1,000 fictitious counties and 1,000 fictitious firms. Each county and firm have 100 facilities and there is no correlation between county-level and firm-level clusters. Each facility is allocated capital stock at time t, denoted k_{ifct}, where i denotes the facility, f the firm, c the county where the facility is located.

$$k_{ifct} = \varepsilon_{ifct} + \rho_k k_{ifc(t-1)} + \gamma_t + \gamma_f + \gamma_c + \Gamma_{ifct} \tag{A1}$$

where ε_{ifct}, is a standard lognormal independent and identically distributed (ln-iid) random variable – $ln(\varepsilon_{ifct}) \sim N(0, 1)$. This random variable captures the impact of omitted time-varying factors. We assume that facility-level capital stock is autocorrelated, with ρ_k the autocorrelation coefficient of an autoregressive process of order one–AR(1). To disentangle the impact of the AR(1) component from that of time-invariant omitted variables, we also include time-specific FE, γ_t, firm-specific FE, γ_f, and county-specific FE, γ_c. These FE follow a uniform distribution over the interval (0,1), and they are orthogonal to each other.

In our simulations, we model the size of these time-, firm-, and county-specific components to simulate a certain weight on the overall explanatory power of the model. We report the details of the weighting in Section 3.1.

Finally, Γ_{ifct} is a Gamma-distributed iid variable with shape parameter ($k_{\Gamma 1}$ and scale parameter $k_{\Gamma 2}$). We choose the parameter values ($k_{\Gamma 1}$ and $k_{\Gamma 2}$) such that the distribution of k_{ifct} is positively skewed and leptokurtic, to mimic the distribution of the real data for firm capital stock in our sample.[6] Although we do not have data for facility-level capital stock, we can simulate this behavior at the firm-level by choosing a shape and scale parameter leading to skewed and leptokurtic values of k_{ifct}. Following Erickson & Whited (2012), to generate the process described in A1, we first generate the first observation, without the first lag of the dependent variable. Then, we update the data for the following periods by incorporating the with AR(1) component for observations for period two to period 10.

We follow a similar process for facility-level emissions. Denote g_{ifct} be emissions in facility i of firm f located in county c in year t. We use the following equation to simulate g_{ifct}:

$$g_{ifct} = \varepsilon_{ifct} + \rho_g g_{ifc(t-1)} + \beta_k k_{ifct} + \lambda_t + \lambda_f + \lambda_c + \Gamma_{ifct} \tag{A2}$$

where ε_{ifct} is ln-iid, as before, Γ_{ifct} is a Gamma-distributed iid variable with shape

[6] There is no closed form solution for the median of a Gamma distribution. Using both a lognormal and a Gamma-distributed component enables us to obtain values of the median and kurtosis that are closer to those of the actual distribution in our sample than if we were to use only a lognormal or only a Gamma distribution.

parameter $g_{\Gamma 1}$ and scale parameter $g_{\Gamma 2}$.

In our simulations, we have used $\rho_g = 0.92$[7] and $\beta_k = 0.01$ in equation (A2). Since we do not have facility-level data for PPENT, we set the value of ρ_k at the same value as ρ_g in equation (A1). This is similar to what found in previous research, which suggests an autocorrelation coefficient close to one for capital stock (e.g., Cogley & Nason (1995)). We have chosen the values of $g_{\Gamma 1}$ and $g_{\Gamma 2}$ such that the distribution of facility-level emissions, has a median similar to the median reported for Firm Emissions-adj (at the facility-firm-year level) in Table 1 Panel A. We have considered the median, rather than the mean, because often researchers winsorize either emissions or Compustat variables (or both). Since the median is invariant under winsorization, unlike the mean, by matching the median we can replicate the behavior of both raw and winsorized variables.

Since we do not have facility-level data on PPENT, we have chosen values of ($k_{\Gamma 1}$ and $k_{\Gamma 2}$) generating a median value of the firm-level value for capital stock, $K_{ft} = \sum_i k_{ifct}$, similar to that reported for PPENT in Table 1. These values are: $k_{\Gamma 1} = 0.1$, $k_{\Gamma 2} = 1500$, $g_{\Gamma 1} = 0.05$ and $g_{\Gamma 2} = 150$.

Firm-level simulations. We start from the same 1,000 simulated samples at the facility level used for above. As said above, K_{ft} (Firm Capital) is the sum of facility-level capital stock at time t for each facility, denoted k_{ifct}. Similarly, Firm Emissions, G_{ft}, are the sum of facility-level emissions for each firm f, and Firm Emissions/Firm Capital is $G_{ft}/K_{ft} = \sum_i \frac{g_{ifct}}{k_{ifct}}$.

At this point, we collapse all the data at the firm-time level, and construct additional firm-level variables. For simplicity, we assume that production costs are linearly correlated with Firm Capital in the previous period (without a stochastic component): $\beta_K K_{ft}$.

[7]The autocorrelation coefficient in our sample for facility-level emissions is 0.96, while the one for firm-level emissions is 0.92.

However, firm sales are stochastic and, as for capital stock and emissions, include a lognormal and a gamma-distributed component, and Firm Profit is:

$$\Pi_{ft} = \varepsilon_{ft} + \Gamma_{ft} + \beta_K K_{ft} \tag{A3}$$

where ε_{ft} is a random Normally-distributed iid (Niid) variable, at the firm-year level, Γ_{ft} is a Gamma-distributed firm-level iid variable with shape parameter $\Pi_{\Gamma 1}$ and scale parameter $\Pi_{\Gamma 2}$. In equation (A3), we set: $\Pi_{\Gamma 1} = 0.5$, $\Pi_{\Gamma 2} = 1200$, $\beta_K = 0.0001$. Thus, the size of profits depend on capital stock and a random error term.

Finally, we compute Firm Value, V_{ft}, assuming a Price/Earnings ratio of 20 and a negative correlation between Firm Emissions and Firm Value:

$$V_{ft} = \varepsilon_{ft} + \rho_V V_{i(t-1)} + 20\Pi_{ft} + \psi_G G_{ft} + \mu_t + \mu_f \tag{A4}$$

where μ_t simulates the effect of time-varying unobserved factors that might affect all firms and μ_f simulates the impact of time-invariant firm-specific unobserved factors.

ONLINE APPENDIX

Modeling facility-level emissions data in corporate finance:

A cautionary tale

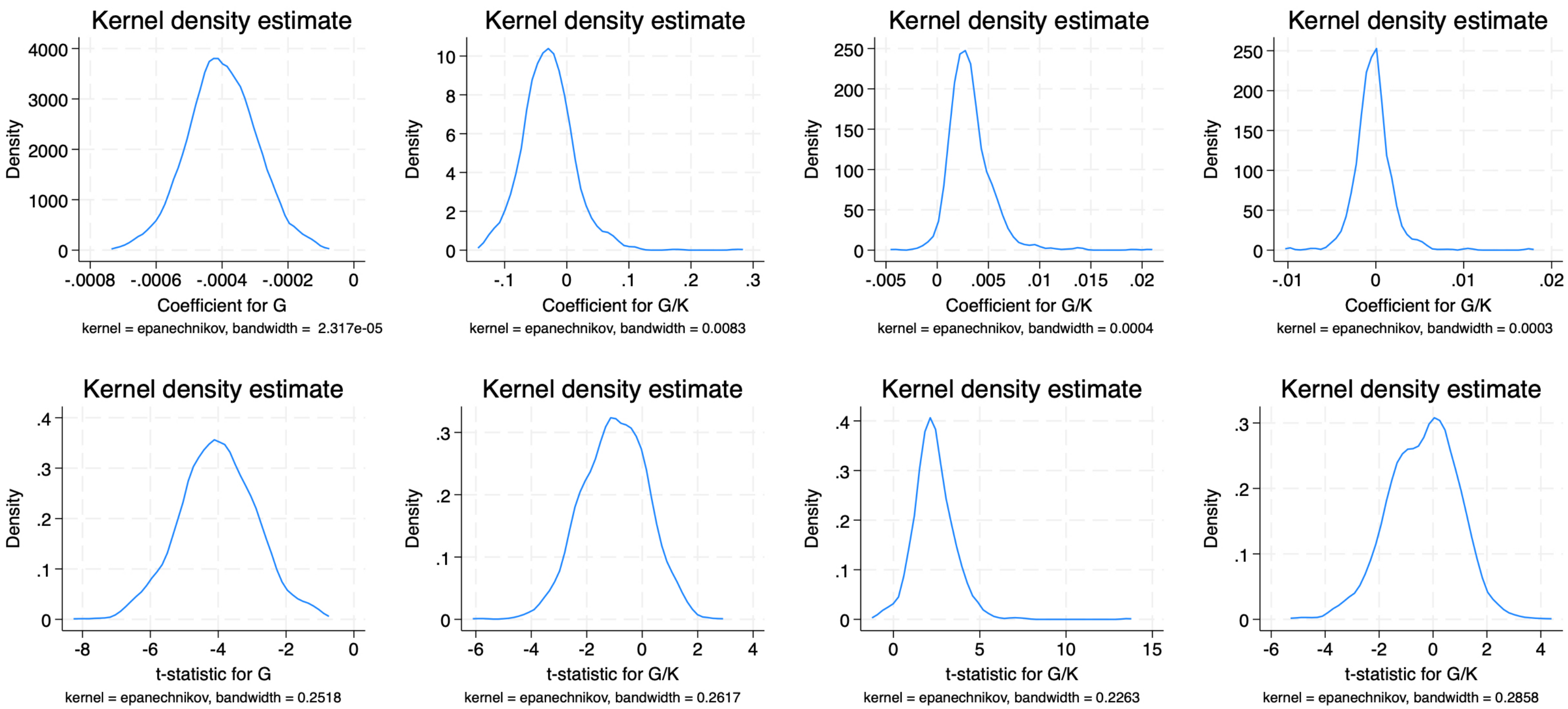

Figure S1: Kernel densities for slope coefficients and t-statistics of Firm Emissions (G) and Firm Emissions/Firm Capital (G/K). The first panel considers the coefficient (β_2) and the corresponding t-statistic for G in regressions run according to equation (2); the second panel considers the coefficient (β_2) and the corresponding t-statistic for Firm G/K in equation (3); the third panel considers the coefficient (β_2) and the corresponding t-statistic for G/K in equation (4); and the last panel considers the coefficient (β_2) and the corresponding t-statistic for G/K in equation (5).

Table S1: Regressions on facility-firm level emissions (adjusted by ownership share): R-squared contribution of different types of FE.

	(1) State FE	(2) County FE	(3) NAICS FE	(4) County & NAICS FE	(5) Gvkey FE	(6) Firm FE	(7) Facility FE	(8) Firm & Facility FE
	Emissions	Emissions	Emissions	Emissions	Emissions	Emissions	Emissions	Emissions
Observations	100,049	96,454	99,914	96,329	31,791	98,562	99,811	98,372
R2-adj	0.0153	0.4446	0.1543	0.4733	0.2499	0.2903	0.8215	0.9078
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	No	No	No	No	No	No	No
County FE	No	Yes	No	Yes	No	No	No	No
NAICS FE	No	No	Yes	Yes	No	No	No	No
Gvkey FE	No	No	No	No	Yes	No	No	No
Firm FE	No	No	No	No	No	Yes	No	Yes
Facility FE	No	No	No	No	No	No	Yes	Yes
	ln(Emissions+1)	ln(Emissions+1)	ln(Emissions+1)	ln(Emissions+1)	ln(Emissions+1)	ln(Emissions+1)	ln(Emissions+1)	ln(Emissions+1)
Observations	100,048	96,453	99,913	96,328	31,791	98,561	99,810	98,371
R2-adj	0.0916	0.4115	0.2465	0.4882	0.2590	0.5959	0.7920	0.8564
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	No	No	No	No	No	No	No
County FE	No	Yes	No	Yes	No	No	No	No
NAICS FE	No	No	Yes	Yes	No	No	No	No
Gvkey FE	No	No	No	No	Yes	No	No	No
Firm FE	No	No	No	No	No	Yes	No	Yes
Facility FE	No	No	No	No	No	No	Yes	Yes
	ln(Emissions)	ln(Emissions)	ln(Emissions)	ln(Emissions)	ln(Emissions)	ln(Emissions)	ln(Emissions)	ln(Emissions)
Observations	99,405	95,816	99,272	95,693	31,572	97,923	99,155	97,721
R2-adj	0.1160	0.4703	0.2847	0.5606	0.3139	0.6978	0.8463	0.9229
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	No	No	No	No	No	No	No
County FE	No	Yes	No	Yes	No	No	No	No
NAICS FE	No	No	Yes	Yes	No	No	No	No
Gvkey FE	No	No	No	No	Yes	No	No	No
Firm FE	No	No	No	No	No	Yes	No	Yes
Facility FE	No	No	No	No	No	No	Yes	Yes

This table reports the results of FE panel regressions on facility-level emissions. Panel A reports the *R2-adj* resulting from regressions of models that employ Total Direct Emissions as a dependent variable, while Panel B and C use $ln(Emissions + 1)$ and $ln(Emissions)$ respectively. For each panel, the FE included are: State FE (Column 1); County FE (Column 2); NAICS FE (Column 3); County and NAICS FE (Column 4); Gvkey FE (Column 5); Firm FE (Column 6); Facility FE (Column 7); Firm and Facility FE (Column 8). All specifications include Year FE, by default.

Table S2: Regressions on facility-level emissions before merge with parent-companies information: R-squared contribution of different types of FE.

	(1) State FE	(2) County FE	(3) NAICS FE	(4) County & NAICS FE	(5) Facility FE
	Emissions	Emissions	Emissions	Emissions	Emissions
Observations	87,562	85,187	87,428	85,063	87,315
R2-adj	0.0125	0.5462	0.1594	0.5722	0.9295
Year FE	Yes	Yes	Yes	Yes	Yes
State FE	Yes	No	No	No	No
County FE	No	Yes	No	Yes	No
NAICS FE	No	No	Yes	Yes	No
Facility FE	No	No	No	No	Yes
	ln(Emissions+1)	ln(Emissions+1)	ln(Emissions+1)	ln(Emissions+1)	ln(Emissions+1)
Observations	87,561	85,186	87,427	85,062	87,314
R2-adj	0.0159	0.3130	0.1926	0.3942	0.7674
Year FE	Yes	Yes	Yes	Yes	Yes
State FE	Yes	No	No	No	No
County FE	No	Yes	No	Yes	No
NAICS FE	No	No	Yes	Yes	No
Facility FE	No	No	No	No	Yes
	ln(Emissions)	ln(Emissions)	ln(Emissions)	ln(Emissions)	ln(Emissions)
Observations	87,082	84,706	86,950	84,584	86,823
R2-adj	0.0206	0.3728	0.2540	0.4745	0.8516
Year FE	Yes	Yes	Yes	Yes	Yes
State FE	Yes	No	No	No	No
County FE	No	Yes	No	Yes	No
NAICS FE	No	No	Yes	Yes	No
Facility FE	No	No	No	No	Yes

This table reports the results of FE panel regressions on facility-level emissions. Panel A reports the *R2-adj* resulting from regressions of models that employ Total Direct Emissions as a dependent variable, while Panel B and C use $ln(Emissions + 1)$ and $ln(Emissions)$ respectively. For each panel, the FE included are: State FE (Column 1); County FE (Column 2); NAICS FE (Column 3); County and NAICS FE (Column 4); Facility FE (Column 5). All specifications include Year FE, by default.

Table S3: Firm-level regressions on Tobin's Q.

	(1) Tobin	(2) Tobin	(3) Tobin	(4) Tobin	(5) Tobin	(6) Tobin	(7) Tobin	(8) Tobin	(9) Tobin	(10) Tobin	(11) Tobin	(12) Tobin
Emissions (ln)	-0.0072	0.0038										
	(-0.6949)	(0.3232)										
Emissions			-0.0000	-0.0000								
			(-1.3343)	(-0.7764)								
Emissions/Assets					-0.0427	-0.1177*						
					(-1.2900)	(-1.9667)						
Emissions/Sales							-0.0276**	-0.0372				
							(-2.2689)	(-1.4968)				
Emissions/PPENT									0.0104	-0.0149		
									(0.4944)	(-0.5719)		
Emissions/ME											-0.0221***	-0.0145**
											(-4.1741)	(-2.5376)
Size		-0.1528***		-0.1493***		-0.1576***		-0.1563***		-0.1511***		-0.1559***
		(-4.4346)		(-4.3751)		(-4.6069)		(-4.5442)		(-4.4052)		(-4.5638)
Total debt ratio		0.2205**		0.2181**		0.2143**		0.2139*		0.2148*		0.2386**
		(2.0097)		(1.9917)		(1.9737)		(1.9497)		(1.9556)		(2.1331)
ROA		1.0349***		1.0346***		0.9978***		1.0119***		1.0214***		0.9762***
		(5.7542)		(5.8092)		(5.6044)		(5.6316)		(5.7158)		(5.6115)
Sales ratio		0.1772***		0.1776***		0.1942***		0.1678***		0.1836***		0.1831***
		(3.4478)		(3.4407)		(4.1997)		(3.3038)		(3.7927)		(3.7163)
Cash ratio		0.4003**		0.3945**		0.3961**		0.3754**		0.4055**		0.3743**
		(2.5167)		(2.4818)		(2.4768)		(2.3527)		(2.5385)		(2.3835)
Firm age (ln)		-0.0978**		-0.0937*		-0.0927*		-0.0948*		-0.0943*		-0.0915*
		(-2.0022)		(-1.9417)		(-1.9215)		(-1.9660)		(-1.8811)		(-1.8909)
Multi-owners		-0.0789		-0.0844		-0.1035		-0.1017		-0.0890		-0.0895
		(-1.0473)		(-1.1356)		(-1.4301)		(-1.4132)		(-1.1619)		(-1.2374)
Constant	0.3739***	1.7368***	0.3383***	1.7177***	0.3419***	1.8043***	0.3489***	1.8080***	0.3230***	1.7384***	0.3471***	1.7700***
	(5.9321)	(4.9930)	(47.4700)	(4.9435)	(33.7521)	(5.1954)	(39.7775)	(5.1213)	(25.8040)	(4.9436)	(79.3592)	(5.1294)
Observations	2,954	2,402	2,962	2,410	2,962	2,410	2,952	2,401	2,953	2,401	2,962	2,410
R2-adj	0.7600	0.7987	0.7611	0.7996	0.7611	0.8011	0.7636	0.8014	0.7609	0.7995	0.7667	0.8017
Firms	396	331	397	332	397	332	396	331	396	331	397	332
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm

This table reports the results of FE panel regressions of emissions on firm valuation. In this table, emissions data are collapsed at the firm level. The dependent variable is the natural logarithm of Tobin's Q – defined as the market value of assets over the book value of assets. Key regressors are: the natural logarithm of emissions (columns (1)–(2)), unscaled emissions (columns (3)–(4)), emissions scaled by total assets (columns (5)–(6)), emissions scaled by total sales (columns (7)–(8)), emissions scaled by Total Property, Plant and Equipment (PPENT) (columns (9)–(10)), and emissions scaled by the Market value of Equity (ME) (columns (11)–(12)). Standard firm-level controls are added to specifications reported in even columns, and include Size (the natural logarithm of total assets), Total debt ratio (total debt over total assets), ROA (net income over total assets), Sales ratio (sales over total assets), Cash ratio (cash and short term investments over total assets), Firm age (the natural logarithm of the difference between the current year and the first year a firm was recorded in Compustat), and Multi-owners (a dummy equal to one when facilities are owned by more than one owner). All specifications include firm and year FE. *t*-statistics are reported in parentheses. Standard errors are clustered at the firm level. ***, **, * denote statistical significance at the 1%, 5%, and 10% level respectively.

Table S4: Regressions on facility-level emissions – with and without the lag of the dependent variable.

	(1) Facility FE no lag	(2) Facility FE no lag	(3) Facility FE lag	(4) No facility FE lag	(5) Dyn-Xtdpdsys
$g_{i(t-1)}$			0.699***	0.949***	0.956***
			(25.859)	(261.549)	(69.744)
Constant	385.291***	388.464***	110.997***	11.447***	27.529***
	(7.028e+14)	(1.152e+15)	(10.345)	(11.482)	(4.096)
Observations	85,745	76,878	76,878	77,128	77,128
R-squared	0.9214	0.9283	0.9647	0.9604	
Facility FE	Yes	Yes	Yes	No	No
Year FE	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	No	No
Cluster	Facility	Facility	Facility	Facility	NA
AR(2) test: p-value					0.1113

This table reports the results of OLS regressions (columns 1–4) and of regressions run according to the estimator by Arellano & Bover (1995) and Blundell & Bond (1998), System-GMM (column 5). The dependent variable is facility-level emissions. Columns (1) and (2) do not include any independent variable (apart from the FE mentioned), while columns (3)–(5) add the lag of the dependent variable ($g_{i(t-1)}$). Column (1) is run on the full sample, while column (2) excludes observations that would disappear when including the lag of the dependent variable. Columns 1–3 include Facility FE. All specifications include Year FE. Standard errors are clustered at the facility level in columns 1–4. ***, **, * denote statistical significance at the 1%, 5%, and 10% level respectively.

Table S5: Effect of scaling by a proxy for firm size: Monte Carlo simulations at the firm level assuming a negative correlation between Firm Emissions and Firm value and persistence in Firm Value (true value of $\psi_G = -0.0004$; true value of $\rho_V = 0.68$).

Panel A: Summary statistics	Obs	Monte Carlo Mean	Sample (Table 1)	Sample winsorized				
G_{ft} - Median	1,000	524.7831	327.745	327.745				
G_{ft} - Kurtosis	1,000	8.5719	33.7462	27.7048				
K_{ft} - Median	1,000	3510.21	3060.5	3060.5				
K_{ft} - Kurtosis	1,000	11.0415	42.7675	41.9282				
G_{ft}/K_{ft} - Median	1,000	.1199	.1483	.1483				
G_{ft}/K_{ft} - Kurtosis	1,000	267.8001	795.8539	20.2895				
V_{ft} - Median	1,000	5621.955	5105.557	5105.557				
V_{ft} - Kurtosis	1,000	16.8736	249.7499	25.4312				
Panel B: equation (2)	Obs	Mean	Std. Dev.	Skewness	Kurtosis	P1	P50	P99
Coefficient on G_{ft} (β_2)	1000	.0094	.5527	.0052	2.8463	-1.2014	.0011	1.2638
T-statistic for G_{ft}	1000	-.0075	1.0466	-.1161	2.9062	-2.4577	.0019	2.2759
Panel C: equation (3)	Obs	Mean	Std. Dev.	Skewness	Kurtosis	P1	P50	P99
Coefficient on G_{ft}/K_{ft} (β_2)	1000	-4.4265	154.3006	-.5519	4.5015	-437.6916	11.447	328.4109
T-statistic for G_{ft}/K_{ft}	1000	.0777	1.075	.0406	2.9654	-2.3783	.0888	2.5252
Panel D: equation (4)	Obs	Mean	Std. Dev.	Skewness	Kurtosis	P1	P50	P99
Coefficient on G_{ft}/K_{ft} (β_2)	1000	1.9835	2.0533	1.3086	6.8078	-1.5581	1.6452	8.5837
T-statistic for G_{ft}/K_{ft}	1000	1.1886	.9399	-.1676	3.2071	-.9747	1.2894	3.4023
Panel E: equation (5)	Obs	Mean	Std. Dev.	Skewness	Kurtosis	P1	P50	P99
Coefficient on G_{ft}/K_{ft} (β_2)	1000	2.6004	2.2	.9123	5.7488	-2.2881	2.3897	9.262
T-statistic for G_{ft}/K_{ft}	1000	1.5818	1.0146	-.1369	3.8678	-.8698	1.625	3.7356
Panel F: equation (6)	Obs	Mean	Std. Dev.	Skewness	Kurtosis	P1	P50	P99
Coefficient on G_{ft} (β_2)	1000	-.0004	.0002	-.0827	2.9178	-.001	-.0004	.0001
T-statistic for G_{ft}	1000	-1.7026	1.0325	-.063	2.7302	-4.0628	-1.6981	.5573

This table reports the results of OLS regressions on 1,000 simulated samples. In Panel A, we report the descriptive statistics for the average value in each of the 1,000 simulated samples for the following variables: Firm Emissions (G_{ft}), Firm Capital (K_{ft}), Firm Emissions scaled by Firm Capital (G_{ft}/K_{ft}), and Firm Value (V_{ft}). For the same variables, we report the corresponding statistic in our sample, without any adjustment, and with a winsorization at the 1^{st} and 99^{th} percentile. In Panel B, we report the distribution of the estimated coefficients and t-statistics for the coefficient on G_{ft} (β_2) in regressions run according to equation (2). In Panels C, D and E we report the distribution of the estimated coefficients and t-statistics for G_{ft}/K_{ft} (β_2) in regressions run according to equations (3), (4), and (5), respectively. The Data Generating Process (DGP) is based on the same 1,000 simulated samples used for Table 5, but the data is aggregated at the firm-year level, although we start from the datasets with facility-firm-county-year structure. Further details of the DGP are described in Section B. All regressions have firm FE and year FE and standard errors are clustered at the firm level.

Table S6: Effect of different types of FE: Monte Carlo simulations at the facility level.

Panel A: True $\rho_g = 0$ – Regressions on facility-level Emissions	Obs	Mean	Std. Dev.	Skewness	Kurtosis	P1	P10	P50	P90	P99
$\hat{\rho_g}$	1000	-.0001	.0041	.3112	3.3841	-.0088	-.0052	-.0004	.0052	.0102
R2-adj Facility FE and Year FE	1000	0	.0015	.0563	2.7648	-.0036	-.002	0	.002	.0036
R2-adj Firm FE, County FE and Year FE	1000	.0017	.0007	.1337	2.8713	.0002	.0009	.0017	.0026	.0033
R2-adj Firm FE and Year	1000	.0018	.0005	.0807	2.905	.0006	.0011	.0018	.0024	.003
R2-adj County FE and Year FE	1000	0	.0005	.0318	2.7664	-.0009	-.0006	0	.0006	.001
Panel B: True $\rho_g = 0$ – Regressions on facility-level Emissions/Capital	Obs	Mean	Std. Dev.	Skewness	Kurtosis	P1	P10	P50	P90	P99
R2-adj Facility FE and Year FE	1000	0	.0014	1.0004	4.8993	-.0024	-.0016	-.0002	.0018	.0042
R2-adj Firm FE, County FE and Year FE	1000	.0003	.0006	.343	2.9817	-.0011	-.0005	.0002	.0011	.0018
R2-adj Firm FE and Year	1000	.0003	.0004	.58	3.4857	-.0006	-.0002	.0002	.0009	.0015
R2-adj County FE and Year FE	1000	0	.0004	.5227	3.7397	-.0009	-.0005	0	.0006	.0012
Panel C: True $\rho_g = 0.5$ – Regressions on facility-level Emissions	Obs	Mean	Std. Dev.	Skewness	Kurtosis	P1	P10	P50	P90	P99
$\hat{\rho_g}$	1000	.4999	.0046	.3456	3.3664	.4902	.494	.4997	.5055	.5118
R2-adj Facility FE and Year FE	1000	.193	.0038	-.0508	2.9924	.1837	.1883	.1931	.1981	.2019
R2-adj Firm FE, County FE and Year FE	1000	.0426	.002	.1453	3.0767	.0381	.0401	.0426	.0452	.0478
R2-adj Firm FE and Year	1000	.0426	.002	.1645	3.04	.0383	.0401	.0426	.0452	.0479
R2-adj County FE and Year FE	1000	.0116	.0009	.0316	3.0956	.0096	.0105	.0116	.0127	.0136
Panel D: True $\rho_g = 0.5$ – Regressions on facility-level Emissions/Capital	Obs	Mean	Std. Dev.	Skewness	Kurtosis	P1	P10	P50	P90	P99
R2-adj Facility FE and Year FE	1000	.0423	.012	-.1085	2.9468	.015	.0259	.0426	.0569	.0694
R2-adj Firm FE, County FE and Year FE	1000	.0154	.0042	-.2337	3.0353	.005	.0097	.0157	.0207	.0245
R2-adj Firm FE and Year	1000	.0154	.0042	-.234	3.0769	.0052	.0098	.0157	.0207	.0244
R2-adj County FE and Year FE	1000	.0105	.0027	-.391	3.1398	.0033	.0068	.0107	.0138	.0162

This table reports the results of Monte Carlo simulations based on 1,000 simulated samples. For each of the 1,000 replications, we run an OLS regression where the dependent variable is either Emissions (Panels A and C) or Emissions/Capital (Panels B and D), and the independent variables are different types of FE. We then collect the adjusted R-squared (*R2-adj*) for each regression and compute the mean, standard deviation and percentiles of the distribution of *R2-adj* for the whole set of 1,000 replications. The Data Generating Process (DGP) is described in Section B. In both Panel A and B, firm FE are correlated with the average facility-level capital stock, k_{ifct}. The first line of Panels A and C reports the estimated autocorrelation coefficient in emissions, $\hat{\rho_g}$, according too the dynamic panel data estimator by Arellano & Bover (1995) and Blundell & Bond (1998). The true value of ρ_g is either 0 (Panels A and B) or 0.5 (Panels C and D).

Table S7: Effect of taking the logarithm of Firm value and Firm Emissions: Monte Carlo simulations at the firm level assuming a negative correlation between Firm Emissions and Firm value (true value of $\psi_G = -0.0004$).

Panel A: Summary statistics	Obs	Mean	Std. Dev.	Skewness	Kurtosis	P1	P50	P99
$\ln(G_{ft})$ - Median	1,000	6.2623	.0135	-.1191	3.2504	6.2291	6.2629	6.2924
$\ln(G_{ft})$ - Kurtosis	1,000	3.7968	.0651	-.056	2.8216	3.645	3.7976	3.9406
$\ln(K_{ft})$ - Median	1,000	8.164	.0168	-.0037	2.9606	8.1249	8.1643	8.2034
$\ln(K_{ft})$ - Kurtosis	1,000	3.0549	.0405	.07	3.3955	2.967	3.0544	3.1531
$\ln(V_{ft})$ - Median	1,000	8.6244	.0231	.054	3.3718	8.5686	8.6248	8.6804
$\ln(V_{ft})$ - Kurtosis	1,000	3.6203	.1326	.9774	5.8639	3.3717	3.6084	4.0286
Panel B: Coefficient estimates	**Obs**	**Mean**	**Std. Dev.**	**Skewness**	**Kurtosis**	**P1**	**P50**	**P99**
Coefficient on $\ln(G_{ft})$	1,000	-.0006	.0016	-.0883	6.8197	-.0048	-.0006	.0036
T-statistic for $\ln(G_{ft})$	1,000	-.465	1.0034	.1069	2.4941	-2.5613	-.5022	1.6923

This table reports the results of OLS regressions on 1,000 simulated samples. In Panel A, we report the descriptive statistics for the average value in each of the 1,000 simulated samples for the following variables: log of Firm Emissions, $\ln(G_{ft})$, log of Firm Capital, $\ln(K_{ft})$, and log of Firm Value, $\ln(V_{ft})$. In Panel B, we report the distribution of the estimated coefficients and t-statistics for the coefficient on $\ln(G_{ft})$–β_2–in regressions run according to equation (2), after taking the logs of both the dependent variable, V_{ft} and the independent variables, G_{ft} and Π_{ft}. The true value of $\psi_G = -0.0004$ in the Data Generating Process (DGP) is for the original (i.e., unlogged) version of equation (2). Further details of the DGP are described in Section B. All regressions have firm FE and year FE and standard errors are clustered at the firm level.