



Meet, beat, and pollute

Jake Thomas¹ · Wentao Yao² · Frank Zhang¹ · Wei Zhu³ 

Accepted: 29 May 2022 / Published online: 21 June 2022
© The Author(s) 2022

Abstract

We investigate two related questions about the trade-off between the short-term pressures on managers to meet earnings targets and the long-term environmental benefits of reduced pollution. Do firms release more toxins by cutting back on pollution abatement costs to boost earnings in years they meet earnings benchmarks? If so, is that relation weaker for firms with higher environmental ratings? Using Environmental Protection Agency (EPA) data on toxic emissions, we find that U.S. firms pollute more when they meet or just beat consensus earnings per share (EPS) forecasts, suggesting that meeting expectations is a more important goal than reducing pollution. We find this relation is stronger, not weaker, for firms with higher environmental ratings: they increase pollution even *more* when meeting earnings benchmarks than firms with lower ratings. This suggests that highly rated firms build regulatory and reputational slack over time and use it when needed to soften the negative impact of increased pollution. We contribute to the real earnings management and environmental economics literatures by documenting a negative externality of financial reporting incentives on the environment and society. We also contribute to the corporate sustainability literature by showing that an environmental, social, and governance (ESG) focus does not curb managerial short-termism.

Keywords ESG · Corporate sustainability · Managerial short-termism · Meeting earnings benchmarks · Real earnings management · Externality

JEL classification G10 · M14 · M41 · Q50

✉ Wei Zhu
zhuwei@illinois.edu

¹ Yale University, New Haven, CT, USA

² Xiamen University, Xiamen, Fujian, China

³ University of Illinois at Urbana-Champaign, Champaign, IL, USA

1 Introduction

Environmental sustainability (E-S) is an increasingly desirable corporate attribute, where firms are expected to deliver their goods and services in a manner that is both environmentally sustainable and profitable.¹ A common theme of E-S is that it promotes a long-term view of shareholder value and thus curbs managerial short-termism (Eccles et al. 2014; Fink 2019; Starks et al. 2020). In this paper, we investigate two related questions about the trade-off between E-S and managerial short-termism, where short-termism is typically reflected as efforts to meet earnings benchmarks (e.g., Dechow and Sloan 1991). Do firms release more toxins by cutting back on pollution abatement costs to boost earnings in years they meet earnings benchmarks? If so, is that relation weaker for firms with higher environmental ratings? We define meeting earnings benchmarks as meeting or just beating (by two cents or less) analysts' consensus earnings per share (EPS) forecasts (e.g., Caskey and Ozel 2017). We use releases of about 400 toxins to capture pollution, and measure environmental ratings as the environmental component (E ratings) of firm-level environmental, social, and governance (ESG) ratings.

One motivation to investigate these two questions is the adverse effects of toxins on public health, an important concern that has received much attention in the U.S. (e.g., Currie et al. 2014). Understanding why firms pollute is thus of interest to the general public, and of particular interest to regulators tasked to preserve clean air, water, and land. Second, while prior research documents the negative externalities of real earnings management (REM) on firm stakeholders such as investors and employees (Caskey and Ozel 2017; Raghunandan 2021), REM's impact on society remains relatively unexplored.² Finally, whether the focus on long-term value promoted by E-S moderates the managerial short-termism underlying REM is also unexplored.

Our data on toxic releases are taken from the EPA's Toxic Release Inventory (TRI) database, which covers chemical releases that may have long-term health and environmental effects.³ While TRI chemicals are toxic, they are not as hazardous as the six *criteria* air pollutants (including SO₂, lead, and ozone), which are monitored by the EPA in real-time to allow immediate responses if limits are breached.⁴ In contrast, TRI disclosures are self-reported and filed with a delay (by June 30 of the next year).

To illustrate how firms might cut pollution abatement costs to meet earnings benchmarks, consider the seven years around 2009 for Texas Instruments (TI), a semiconductor manufacturer in our sample. TI just met the consensus analyst forecast

¹ Consistent with the definition of corporate sustainability in Grewal and Serafeim (2020), we define E-S as an intentional strategy to create long-term value through improved measurable environmental impact.

² A notable exception is Liu et al. (2021). We discuss their findings later in this section and in Section 2.2.

³ TRI is the most comprehensive industrial toxic emission disclosure program in the U.S. It requires the disclosure of over 400 toxic chemicals that are released to the air, water, and land by thousands of plants. It was established by Section 313 of the Emergency Planning and Community Right-to-Know Act (EPCRA) and provides data beginning in 1987. It is the most widely used EPA pollution database in the academic literature (e.g., Akey and Appel 2021; Kim and Xu 2022). Given that our primary analyses are conducted at the firm level, we aggregate annual plant-level releases reported to the EPA to generate firm-level pollution data.

⁴ For example, SO₂, along with particulates formed by SO₂, are associated with respiratory disease, heart disease, and premature mortality. They also cause acid rain that contributes to the acidification of forests and lakes. Lead and ozone are unusual, as they are criteria chemicals that are also covered under TRI.

in 2009, and in the other six years it either missed or comfortably beat the consensus (by more than 2 cents). In 2009, TI released an incremental 6.4 tons of hydrogen fluoride (HF), a toxin generated when fluorine etches silicon used to make semiconductor wafers.⁵ Packed bed wet scrubbers, the most common abatement technology for HF, can be switched off to reduce abatement levels and later switched on. Using an estimate of \$20.6 million for the abatement cost per ton of HF, pre-tax earnings per share for TI increased by about 10 cents.⁶ We note that TI has a high E rating, which runs counter to the view underlying our second research question: managers of environmentally friendly firms with a long-term focus should be less likely to pollute to meet short-term earnings benchmarks.

Consistent with the TI illustration, we find that toxin releases for “suspect” firm-years in our sample—firm-years that meet or beat the consensus by 2 cents or less—are higher by about 15%, an economically and statistically significant increase, relative to firm-years that miss or comfortably beat forecasts. We control for a variety of firm-level proxies for normal levels of production, various fixed effects that account for unobservable permanent firm/plant characteristics, and time-varying industry-wide shocks. These results suggest that pollution rises in years when managers who are under pressure to meet earnings benchmarks cut pollution abatement costs to boost earnings.

We confirm that this first finding is robust. We observe similar results when we scale toxic releases by revenues or weight released toxins by their relative toxicity (assigned by the EPA’s Risk-Screening Environmental Indicator (RSEI) program). Our results are not due to pollution increases caused by firms overproducing to boost earnings, as production does not increase abnormally in years when firms meet earnings benchmarks. And the observed rise in pollution levels is not a byproduct of (unexpected) performance improvements, because pollution levels do not increase for firms that comfortably beat forecasts. Finally, as a falsification test, we show that the positive relation between pollution and meeting earnings benchmarks declines when we use noisier proxies (from earlier in the fiscal year) for those benchmarks.

Turning to our second research question, we find that the additional toxic releases observed in suspect firm-years *increase* with a firm’s E rating from MSCI ESG KLD STATS (shortened to MSCI). As mentioned in the TI illustration above, this result is surprising because managers of firms with higher E ratings—which we find reliably reflect environmental sustainability track records—should focus on the long term and avoid pollution increases. Although these firms pollute less on average, which is consistent with a long-term focus, they increase pollution levels when needed to meet earnings benchmarks. Managers of firms with high E ratings appear to be trading off multiple conflicting goals (Levinthal and Rerup 2021): they maintain lower pollution

⁵ While HF accounts for most (86%) of the increase in toxic release for 2009, TI also released excess amounts of other toxins. The ratio of total toxic release to revenue is 4.82 (thousands of lbs. per million dollars of revenue) in 2009, which is 42% higher than the average ratio of 3.40 over the other six years.

⁶ Pollution abatement costs include salaries/wages, costs of materials/supplies, utility/energy costs, and costs of purchased services (e.g., contract work and lease rentals) used to test/monitor, recycle, treat, and dispose pollutants. Under U.S. GAAP, these operating costs are allocated to cost of goods sold (White et al. 1995). Because TI, like other firms, does not provide pollution abatement costs, the HF abatement cost estimate we use is inferred from data provided by Intel (<https://digital.osl.state.or.us/islandora/object/osl:21322>). We therefore caution that our estimate may not represent TI’s actual abatement cost for HF in 2009. To compute the per share impact, we use 1.269 billion shares, the weighted average number of shares outstanding for TI for 2009.

levels over the long term, yet increase pollution to meet market expectations for short-term earnings.

To confirm the robustness of this second result, we conduct additional tests leveraging other information provided by MSCI. First, in addition to E ratings, MSCI provides S&G ratings, which serve as placebos because they should have weaker links with toxic releases. We find that the S&G ratings do not explain variation in the positive relation between pollution and meeting benchmarks. Second, these E ratings cover a wide range of environmental performance indicators, some that are directly related to toxic releases generated by production processes (e.g., pollution prevention and past toxic emissions) and others that are not (e.g., producing environmentally beneficial products). We find that the moderating effect of the overall E rating is driven by the subset of environmental indicators that are directly related to toxic releases from production processes. Finally, MSCI separates environmental indicators into environmental strengths (e.g., pollution prevention programs) and concerns (e.g., quantities of past toxic emissions), representing mainly inputs to and outcomes from environmental programs, respectively. We find that the relation between pollution and meeting earnings benchmarks only increases (decreases) with strengths (concerns) that are directly related to toxic releases.

One plausible explanation for our surprising second result is that firms with better environmental track records (reflected in higher E ratings) generate “pollution slack” that allows managers to increase pollution, when needed, at a lower cost than for firms with poor E ratings. Slack could arise from higher accumulated emission allowances that offer firms the flexibility to temporarily increase pollution without exceeding regulatory emission limits (regulatory slack). Slack could also arise if a good reputation that firms have built over the long term with stakeholders such as consumers, communities, and investors softens the negative impact of occasional pollution increases (reputation slack).

Consistent with this explanation, we find that higher E ratings are associated with less toxic releases, fewer EPA enforcement cases, and smaller EPA enforcement penalties in the past. We also find that the moderating effect of environmental track records on the positive relation between pollution and meeting earnings benchmarks observed across firms also holds within firms at the plant level. Firms that are polluting more to meet earnings benchmarks increase pollution at plants that have better track records.

We consider the impact of our findings on sustainable investing, which refers to the practice of incorporating ESG factors in portfolio decisions (Amel-Zadeh and Serafeim 2018). Prior studies (Starks et al. 2020; Cao et al. 2021) show that sustainable funds are more patient—i.e., they take a longer-term perspective—and respond less to short-term earnings information.⁷ However, if sustainable funds invest in firms with higher ESG ratings (Curtis et al. 2021; Heath et al. 2021), our findings above suggest that these investees are likely to pollute more to achieve short-term earnings benchmarks. We find that this is indeed the case: firms with higher levels of sustainable institutional ownership tend to pollute more when meeting earnings benchmarks.

⁷ The notion that ESG strategies cure managerial short-termism is popular in sustainable investing. See, for example, <https://www.design-portfolio.co.uk/insights/esg-short-term-pain-long-term-value/>.

It is useful to contrast our findings with those in Liu et al. (2021), who document increased pollution by Chinese firms meeting earnings benchmarks. On the surface, the analyses of our first research question may seem similar to Liu et al. (2021), but there are important differences between our inferences and theirs. Their result only holds for state-owned (SOE) firms controlled by the central government. Chinese non-SOE firms, which are the relevant peer group for our U.S. sample, do *not* pollute more to meet earnings benchmarks. To better understand the different inferences from the two studies, we focus on the respective pollutants investigated: we study a broad group of TRI chemicals, whereas they study SO₂, a criteria chemical that the EPA monitors in real time because it is so hazardous.⁸ We too find that U.S. firms do not emit more SO₂ when meeting earnings benchmarks.⁹ By extending our study to incorporate the second research question and ESG rating data, we are able to provide a fuller picture of cross-sectional variation in pollution levels around suspect firm-years.

We contribute to the real earnings management (REM) literature by showing that REM is associated with increased pollution by U.S. firms. Early REM research, which considers decisions such as overproduction and R&D cuts, suggests that the costs of managerial short-termism are borne by shareholders. Subsequent research finds that the costs of REM are also borne by employees, due to cuts in spending on employee safety and compensation (Caskey and Ozel 2017; Raghunandan 2021). Our results suggest that the negative effects of REM extend beyond firm boundaries to the environment and society at large.¹⁰

Second, we add to the corporate sustainability/ESG literature by showing that even though firms that are rated as being more environmentally friendly pollute less in general, they increase pollution when meeting earnings benchmarks. This runs counter to the view that corporate sustainability curbs managerial short-termism. Finally, our study contributes to the environmental economics literature by adding financial reporting incentives to the previously documented determinants of toxic releases, which include financial constraints (Cohn and Deryugina 2018; Goetz 2018; Kim and Xu 2022), legal liability (Akey and Appel 2021), and third-party auditors (Duflo et al. 2013).

The remainder of our paper is laid out as follows. Section 2 develops hypotheses after reviewing institutional details and prior research. Section 3 provides details of our data and samples. Section 4 examines the effect of meeting benchmarks on toxic releases. Section 5 investigates the role of E-S. Section 6 conducts additional analyses, and Section 7 concludes.

⁸ Liu et al. (2021) also find no link between earnings management and pollution for SOEs controlled by local governments. The SO₂ pollution–earnings management link for SOEs controlled by the central government likely reflects some form of regulatory enforcement exemption that such entities enjoy (Eaton and Kostka 2017).

⁹ As further confirmation, we find firms meeting earnings benchmarks do not release more ozone and lead, which are the two TRI toxins also classified by the EPA as criteria chemicals and monitored in real time.

¹⁰ We investigate whether cutting pollution abatement in suspect years is related to other forms of REM (overproduction and cuts to R&D, advertising, and SG&A), earnings guidance designed to increase the odds of meeting benchmarks, and accruals-based earnings management. We find no evidence of other types of REM, which suggests that cutting pollution abatement activities is the lowest-cost REM option for our sample. We also find no evidence of firms guiding analyst forecasts to allow meeting earnings benchmarks. We do find some evidence consistent with firms engaging in pollution-related REM having accumulated more accruals in prior years (indicated by higher net asset balances).

2 Institutional details, related literature, and hypothesis development

2.1 Environmental regulation, pollution abatement, and accounting for pollution abatement

2.1.1 Environmental regulation

Amid elevated concerns about environmental pollution in the 1970s, Congress passed several laws (e.g., Clean Air Act, Clean Water Act, Safe Drinking Water Act, and Resource Conservation and Recovery Act) and consolidated many environmental responsibilities under one federal agency, the U.S. Environmental Protection Agency (EPA). The EPA has since served as the central authority for federal research, standard-setting, monitoring, and enforcement of environmental protection.¹¹ Compliance with environmental rules and standards causes firms to partially internalize environmental costs created by their operations.

Disclosure serves as an important tool within the environmental regulatory system. In 1986, Congress passed the Emergency Planning and Community Right-to-Know Act (EPCRA) to support and promote emergency planning and to provide the public with information on the release of toxic chemicals in their communities.¹² In particular, Section 313 of the EPCRA established the Toxic Release Inventory (TRI) program, which requires plants to disclose annually, to the EPA, the release of toxic chemicals that may cause negative chronic (long-term) health and environmental effects.¹³ Under Section 313, reporting the amount of each chemical released to the environment is required for U.S. facilities that have at least ten full-time employees; operate in one of roughly 400 industries defined at the six-digit NAICS level; and manufacture, process, or otherwise use listed chemicals in amounts above established levels.¹⁴ By making information about toxic chemicals publicly available, TRI creates a strong incentive for firms to improve their waste management and environmental performance. Although TRI data are self-reported, ample evidence suggests that they are generally reliable.¹⁵ As such, TRI data are widely used in the literature to measure pollution and environmental performance (e.g., Chatterji et al. 2009; Akey and Appel 2021; Kim and Xu 2022).

Some toxins are more hazardous and have acute (immediate) effects, and consequently are subject to real time monitoring. The six pollutants that fall in this category,

¹¹ The EPA develops and enacts rules and standards that provide details regarding the implementation of environmental laws. It also works with state and local authorities to ensure compliance with environmental regulation by conducting inspections/investigations and enforcing sanctions.

¹² A “release” of a chemical means that it is emitted to the air or water, or placed in some type of land disposal.

¹³ See <https://www.epa.gov/sites/default/files/documents/1998qa.pdf>.

¹⁴ In addition to reporting total release, plants are also required to provide separately the amount (in pounds) released into the air, water, and ground. In 2020, TRI added another 300-plus toxic chemicals to the 400 already being reported.

¹⁵ While there are no penalties associated with legitimate releases of TRI chemicals, there are penalties associated with misreporting (Greenstone 2003). See the following examples of settlements on misreporting: <https://www.epa.gov/newsreleases/three-connecticut-companies-provide-public-chemical-information-under-epa-settlements>. The EPA also conducts periodic audits to verify data integrity. Bui and Mayer (2003) note that there is little evidence of systematic over- or underestimation in the TRI data. EPA (1998), an audit of TRI facilities, finds that facility and surveyor estimates are within 3% of each other for most industries.

labeled *criteria* air pollutants, are sulfur dioxide (SO₂), carbon monoxide (CO), nitrogen dioxide (NO₂), ozone (O₃), lead (Pb), and particulate matter (PM₁₀ and PM_{2.5}). Ambient concentrations of these pollutants are measured and monitored at more than 4,000 monitoring stations owned and operated by state environmental agencies.¹⁶ The agencies send hourly or daily measurements of pollutant concentrations to the AQS (Air Quality System), an EPA database. Firms that are noncompliant with national ambient air quality standards can face sanctions under the Clean Air Act, which may result in severe economic consequences.

2.1.2 Pollution abatement

Environmental standards typically require plants to integrate pollution abatement technologies and measures into their production processes. As a result, a plant's pollution level is jointly determined by its production processes and pollution abatement activities. Pollution abatement activities fall into two broad categories: pollution prevention and postproduction processing (see Appendix A).

Pollution prevention generally refers to practices that reduce the amount of any hazardous substance, pollutant, or contaminant entering any waste stream or otherwise released into the environment prior to recycling, energy recovery, or treatment. Examples of pollution prevention include process or procedure modifications, equipment or technology modifications, reformulation or redesign of products, substitution of raw materials, and improvements in housekeeping, maintenance, training, and inventory control.

Postproduction processing generally refers to recycling, energy recovery, and treatment of any hazardous substance, pollutant, or contaminant chemical. Recycling is the recovery for reuse of a toxic chemical. Energy recovery is the combustion of a toxic chemical in a combustion unit that is integrated into an energy recovery system (i.e., industrial furnaces, industrial kilns, and boilers). Treatment includes a variety of methods (biological treatment, chemical oxidation, and incineration) to reduce the toxicity of chemicals in waste streams,

2.1.3 Accounting for pollution abatement costs

As pollution abatement is integrated into production processes, its costs are allocated to product costs.¹⁷ The abatement costs generally fall into two categories: depreciation of capital expenditures and operating costs. Capital expenditures include the purchase, installation, and startup costs of abatement technologies and equipment to prevent, recycle, energy recover, treat, and dispose of chemical waste. Costs of production redesign, process modification, and material substitution are also capitalized, as they typically yield reduced emission costs over the long term.

¹⁶ Criteria air pollutants are monitored separately because they are extremely hazardous, especially to vulnerable populations such as asthmatics, children, and the elderly. Ozone and lead are also covered under TRI.

¹⁷ These costs are initially assigned to cost pools (most often an overhead pool and occasionally a separate environmental pool). They are then allocated to products at the end of accounting periods based on cost drivers such as labor hours, production volume, materials used, and square footage of facility space (White et al. 1995).

Operating costs of pollution abatement occur during both pollution prevention and postproduction processing. Examples of pollution prevention costs include the cost of running and maintaining pollution prevention equipment (e.g., leak detection equipment) and the cost of modifying production processes. Examples of postproduction processing costs include employee salaries/wages, costs of materials/supplies, utility/energy costs, and costs of purchased services (e.g., contract work and lease rentals) used to test/monitor, recycle, treat, and dispose of pollutants.

The annual cost of pollution abatement (i.e., operating cost and depreciation expense) can be a nontrivial portion of total product cost, especially in certain industries. Joshi et al. (2001) offer a reliable estimate, likely at the high end of the range. Using confidential plant-level data for 55 steel mills from 1979 to 1988 drawn from the Pollution Abatement Costs and Expenditures (PACE) surveys, they estimate that periodic abatement costs account for 46%–53% of total product cost. Furthermore, according to the 2005 PACE Survey conducted by the U.S. Census Bureau and EPA, operating costs of pollution abatement are more than six times the depreciation expense of abatement investment. Given that most periodic pollution abatement costs are variable, we believe that cutting pollution abatement costs is a viable and effective way to boost short-term earnings.

2.2 Managerial short-termism, real earnings management, and externalities

The real earnings management (REM) literature provides compelling evidence, both survey-based and empirical, of managers manipulating real activities to meet earnings benchmarks even if these actions sacrifice long-term shareholder value.¹⁸ Graham et al. (2005) report that about 80% of surveyed U.S. executives admit to sacrificing long-term value when engaged in REM. Consistent with this survey, Bhojraj et al. (2009) and Cohen and Zarowin (2010) empirically show that REM negatively impacts long-run performance.

In addition to negative effects on long-term shareholder value, recent studies show that REM can also result in negative externalities on firm employees. Specifically, efforts to meet or beat analyst forecasts lead to more workplace injuries (Caskey and Ozel 2017) and lower wages (Raghunandan 2021). We extend this emerging literature on externalities due to REM by examining the effects of meeting earnings benchmarks on the environment and society.

Using data on Chinese listed companies, Liu et al. (2021) show that Chinese state-owned enterprises (SOEs) controlled by the central government emit more SO₂ when meeting or just beating analyst forecasts, but locally controlled SOEs and non-SOEs do not. Liu et al. (2021) posit that central SOEs are protected by the central government from the environmental monitoring and enforcement carried out by local environmental protection bureaus (Eaton and Kostka 2017). More relevant to our study is their finding that non-SOE firms, which represent the subset of Chinese firms that are most comparable to our U.S. sample, do not increase SO₂ emissions when they meet earnings benchmarks. Compared with the TRI toxins that we focus on, SO₂ is far more hazardous and monitored in real-time for U.S. firms.

¹⁸ Reasons for managerial myopia include career concerns and the dependence on short-term profits in efficient compensation contracts (Bénabou and Tirole 2010).

2.3 Corporate sustainability and ESG

As regulation may be insufficient to reduce negative externalities—especially those driven by managerial short-termism—corporate sustainability has recently been suggested as a market-based solution (Bénabou and Tirole 2010). Grewal and Serafeim (2020) define corporate sustainability as an intentional strategy to create long-term value through improved measurable social and environmental impact.

According to this view, which is also referred to as “doing well by doing good,” corporate sustainability is a long-term strategy that leads to long-run shareholder value maximization, even though it may come at a short-run cost (e.g., Bénabou and Tirole 2010; Eccles et al. 2014; Lins et al. 2017; Starks et al. 2020).¹⁹ Consistent with this view, Friede et al. (2015) document that 90% of academic studies find a nonnegative relationship between ESG and financial performance. In particular, Dowell et al. (2000) and Konar and Cohen (2001) find a positive relation between the environmental component of ESG and firm value. As described in Section 2.4.2 below, this view is not universally held; there is disagreement about the benefits of corporate sustainability.²⁰

We extend the growing sustainability/ESG literature by examining whether corporate environmental sustainability (E-S) attenuates the potential negative effect of managerial short-termism on pollution. Finding evidence consistent with this proposition reinforces the view that commitment to and investment in E-S curbs managerial short-termism.

2.4 Research hypotheses

2.4.1 Meeting benchmarks and toxic releases

The above review of institutions and literature suggests that managers trade-off the short-term benefits of meeting earnings benchmarks with the potential long-term benefits of reduced pollution and improved E-S.²¹ Cutting pollution abatement activities such as recycling, energy recovery, and treating pollutants reduces the costs of goods sold and increases earnings, and thereby increases the chance of meeting earnings benchmarks. Not meeting those benchmarks could have an immediate, large negative impact on share price (Skinner and Sloan 2002). While polluting to beat an earnings benchmark provides a short-term benefit for managers, the increased pollution imposes a long-term cost on the environment and society and can negatively affect the

¹⁹ For example, Bill McNabb, former CEO of Vanguard, argues that “by taking a broader, more complete view of corporate purpose, boards can focus on creating long-term value, better serving everyone—investors, employees, communities, suppliers and customers.” Potential channels through which ESG activities increase long-term value include avoiding value-destroying myopic decisions (Bénabou and Tirole 2010), strengthening product market position (Baron 2001), attracting customers (Navarro 1988) and employees (Greening and Turban 2000), and decreasing firm risk (Hoepner et al. 2019).

²⁰ Two other views of E-S, which suggest that it *reduces* long-term shareholder value, are that managers in liberal states engage in E-S to connect with liberal stakeholders (Di Giuli and Kostovetsky 2014) or do so for private benefits (Ferrell et al. 2016).

²¹ The underlying tension is akin to that in agency conflicts with shareholders, causing managers to emphasize short-term over long-term financial performance (e.g., Graham et al. 2005).

firm's E ratings and reputation, possibly even leading to environmental enforcement and litigation.²²

As managers tend to have shorter horizons than regulators and important stakeholders (Bénabou and Tirole 2010), we expect them to overemphasize the benefits of meeting short-term earnings benchmarks. That is, we expect higher toxic releases when managers cut pollution abatement costs to boost earnings in an effort to meet those benchmarks. This expectation implies the following testable prediction:

Hypothesis 1: For firms that release TRI toxins, meeting earnings benchmarks is positively associated with toxic releases.

Several factors may prevent us from observing a significant positive relation between meeting benchmarks and toxic releases. The long-term reputational or regulatory costs of increased pollution may be large enough to overwhelm the short-term benefits of earnings management. Consider, for example, the evidence in Liu et al. (2021) for non-SOE Chinese firms: the cost of SO₂ emissions appears to be prohibitive. Even in cases where the costs of increased pollution are lower, they might still exceed the costs of alternative ways to meet earnings benchmarks, such as guiding down analysts' earnings forecasts or engaging in other forms of earnings management, both real and accruals-based (Dechow et al. 2010).

2.4.2 The moderating effect of environmental sustainability

Observing a positive relation between pollution and meeting earnings benchmarks, as predicted by H1, raises a follow-on question: Is that relation moderated by a firm's commitment to E-S? The long-term shareholder view of E-S (reviewed in Section 2.3) predicts a weaker link between meeting benchmarks and toxic releases for firms that are more environmentally sustainable. This is because firms with high E ratings commit to a long-term strategy to reduce their environmental impact. They are thus less likely to increase pollution to meet short-term earnings benchmarks, which leads to the following hypothesis:

Hypothesis 2: If there is a positive association between meeting earnings benchmarks and toxic release, that association is weaker among firms with higher environmental ratings.

This prediction too is not without tension. Specifically, some practitioners and academics suggest that firms engage in "greenwashing" and exaggerate their environmental sustainability efforts (Bénabou and Tirole 2010). If so, E ratings are unreliable, as they primarily reflect voluntary disclosures of environmental policies (Lopez-de-Silanes et al. 2020), not outcomes. Consistent with greenwashing, Raghunandan and Rajgopal (2021) show that Business Roundtable signatories that claim to be socially responsible are in fact more likely to violate environmental and labor regulations. To the extent that firms engaging in greenwashing are more likely to emphasize short-term

²² See EPA enforcement action against Valero Energy Corporation (<https://www.epa.gov/enforcement/valero-energy-corporation-et-al-clean-air-act-settlement#injunctive>).

goals, we expect such firms to pollute more in general (despite their high E ratings) and to pollute even more when meeting earnings benchmarks.²³

3 Data, variable measurement, and sample

3.1 Data and variable measurement

We collect plant-level data on toxic releases for different chemicals from the EPA's TRI database. Our dependent variable is the total weight of toxic release (*Toxic Release*) (e.g., Akey and Appel 2021; Kim and Xu 2022). We first calculate *Toxic Release* at the plant level as the sum of the amount (in thousands of pounds) of all toxic chemicals released to air, water, and ground by a plant each year (Appendix B contains variable definitions). To calculate *Toxic Release* at the firm level, we aggregate *Toxic Release* across all plants owned by that firm. We assign an equal weight to each chemical for our primary measure of *Toxic Release*. Given that chemicals differ in toxicity, we also compute a toxicity-weighted sum of chemical releases (*Toxicity-weighted Release*). To do so, we use the toxicity index from the EPA's RSEI program, which describes each chemical's toxicity relative to other TRI-reported chemicals in terms of their chronic human health effects.²⁴ Following RSEI, we calculate the product of the release amounts (in thousands of pounds) and the toxicity index for each chemical, then aggregate it across all chemicals released by a plant-year or firm-year to obtain *Toxicity-weighted Release*.²⁵

Following recent studies of real earnings management (Caskey and Ozel 2017; Raghunandan 2021), we measure our primary explanatory variable—managers' efforts to meet earnings benchmarks—as the realized incidence of meeting or just beating consensus analyst EPS forecasts. We define firm-years as suspected of managing earnings to meet benchmarks (*Suspect*) if the actual EPS reported by I/B/E/S minus the latest consensus forecast is between zero and two cents.²⁶ We construct the latest consensus forecast as the average of all analysts' most recent forecasts issued within the [−90, −4] day window prior to the earnings announcement date. Using more recent forecasts reduces the distortion due to stale forecasts that are unlikely to be included in

²³ One unique prediction of the greenwashing view is that the moderating effect of E ratings should be driven by indicators derived from environmental inputs (such as policies and programs), which are easier to manipulate than indicators derived from outcomes (such as past toxic releases) (Grewal and Serafeim 2020).

²⁴ The RSEI toxicity index is based solely on human health effects, both cancer and noncancer (such as reproductive, developmental, and respiratory), associated with *long-term* exposure to chemicals. It ranges from 0.02 to 1,400,000,000. The level of toxicity index has no meaning by itself, as it describes each chemical's toxicity relative to other TRI-reported chemicals. Chemicals with a missing toxicity index are dropped from the calculation of *Toxicity-weighted Release*.

²⁵ RSEI may report more than one toxicity index for a chemical. We follow the RSEI's approach of selecting a toxicity index for calculating RSEI hazard (<https://www.epa.gov/rsei/rsei-toxicity-data-and-calculations>).

²⁶ We focus on meeting annual EPS forecasts because toxic release is reported annually. The use of *Suspect* to measure incentives to meet benchmarks relies on two implicit assumptions. First, more *Suspect* firms managed earnings upward than downward. This assumption is justified by the shape of discontinuity of the forecast error distribution around 0 (e.g., Degeorge et al. 1999; Cheong and Thomas 2018). Second, the final consensus is an ex post proxy for the number that managers during the year expected the final benchmark to be (e.g., Roychowdhury 2006). This assumption is required because managers manipulate real activities during the year, based on their expectation of the final consensus forecast.

the targets relevant to managers. To assess the robustness of our findings, we consider alternative measures of consensus forecasts, including the average of all analysts' most recent forecasts issued within a longer $[-180, -4]$ window prior to the earnings announcement date, as well as the most recent consensus forecast (taken from the I/B/E/S summary files) before the earnings announcement date. We collect firm-level accounting data from Compustat and stock market data from CRSP.

We follow the sustainability/ESG literature and measure E-S with E ratings (Gillan et al. 2021). We use data from MSCI ESG KLD STATS to construct E ratings of firms' ongoing efforts to reduce their environmental impact. This database relies on analysts that assess firms on a wide array of performance indicators pertaining to inputs and outcomes on environmental, social, and governance factors (Chatterji et al. 2009; Cao et al. 2021). MSCI scans public databases for ESG information and updates performance indicator values each year (Krüger 2015). It classifies these performance indicators into seven broad categories: environment, community, diversity, employee relations, human rights, product, and corporate governance. Within each category, MSCI further divides the performance indicators into strengths and concerns, which roughly correspond to inputs and outcomes, respectively.

Krüger (2015) shows that the ESG ratings from MSCI are highly positively autocorrelated over time. For example, an environmental accident generates an environmental concern that remains for several years thereafter. This stickiness implies that ESG ratings reflect the "stock" of a firm's environmental and societal attributes, which suggests that the ratings proxy well for firms' strategy on sustainability.

Regarding coverage, MSCI contains yearly ESG ratings for 650 to 3,000 of the largest U.S. companies and has been widely used in the literature (e.g., Lins et al. 2017). It has two major advantages relative to other ESG databases such as Refinitiv (formerly ASSET4): it covers more U.S. firms over our sample period, and it provides information about outcomes on specific issues in a standardized format (rather than merely indicating the presence or absence of disclosure) (Khan et al. 2016).

To build our primary E rating, we use all indicators from MSCI's *environment* category to calculate a net environmental score (*E Score (Net)*), equal to the number of E strengths minus the number of E concerns. We adjust *E Score (Net)* in two ways to make it more relevant for tests of Hypothesis 2. First, the total number of E indicators rated changes over time, making *E Score (Net)* less comparable over time. To reduce this source of measurement error, we refine *E Score (Net)* so that it is based only on E indicators with sufficiently long time-series data (ten or more years). We provide the list of these E indicators, along with their definitions, in Table IA.1 of the Internet Appendix.²⁷

Second, MSCI attempts to cover a comprehensive set of environmental issues, not all of which are directly related to toxic releases generated by production processes.

²⁷ See https://drive.google.com/file/d/117DRFeKgc4ICkuvAb_a23Md5yw1wEWoO/view?usp=sharing. Another solution to address changes over time in rating scope is to scale the number of strengths (concerns) by the maximum number of strengths (concerns) possible for that category in that year (e.g., Di Giuli and Kostovetsky 2014; Lins et al. 2017). We show that our inferences for *E Score (Net)* remain unchanged using this alternative solution in Table IA.5 of the Internet Appendix. One benefit of using E indicators with long time-series is that it facilitates the decomposition of *E Score (Net)* along two dimensions: partitions directly related and not directly related to toxic releases generated by production processes, and partitions corresponding to strengths and concerns.

Given our focus on releases generated by production processes, when testing Hypothesis 2 we create two partitions: those related to and those unrelated to toxic releases generated by production processes. The first group includes pollution prevention (Env-Str-B), recycling (Env-Str-C), hazardous waste (Env-Con-A), and regulatory compliance (Env-Con-B). The second group includes manufacturer of beneficial products and services (Env-Str-A), communications (Env-Str-E), and manufacturer of agricultural chemicals (Env-Con-E). We calculate a net environmental score—*TR-E Score (Net)* and *NTR-E Score (Net)* for the first and second partition, respectively—equal to the number of strengths minus the number of concerns (see Appendix B for details). We expect *TR-E Score (Net)* to be more relevant for testing Hypothesis 2, whereas *NTR-E Score (Net)* serves as a placebo used in falsification tests.

3.2 Sample selection and descriptive statistics

We first collect all plants reporting to TRI that can be linked to firms in the Compustat/CRSP merged database over the 1994 to 2018 period. We use the mapping table from Harvard Dataverse shared by Xiong and Png (2019) to link a plant in TRI to the firm in Compustat/CRSP that owns the plant. We restrict our sample to firm-years after 1994 because of the documented inconsistency between actual and forecasted EPS in I/B/E/S prior to 1994 (e.g., Clement and Tse 2003; Cohen et al. 2007; Kirk et al. 2014). This restriction does not reduce sample size much, as we find little overlap between TRI and Compustat/CRSP before 1994.

Following Akey and Appel (2021) and Kim and Xu (2022), we drop plant-year observations reporting zero values for *Toxic Release*.²⁸ We restrict our sample to firms that have December fiscal year ends to align financial data with TRI data, which are reported by calendar year. We also require observations in our sample to have non-missing values for all control variables used in our regression (1) discussed below. After imposing these data requirements, our final sample includes 39,090 plant-year observations for 5,149 firm-years, representing 4,518 plants for 559 unique firms. We winsorize all non-indicator variables at the 1st and 99th percentiles of their pooled distributions.

Table 1, Panel A describes the industry composition of our sample at the plant-year level, sorted by the number of observations. Consistent with our anecdotal understanding of across-industry variation in pollution levels, chemical & allied products, transportation equipment, and industrial machinery & equipment have the largest number of plant-years discharging toxic chemicals. They are followed by primary metal industries, paper & allied products, petroleum & coal products, fabricated metal products, and electronic & other electrical equipment.

Panels B and C in Table 1 provide summary statistics for plant- and firm-level variables. Each year, the average plant releases 119.84 thousand pounds of toxic chemicals (*Toxic Release*) and generates 0.10 thousand pounds of nonroutine toxic release (*Nonroutine Release*), with the latter generally resulting from catastrophic events or accidents. We delay to later sections the discussion of plant-level variables

²⁸ Zero values in TRI database are not meaningful as TRI inserts zeros into blank fields. (https://19january2021snapshot.epa.gov/sites/static/files/2019-08/documents/basic_data_files_documentation_aug_2019_v2.pdf).

Table 1 Summary statistics**Panel A: Industry membership (sorted by decreasing membership)**

2-digit SIC Code	Industry Description	# of obs.	Percentage (%)
28	Chemical & Allied Products	5,749	14.71%
37	Transportation Equipment	4,877	12.48%
35	Industrial Machinery & Equipment	4,233	10.83%
33	Primary Metal Industries	3,192	8.17%
26	Paper & Allied Products	3,112	7.96%
29	Petroleum & Coal Products	3,018	7.72%
34	Fabricated Metal Products	2,833	7.25%
36	Electronic & Other Electric Equipment	2,062	5.28%
32	Stone, Clay, & Glass Products	1,597	4.09%
14	Nonmetallic Minerals, Except Fuels	1,404	3.59%
38	Instruments & Related Products	1,327	3.39%
20	Food & Kindred Products	983	2.51%
24	Lumber & Wood Products	872	2.23%
13	Oil & Gas Extraction	713	1.82%
25	Furniture & Fixtures	626	1.60%
	Other industries contributing less than 1.50%	2,492	6.38%
Total		39,090	100.00%

Panel B: Plant-level variables

Variables	Mean	STD	P25	P50	P75	N
Toxic Release (lbs. 000)	119.84	399.29	0.09	2.77	33.15	39,090
Nonroutine Release (lbs. 000)	0.10	0.68	0.00	0.00	0.00	39,090
Past Enforcement Count	0.02	0.14	0.00	0.00	0.00	39,090
Past Settlement Amount (\$000)	0.24	2.13	0.00	0.00	0.00	39,090
Prod Ratio	1.02	0.60	0.87	1.00	1.11	39,090

Panel C: Firm-level variables

Variables	Mean	STD	P25	P50	P75	N
Toxic Release (lbs. 000)	1,097.27	3,812.35	3.15	43.10	401.18	5,149
Nonroutine Release (lbs. 000)	3.20	15.70	0.00	0.00	0.00	5,149
Suspect	0.12	0.33	0.00	0.00	0.00	5,149
Assets	10,557.56	26,373.88	899.90	2,629.10	7,740.00	5,149
Leverage	0.27	0.16	0.16	0.25	0.36	5,149
PPE/Assets	0.32	0.18	0.18	0.29	0.42	5,149
Turnover	1.17	0.64	0.78	1.04	1.39	5,149
CapEx/Assets	0.05	0.04	0.03	0.04	0.07	5,149
M/B	1.46	0.93	0.86	1.22	1.77	5,149

Table 1 (continued)

<i>E Score (Net)</i>	0.24	1.14	0.00	0.00	1.00	3,710
<i>S&G Score (Net)</i>	0.02	2.39	-1.00	0.00	1.00	3,710
<i>TR-E Score (Net)</i>	0.02	0.80	0.00	0.00	0.00	3,710
<i>NTR-E Score (Net)</i>	0.21	0.64	0.00	0.00	1.00	3,710
<i>TR-E Score (Strength)</i>	0.34	0.67	0.00	0.00	0.00	3,710
<i>TR-E Score (Concern)</i>	0.32	0.62	0.00	0.00	0.00	3,710
<i>NTR-E Score (Strength)</i>	0.32	0.53	0.00	0.00	1.00	3,710
<i>NTR-E Score (Concern)</i>	0.10	0.34	0.00	0.00	0.00	3,710
<i>Past Enforcement Count</i>	0.21	0.68	0.00	0.00	0.00	5,149
<i>Past Settlement Amount (\$000)</i>	19.59	116.89	0.00	0.00	0.00	5,149
<i>E-S IO</i>	0.15	0.12	0.05	0.11	0.21	5,035
<i>S&G-S IO</i>	0.15	0.10	0.07	0.13	0.20	5,035
<i>Production</i>	0.85	0.60	0.48	0.71	1.04	5,149
<i>SG&A/Sales</i>	0.16	0.12	0.08	0.14	0.22	5,149
<i>R&D/Sales</i>	0.03	0.05	0.00	0.01	0.03	5,149
<i>AD/Sales</i>	0.01	0.01	0.00	0.00	0.00	5,149
<i>COGS/Sales</i>	0.68	0.16	0.60	0.71	0.79	5,149

This table provides summary statistics for our sample. Panel A presents industry composition for our sample. Panel B provides summary statistics for plant-level variables from EPA data. Panel C provides summary statistics for firm-level variables from EPA, I/B/E/S, and Compustat data. All continuous variables are winsorized at 1% and 99% of the pooled distributions. Variable definitions are in Appendix B

capturing enforcement track records (*Past Enforcement Count* and *Past Settlement Amount*) and production levels (*Prod Ratio*). At the firm level, around 12% of observations in our sample are classified as firms meeting benchmarks based on *Suspect*. The average *E Score (Net)* based on E indicators with at least ten years of MSCI coverage over our sample period is 0.24.²⁹ The distributions of firm-level control variables (discussed next in Section 4) are comparable to those reported in Caskey and Ozel (2017).

4 Relation between toxic release and meeting earnings benchmarks

4.1 Baseline results for hypothesis 1

To test the relation between meeting earnings benchmarks and toxic releases outlined in Hypothesis 1, we estimate the following regression at both the firm and plant levels.

$$\begin{aligned} \ln(1 + \text{Toxic Release}) = & \text{Suspect} + \text{Controls} + \text{Firm or Plant FE} \\ & + \text{Industry-by-year FE} + \varepsilon \end{aligned} \quad (1)$$

²⁹ The percentages of non-zero observations for four key E rating measures—*E Score (Net)*, *TR-E Score (Net)*, *TR-E Score (Strength)*, and *TR-E Score (Concern)*—are 48%, 35%, 25%, and 24%, respectively. These relatively higher fractions provide comfort that our results are not due to a handful of outliers.

Given that *Suspect* is a firm-level reporting decision, we focus on firm-level analyses for Eq. (1). We also conduct plant-level analyses to confirm that the firm-level results are not due to a few plants with large toxic releases. Finding similar results across plants suggests that the decision to boost earnings is taken at the firm level and then communicated to plants. Plant-level analyses also allow us to examine where pollution occurs within a firm, which enables investigation of predictions about across-plant variation in pollution.³⁰ As *Toxic Release* is right-skewed, we follow prior studies (e.g., Akey and Appel 2021; Chatterji et al. 2009; Kim and Xu 2022) and use the natural logarithm of one plus *Toxic Release* as the dependent variable in regression (1).³¹ We also follow prior studies and control for the normal level of toxic release by including various firm characteristics that reflect normal production levels. These include the natural logarithm of firm size measured by total assets (*Assets*); natural logarithm of the ratio of total revenue to total assets (*Turnover*); property, plant, and equipment scaled by total assets (*PPE/Assets*); capital expenditures scaled by total assets (*CapEx/Assets*); market-to-book ratio (*M/B*); and book leverage (*Leverage*). We also include nonroutine release (*Nonroutine Release*) to control for the effects of one-time pollution events.

Following prior studies (e.g., Greenstone 2002; Akey and Appel 2021; Tomar 2019), we include firm (plant) fixed effects and industry-by-year fixed effects in firm-level (plant-level) regressions. The firm/plant fixed effects purge the estimated effects of meeting benchmarks of all persistent firm/plant characteristics that determine toxic releases. The industry-by-year fixed effects remove all transitory differences in toxic releases across industries. As Greenstone (2002) explains, these fixed effects are important because macroeconomic changes over time (e.g., oil crises, recessions, and increases in foreign competition) have different effects on different manufacturing industries. While TRI provides data on 334 toxic chemicals, the mean (median) number of chemicals reported is 20.9 (13) at the firm level and 7.5 (5) at the plant level. We cluster all standard errors at the firm level and present significance for two-tailed tests.

Panel A of Table 2 presents firm and plant-level estimates of regression (1) in columns (1) and (2), respectively. As we include firm (plant) fixed effects, the coefficients on *Suspect* capture incremental toxic releases for suspect firm-years relative to non-suspect firm-years for the same firm (plant). Consistent with Hypothesis 1, we find significantly higher toxic releases for *Suspect* firm-years in both columns of Table 2, Panel A. The magnitude of the coefficient on *Suspect* is larger in the firm-level regression than in the plant-level regression because firm-level pollution combines pollution across plants for each firm.

In terms of economic magnitude, the estimates from the firm-level regression reported in column (1) suggest that real earnings management to meet benchmarks leads to about a 15% increase in toxic release relative to the mean (median) level of

³⁰ In Table 6 we show that firms are less likely to increase pollution in suspect years from plants for which increased pollution is more costly—high polluters or plants faced with a prior enforcement action. In Table 9, we use plant-level production data to confirm that suspect firm-years are not associated with overproduction.

³¹ Chatterji et al. (2009) recommend adding a constant before taking the natural logarithm to avoid overemphasizing changes from small initial values. In untabulated results, we confirm that *Suspect* remains significant in regression (1) when the dependent variable is $\ln(\text{Toxic Release})$. Specifically, the coefficient estimate for *Suspect* is 0.222 (0.078) with a *t*-statistic of 3.12 (1.98) for the firm (plant) level regressions.

Table 2 The effect of meeting earnings benchmarks on toxic release

Panel A: All TRI pollutants				
	Firm-level		Plant-level	
	(1)		(2)	
<i>Suspect</i>	0.137***		0.065***	
	(3.19)		(3.47)	
<i>Ln(Assets)</i>	0.532***		0.134**	
	(4.74)		(2.24)	
<i>Leverage</i>	0.189		0.403***	
	(0.65)		(2.68)	
<i>PPE/Assets</i>	0.688		0.250	
	(1.42)		(1.03)	
<i>Ln(Turnover)</i>	0.436***		0.141**	
	(3.16)		(2.20)	
<i>CapEx/Assets</i>	-1.640		-0.963**	
	(-1.33)		(-2.05)	
<i>M/B</i>	0.007		0.035	
	(0.13)		(1.16)	
<i>Nonroutine Release</i>	0.011		0.121***	
	(0.45)		(2.84)	
<i>Firm or Plant FE</i>	Firm		Plant	
<i>Industry-Year FE</i>	YES		YES	
<i>Observations</i>	4,881		38,442	
<i>Adjusted R-squared</i>	0.910		0.892	

Panel B: Separately for criteria air pollutants (ozone and lead) and all other TRI pollutants				
	Firm-level		Plant-level	
	(1)	(2)	(3)	(4)
Criteria air pollutants	Yes	No	Yes	No
<i>Suspect</i>	-0.013	0.121***	-0.009	0.066***
	(-0.44)	(2.83)	(-0.76)	(3.39)
<i>Controls</i>	YES	YES	YES	YES
<i>Firm or Plant FE</i>	Firm	Firm	Plant	Plant
<i>Industry-Year FE</i>	YES	YES	YES	YES
<i>Observations</i>	1,374	4,792	4,998	36,538
<i>Adjusted R-squared</i>	0.860	0.908	0.945	0.890

The dependent variable is $\ln(1 + \text{Toxic Release})$, and regressions are estimated for firm- and plant-year observations. The independent variable *Suspect* is an indicator variable equal to one if a firm meets or beats the latest annual consensus analyst EPS forecast by two cents or less. Panel A presents the results for all TRI pollutants combined, and Panel B provides results separately for two criteria air pollutants also covered under TRI (ozone and lead) and all other TRI pollutants. *t*-statistics are presented in parentheses with robust standard errors clustered at the firm level. All variables are defined in Appendix B. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively

Toxic Release reported in Table 1, Panel C.³² In addition to confirming H1, this evidence is consistent with our rationale for H1: pollution increases when managers cut pollution abatement costs to boost earnings in order to meet earnings benchmarks.³³ Regarding control variables, the coefficients on *Assets* and *Turnover* are positive and significant, consistent with both variables capturing total assets and sales (two measures of scale). *Nonroutine Release* is also positively associated with *Toxic Release* at the plant level, suggesting that pollution due to one-time events should be controlled for.

As discussed in Section 2.1.1, two TRI toxins—ozone and lead—are subject to real time monitoring because the EPA includes them among the six criteria air pollutants. When meeting earnings benchmarks, firms are less likely to increase releases of these two toxins, relative to other TRI toxins, because the costs of increased pollution are considerably higher. To shed light on this issue, we re-estimate regression (1) separately for released amounts of ozone and lead. Consistent with the costs of pollution being higher for ozone and lead, the estimates for *Suspect* in columns (1) and (3) of Table 2, Panel B are insignificant.³⁴ In contrast, *Suspect* continues to be significantly positively associated with the released amounts of other TRI toxins, as shown in columns (2) and (4) of Table 2, Panel B. The different results observed for ozone and lead confirm that the costs of increased pollution are considerably higher for those two toxins—sufficiently high to deter managers from cutting back on their pollution control costs to meet earnings goals.

4.2 Robustness of baseline results

In this section, we examine if the main result in Table 2—our finding that toxic releases increase during *Suspect* firm-years—is robust to alternative measures of toxic release and *Suspect*.

4.2.1 Alternative measures of toxic release

We examine the effect of meeting benchmarks for two alternative measures of toxic release. The first measure is toxicity-weighted chemical release (*Toxicity-Weighted Release*), which accounts for variation in the chronic health effects of chemicals reported to TRI. The second measure is *Toxic Release* scaled by revenues (*Toxic Release/Sales*), which allows for potential non-linearity in the relation between toxic release and production levels.³⁵ We replace *Toxic Release* in regression (1) with

³² Assuming a baseline level of *Toxic Release* equal to the firm-level mean (median) value of 1303.89 (43.10), the coefficient 0.137 on *Suspect* implies an increase of toxic release by 194.22 (6.60), which equals $\exp\{\ln[1 + 1303.89] + 0.137\} - 1 - 1303.89 (\exp\{\ln[1 + 43.10] + 0.137\} - 1 - 43.10)$ or roughly 14.7% (15.0%) of the mean (median) value.

³³ In Table IA.2 of the Internet Appendix, we confirm that the positive association between *Suspect* and *Toxic Release* remains robust after dropping potential outliers based on Cook's distance and studentized residuals.

³⁴ The sample sizes for these regressions are smaller, as many plants/firms do not release ozone or lead. It is unlikely, however, that lead and ozone are in fact similar to other toxins, but the small sample sizes cause us to infer otherwise. Not only do we observe a negative, not positive, sign on *Suspect* in both columns, the slopes on the control variables *Assets*, *Leverage*, and *Turnover* are significant (untabulated).

³⁵ In Regression (1) we regress undeflated toxic releases on *Suspect* and controls for scale because that specification is preferred to deflating by scale (Lev and Sunder 1979). Deflating toxic releases by scale and retaining controls for scale allow for a nonlinear relation between toxic release and scale.

Toxicity-Weighted Release and *Toxic Release/Sales* and report the regression results in Table 3, Panel A.

We continue to find a significant positive correlation between *Suspect* and *Toxicity-Weighted Release*, suggesting that the effect of meeting benchmarks documented in Table 2 is not mainly driven by chemicals that are relatively benign. Similarly, *Suspect* and *Toxic Release/Sales* are also significantly positively associated, suggesting that our finding is unlikely to be due to insufficient control for scale. To investigate the potential concern that our results so far are driven by a few chemicals in a handful of industries, we repeat the analyses separately for a subset of 20 chemicals with the largest release quantities in our sample, as well as for the remaining chemicals that are not as prevalent. Our results in Table IA.3 of the Internet Appendix indicate significant positive slopes on *Suspect* for both subgroups.

We also examine potential variation in the effect of meeting benchmarks on toxic release across pollution media: air emissions, water discharge, and solid waste disposal. Table IA.4 of the Internet Appendix shows that the effect of meeting earnings benchmarks is concentrated in air emissions and is absent for water discharge and solid waste disposal. One potential explanation for this finding is that toxic air emissions may be less salient (hard to trace and not as visible) than toxic water discharges and ground disposal.

4.2.2 Alternative measures of consensus analyst forecasts used to construct suspect

We next consider different measures of consensus forecasts. In Table 3, Panel B, we redefine *Suspect* using the average of each analyst's latest forecast issued within a wider $[-180, -4]$ day window prior to the earnings announcement date (*Suspect (180)*). Like *Suspect*, which is based on forecasts issued within the $[-90, -4]$ window, *Suspect (180)* is significantly positively associated with *Toxic Release*.

As discussed earlier, one assumption motivating the use of *Suspect* to capture earnings management is that managers manipulate real activities during the year based on their expectations of both performance during the year and the final analyst consensus against which actual EPS is compared (benchmark). Managers likely ignore consensus forecasts available earlier in the year because the forecasts are generally biased upward and are based on less information than is available to managers. In effect, those earlier consensus forecasts measure managers' expectations of the final consensus with error, and that measurement error is larger earlier in the year. When consensus forecasts are used to compute *Suspect*, the relation with toxic release should therefore become successively weaker with earlier forecast dates.

To test the above prediction, we collect the consensus forecast of *annual* EPS (constructed by I/B/E/S in its summary files) at the end of each fiscal quarter as well as just before the earnings announcement, and redefine *Suspect* relative to each of these consensus forecasts. Consistent with earlier forecasts being a noisier proxy for the final consensus managers expect during the year, the results in Table 3, Panel C show that the relation between *Suspect* and *Toxic Release* weakens monotonically as the date of the consensus moves back in time from the month right before earnings announcement to the end of the first fiscal quarter. The relation is equally strong and significant for the month prior to the earnings announcement and the end of Q4, but declines in magnitude and significance for consensus forecasts available as of the end of Q3, Q2, and Q1.

Table 3 Robustness analyses: effect of meeting earnings benchmarks on toxic release

Panel A: Effect of meeting benchmarks on toxicity-weighted and sales-deflated toxic release

	Firm-level		Plant-level	
	Toxicity-weighted	Sales-deflated	Toxicity-weighted	Sales-deflated
	(1)	(2)	(3)	(4)
<i>Suspect</i>	0.181*** (2.82)	0.052*** (3.29)	0.091*** (2.78)	0.005*** (3.85)
Controls	YES	YES	YES	YES
<i>Firm FE</i>	YES	YES		
<i>Plant FE</i>			YES	YES
<i>Industry-Year FE</i>	YES	YES	YES	YES
<i>Observations</i>	4,881	4,881	37,735	38,442
<i>Adj. R-squared</i>	0.883	0.798	0.864	0.794

Panel B: Alternate *Suspect* measure based on a wider window of forecasts.

	Firm-level	Plant-level
	(1)	(2)
<i>Suspect (180)</i>	0.136*** (3.07)	0.058*** (2.99)
Controls	YES	YES
<i>Firm FE</i>	YES	
<i>Plant FE</i>		YES
<i>Industry-Year FE</i>	YES	YES
<i>Observations</i>	4,881	38,442
<i>Adjusted R-squared</i>	0.910	0.892

Panel C: Alternative *Suspect* measures based on IBES consensus forecast at different dates.

	Firm-level				
	I/B/E/S consensus from month				
	before EA (1)	at end of Q4 (2)	at end of Q3 (3)	at end of Q2 (4)	at end of Q1 (5)
<i>Suspect</i>	0.111** (2.48)	0.100* (1.95)	0.068 (1.27)	-0.003 (-0.04)	-0.028 (-0.31)
Controls	YES	YES	YES	YES	YES
<i>Firm FE</i>	YES	YES	YES	YES	YES
<i>Industry-Year FE</i>	YES	YES	YES	YES	YES
<i>Observations</i>	4,881	4,881	4,881	4,881	4,881
<i>Adjusted R-squared</i>	0.910	0.910	0.910	0.910	0.910
	Plant-level				
	I/B/E/S consensus from month				
	before EA (6)	at end of Q4 (7)	at end of Q3 (8)	at end of Q2 (9)	at end of Q1 (10)
<i>Suspect</i>	0.067*** (3.51)	0.067*** (3.47)	0.054** (2.19)	0.027 (0.71)	0.009 (0.27)

Table 3 (continued)

<i>Controls</i>	YES	YES	YES	YES	YES
<i>Plant FE</i>	YES	YES	YES	YES	YES
<i>Industry-Year FE</i>	YES	YES	YES	YES	YES
<i>Observations</i>	38,442	38,442	38,442	38,442	38,442
<i>Adjusted R-squared</i>	0.892	0.892	0.891	0.891	0.891

Panel D: Toxic release for different bands of forecast error

	Firm-level (1)	Plant-level (2)
<i>Large Miss</i>	0.040 (0.60)	−0.027 (−0.99)
<i>Suspect</i>	0.165*** (2.99)	0.048** (2.31)
<i>Beat 2-5cents</i>	0.083 (1.55)	0.013 (0.49)
<i>Large Beat</i>	−0.008 (−0.14)	−0.037 (−1.58)
<i>Controls</i>	YES	YES
<i>Firm FE</i>	YES	
<i>Plant FE</i>		YES
<i>Industry-Year FE</i>	YES	YES
	4,881	38,442
<i>Adjusted R-squared</i>	0.910	0.892

To investigate the robustness of the firm- and plant-year level results in Table 2, this table presents results for alternative measures of the dependent variable— $\ln(1 + \text{Toxic Release})$ —in Panel A, and alternate measures of the consensus annual EPS forecast underlying the independent variable—*Suspect*—in Panels B, C, and D. The two alternative measures of the dependent variable are toxicity-weighted release ($\ln(1 + \text{Toxicity-Weighted Release})$), which incorporates chemical-specific toxicity, and sales-deflated toxic release ($\text{Toxic Release}/\text{Sales}$). The alternate *Suspect* measure in Panel B, *Suspect* (180), is defined relative to the average of analysts' latest forecasts issued within [−180, −4] days relative to the earnings announcement date. In Panel C, the alternate measures of *Suspect* are based on consensus analyst forecasts constructed by I/B/E/S in its summary history files at different points of time: the last month before year-end earnings announcements (EA) and at the end of each fiscal quarter (Q1 to Q4). In Panel D, we report levels of toxic release for three forecast error bands not included in *Suspect*. *Large Miss/Beat 2-5cents /Large Beat* is an indicator variable equal to one if a firm misses the latest consensus forecast by more than 3 cents/beats it by more than two but less than five cents/beats it by at least five cents. *Controls* include $\ln(\text{Assets})$, *Leverage*, PPE/Assets , $\ln(\text{Turnover})$, *CapEx/Assets*, *M/B*, and *Nonroutine Release*. *t*-statistics are presented in parentheses with robust standard errors clustered at the firm level. Variables are defined in Appendix B. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively

Note that these results have no implications for the timing of the release of toxins, as the pollution data we use only provide the total amount released over the year.

Our final analysis in Table 3 examines the pattern of pollution levels across different levels of forecast error. Table 2 shows that pollution is higher if forecast errors lie between 0 and 2 cents, relative to the average for all other forecast errors combined. However, hypothesis H1 makes a sharper prediction: pollution levels should increase only for the narrow range of forecast errors corresponding to meet or just beat. The

results in Table 2 are also consistent with other patterns though. For example, pollution increases nonlinearly with performance: pollution levels for meet/just beat are considerably higher than those for negative forecast error but only slightly lower than those for large positive forecast errors.

We construct three other indicator variables based on forecast error relative to the latest consensus forecast used to measure *Suspect*: *Large Miss* equals one if the forecast error is more negative than -3 cents, as in Caskey and Ozel (2017); *Beat 2-5cents* equals one if the forecast error is more than 2 cents but no more than 5 cents; *Large Beat* equals one if the forecast error is more than 5 cents. We add *Large Miss*, *Beat 2-5cents*, and *Large Beat* as three additional explanatory variables to regression (1). In effect, these indicator variables, along with *Suspect*, describe pollution levels for the respective ranges relative to the level for a “just miss” base group with negative forecast error higher than -3 cents.

Results reported in Panel D of Table 3 show that only the coefficient on *Suspect* is significantly positive, and the coefficients on *Large Miss*, *Beat 2-5cents*, and *Large Beat* are all insignificant. While firms that meet or just beat consensus forecasts release more toxic chemicals than the base group that just misses the benchmark, firms that substantially miss the benchmark or comfortably beat the benchmark do not. The clear pattern observed in these results is inconsistent with the possibility that pollution levels increase with forecast error (better unexpected performance).

5 The moderating effect of environmental sustainability

To test Hypothesis 2 (H2) regarding the moderating effect of a firm’s E rating on the relation between meeting benchmarks and toxic releases, we estimate the following regression:

$$\begin{aligned} \ln(1 + \text{Toxic Release}) = & \text{Suspect} + \text{Suspect} \times E \text{ Score (Net)} + E \text{ Score (Net)} \\ & + \text{Controls} + \text{Firm FE} + \text{Industry-by-year FE} + \varepsilon \quad (2) \end{aligned}$$

where *E Score (Net)*, our primary measure of E rating, reflects the net effects of the strengths and concerns in a firm’s environmental indicators taken from MSCI. We estimate regression (2) on firm-year observations, as E indicators are only available at the firm level. H2 predicts a negative coefficient on the interaction term *Suspect* \times *E Score (Net)*.

Panel A of Table 4 presents estimates from regression (2). Results in Column (1) are for *E Score (Net)* based on all E indicators available, and the results in Column (3) are for *E Score (Net)* based on indicators with at least ten years of MSCI data. As discussed in Section 3.1, *E Score (Net)* in Column (3) is more comparable over time because it excludes E indicators with sparse coverage. Contrary to H2, we find a significantly positive coefficient on the interaction term *Suspect* \times *E Score (Net)* in both columns (1) and (3). This finding suggests that firms with higher E ratings—i.e., firms rated by MSCI as being more environmentally sustainable—increase pollution more, rather than less, than firms with low E ratings when meeting earnings benchmarks.

Table 4 The moderating effect of firm-level environmental ratings**Panel A: Moderating effects of net environmental score**

	All E items		E items with long time-series	
	(1)	(2)	(3)	(4)
<i>Suspect</i> × <i>E Score (Net)</i>	0.067** (2.12)	0.068** (2.10)	0.081** (2.19)	0.082** (2.16)
<i>E Score (Net)</i>	-0.053 (-1.53)	-0.049 (-1.49)	-0.038 (-1.11)	-0.034 (-1.01)
<i>Suspect</i> × <i>S&G Score (Net)</i>		-0.001 (-0.10)		-0.001 (-0.05)
<i>S&G Score (Net)</i>		-0.018 (-1.49)		-0.019 (-1.58)
<i>Suspect</i>	0.033 (0.78)	0.033 (0.77)	0.025 (0.56)	0.024 (0.54)
<i>Controls</i>	YES	YES	YES	YES
<i>Firm FE</i>	YES	YES	YES	YES
<i>Industry-Year FE</i>	YES	YES	YES	YES
<i>Observations</i>	3,464	3,464	3,464	3,464
<i>Adjusted R-squared</i>	0.929	0.929	0.929	0.929

Panel B: Moderating effects of components of net E score related to (TR) and not related to (NTR) toxic releases from production processes and the separate effect of strengths and concerns underlying those net scores

	E items with long time-series	
	(1)	(2)
<i>Suspect</i> × <i>TR-E Score (Net)</i>	0.115** (2.09)	
<i>Suspect</i> × <i>TR-E Score (Strength)</i>		0.092* (1.71)
<i>Suspect</i> × <i>TR-E Score (Concern)</i>		-0.185** (-2.19)
<i>TR-E Score (Net)</i>	-0.059 (-1.55)	
<i>TR-E Score (Strength)</i>		-0.088 (-1.52)
<i>TR-E Score (Concern)</i>		0.029 (0.55)
<i>Suspect</i> × <i>NTR-E Score (Net)</i>	0.035 (0.71)	
<i>Suspect</i> × <i>NTR-E Score (Strength)</i>		0.021 (0.35)
<i>Suspect</i> × <i>NTR-E Score (Concern)</i>		-0.149 (-0.90)
<i>NTR-E Score (Net)</i>	-0.005 (-0.08)	

Table 4 (continued)

<i>NTR-E Score (Strength)</i>		-0.004 (-0.06)
<i>NTR-E Score (Concern)</i>		-0.007 (-0.06)
<i>Suspect</i>	0.032 (0.71)	0.078 (1.51)
<i>Controls</i>	YES	YES
<i>Firm FE</i>	YES	YES
<i>Industry-Year FE</i>	YES	YES
<i>Observations</i>	3,464	3,464
<i>Adjusted R-squared</i>	0.929	0.929

The dependent variable is $\ln(1 + \text{Toxic Release})$. The independent variable of interest is the interaction term between *Suspect* and an environmental score based on E ratings (from MSCI-ESG KLD STATS). *Suspect* equals one if a firm meets or beats the latest annual consensus analyst forecast by two cents or less. In Panel A, environmental score (*E Score (Net)*) equals the number of E strengths minus the number of E concerns. In columns “All E items,” *E Score (Net)* is based on all E strength and concern indicators. In columns “E items with long time-series,” *E Score (Net)* is based on E strength and concern indicators with at least ten years’ data during our sample period. A higher value of *E Score (Net)* indicates that the firm has more environmentally friendly policies/programs and/or fewer events of negative environmental impact. As a placebo test, we also include the interaction between *Suspect* and *S&G Score (Net)*, which is defined as the total number of strengths minus the total number of concerns for the community, human rights, employee relations, diversity, product, and governance categories of MSCI ratings. In Panel B, we decompose *E Score (Net)* into scores for subcategories. We first decompose *E Score (Net)* from columns (3) and (4) of Panel A into *TR-E Score (Net)* and *NTR-E Score (Net)*, which are based on the subset of E indicators that are and are not directly related to toxic release generated by production processes, respectively. We then decompose *TR-E Score (Net)* into *TR-E Score (Strength)*, which is based on the E strength indicators, and *TR-E Score (Concern)*, which is based on the E concern indicators underlying *TR-E Score (Net)*. By construction, $\text{TR-E Score (Net)} = \text{TR-E Score (Strength)} - \text{TR-E Score (Concern)}$. We similarly decompose *NTR-E Score (Net)* into *NTR-E Score (Strength)* and *NTR-E Score (Concern)*. Control variables (*Controls*) include $\ln(\text{Assets})$, *Leverage*, *PPE/Assets*, $\ln(\text{Turnover})$, *CapEx/Assets*, *M/B*, and *Nonroutine Release*. *t*-statistics are presented in parentheses with robust standard errors clustered at the firm level. Variables are defined in Appendix B. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively

To further confirm the unexpected positive coefficient on $\text{Suspect} \times \text{E Score (Net)}$, we use the S&G components of ESG indicators as a placebo test. Given that the dependent variable is toxic releases, S&G ratings should have little incremental explanatory content over E ratings. We construct a corresponding social and governance net score (*S&G Score (Net)*) by subtracting the number of concerns for community, diversity, employee relations, human rights, product, and corporate governance categories from the number of strengths for these categories. We add *S&G Score (Net)* and $\text{Suspect} \times \text{S&G Score (Net)}$ as additional explanatory variables to regression (2) and report the results in columns (2) and (4) of Table 4, Panel A. If the coefficient on $\text{Suspect} \times \text{E Score (Net)}$ reflects some general features of ESG ratings that are unrelated to E-S, we expect to see similar results for *S&G Score (Net)*. If, however, the relation between pollution and meeting benchmarks is solely a function of E ratings, the coefficient on the interaction with *S&G Score (Net)* should not load. We find the

interaction between *Suspect* and *S&G Score (Net)* is indeed insignificant in columns (2) and (4).³⁶

To provide additional evidence on the role of E ratings in the positive relation between toxic release and meeting earnings benchmarks, we decompose *E Score (Net)* in columns (3) and (4) of Panel A into *TR-E Score (Net)* and *NTR-E Score (Net)*, reflecting factors that are and are not directly related, respectively, to toxic releases generated by production processes. We expect the moderating effect of *E Score (Net)* on the link between *Suspect* and *Toxic Release* to be driven by *TR-E Score (Net)*. The components of the E rating that are unrelated to production processes, *NTR-E Score (Net)*, should be less relevant for the toxic releases that we examine here. We re-estimate regression (2) by replacing *E Score (Net)* with *TR-E Score (Net)* and *NTR-E Score (Net)* and report the results in column (1) of Panel B. Consistent with our expectation, the interaction between *Suspect* and *TR-E Score (Net)* is significantly positive, but the interaction between *Suspect* and *NTR-E Score (Net)* is insignificant.

We confirm that this result for the net effect of factors that are directly related to toxic releases generated by production processes is also observed separately for strengths and concerns. We further decompose *TR-E Score (Net)* into *TR-E Score (Strength)* and *TR-E Score (Concern)*, which are defined as the number of strengths and concerns, respectively, among E indicators directly related to toxic releases generated by production processes. By definition, $TR-E\ Score\ (Net) = TR-E\ Score\ (Strength) - TR-E\ Score\ (Concern)$. The results reported in Column (2) of Table 4, Panel B suggest that both *TR-E Score (Concern)* and *TR-E Score (Strength)* have a significant moderating effect on the link between *Suspect* and *Toxic Release*.

We consider alternative explanations for our surprising result relating to H2. Why do firms with higher E ratings, especially for indicators directly related to toxic releases from production processes, pollute more when facing short-term earnings pressure? The explanation that is most consistent with results from the additional analyses we conduct is based on firms with high E ratings building regulatory and reputational “pollution slack,” which allows them to increase pollution at a lower cost when it is needed to boost earnings. This slack could result from more emission allowances (generated by lower toxic releases in the past), which allow firms to temporarily increase pollution without exceeding regulatory limits. The slack could also arise from a better reputation, built over time among stakeholders (e.g., consumers, communities, and investors), that excuses the firms for temporary increases in pollution.³⁷

One key assumption underlying our pollution slack explanation is that the MSCI data provide reliable E-S measures: higher E ratings reflect better environmental track records on toxic releases. To test this assumption, we construct the following measures of environmental track records using EPA data: toxic releases (*Toxic Release*) over the past two years; the number of EPA enforcement cases on plants owned by the firm over the past five years (*Past Enforcement Count*); and the dollar amount of settlements for all EPA enforcement cases on plants owned by the firm over the past five years (*Past*

³⁶ In untabulated results, we find that neither the S component nor the G component of *S&G Score (Net)* has a significant interaction effect with *Suspect*.

³⁷ The slack could also be psychological. For example, Bénabou and Tirole (2010) propose a “moral credentialing” argument where people who have recently “done good” feel immunized against negative (social or self) inferences and thus later act less morally constrained. Monin and Miller (2001) document experimental evidence consistent with this argument.

Settlement Amount). Higher values of these measures indicate worse environmental track records. We regress each of these measures on *TR-E Score (Strength)*, *TR-E Score (Concern)*, *NTR-E Score (Strength)*, and *NTR-E Score (Concern)* to determine if these four measures of E ratings reflect environmental track records.³⁸

Table 5 presents the regression results for environmental track records. Consistent with our assumption, *TR-E Score (Concern)* is significantly positively associated with all four track record measures. *TR-E Score (Strength)* is significantly negatively associated with *Toxic Release* in the past two years, but not significantly related to past enforcements and settlements. In contrast, *NTR-E Score (Strength)* and *NTR-E Score (Concern)* are generally unrelated to the track record measures. These findings indicate that the E ratings based on MSCI indicators directly related to toxic releases capture reasonably well the track records of firms' toxic releases. They also suggest that MSCI relies on EPA data to construct its E indicators.

As a final test of the pollution slack explanation, we examine within-industry-year and within-firm-year variation in plant-level measures of track records for toxic releases. Exploiting within-firm-year variation is a tighter specification that asks the following question: When firms decide to increase pollution to meet earnings benchmarks, are they more likely to increase pollution in plants with better toxic release track records (i.e., larger pollution slack)? We construct plant-level versions of *Past Enforcement Count* and *Past Settlement Amount* based on enforcement cases for each plant. We also construct a dummy variable, *High Past Pollution*, that equals one (zero) if a plant's toxic release in the previous year is higher (lower) than the sample median for that year. We modify regression (2) by replacing *E Score (Net)* with each of these three plant-level track record variables and estimate it at the plant level with plant fixed effects.

Panel A of Table 6 presents the results for *High Past Pollution*, and Panel B presents results for *Past Enforcement Count* and *Past Settlement Amount*. The left (right) column in each subgroup includes industry-year (firm-year) fixed effects. The coefficients on the interaction term are all negative and significant (at the 10% level or better). These results confirm that plants with better (worse) track records of toxic release, relative to other plants in that industry-year or firm-year, are more (less) likely to cut pollution abatement costs so that the firm can meet earnings benchmarks. The significant results observed for the tighter specification with firm-year fixed effects are consistent with the pollution slack explanation, because firms leverage where the slack resides when deciding to cut pollution abatement costs.³⁹

Our results suggest the following description: while high E rating firms pollute less in general than low E rating firms, they pollute more relative to their historical averages when it is needed to meet earnings benchmarks. This finding contradicts the view that environmental sustainability curbs managerial short-termism. It is also inconsistent with the greenwashing explanation, which holds that firms that obtain high E ratings by exaggerating their environmental efforts also pollute more when meeting earnings

³⁸ This specification allows easier comparison with results reported in Column (2) of Table 4, Panel B.

³⁹ While we use E ratings from MSCI because they offer a comprehensive set of indicators for both E and S&G ratings, we recognize that there are other sources and that ESG ratings vary across those sources (Berg et al. 2020; Gibson et al. 2021; Christensen et al. 2021). We are comforted that the MSCI E ratings are highly correlated with EPA data on firms' environmental track records.

Table 5 Environmental ratings and environmental impact track records

	<i>Ln(1 + Toxic Release)</i> , 1 year ago (1)	<i>Ln(1 + Toxic Release)</i> , 2 years ago (2)	<i>Past Enforcement Count</i> , prior 5 years (3)	<i>Ln(1 + Past Settlement Amount)</i> , prior 5 years (4)
<i>TR-E Score (Strength)</i>	-0.362*** (-2.72)	-0.284** (-2.16)	0.049 (0.78)	0.123 (0.85)
<i>TR-E Score (Concern)</i>	0.711*** (3.64)	0.718*** (3.71)	0.132* (1.74)	0.493** (2.53)
<i>NTR-E Score (Strength)</i>	-0.045 (-0.23)	-0.021 (-0.11)	-0.057 (-1.50)	-0.210 (-1.32)
<i>NTR-E Score (Concern)</i>	0.454 (1.62)	0.493* (1.78)	0.015 (0.15)	0.122 (0.41)
<i>Ln(Assets)</i>	0.736*** (8.51)	0.755*** (8.97)	0.061*** (3.11)	0.222*** (3.07)
<i>Industry-Year FE</i>	YES	YES	YES	YES
<i>Observations</i>	3,347	3,280	3,505	3,505
<i>Adjusted R-squared</i>	0.413	0.394	0.140	0.128

This table presents results for the relation between a firm's environmental (E) rating from MSCI ESG KLD STATS and its environmental impact track record. We summarize E ratings with the strength and concern scores—*TR-E Score (Strength)* and *TR-E Score (Concern)*—for strength and concern indicators directly related to toxic release from production processes. We also prepare two corresponding scores—*NTR-E Score (Strength)* and *NTR-E Score (Concern)*—for strength (concern) indicators not directly related to toxic release from production processes. We only consider E indicators with at least ten years' data. We measure environmental track records with toxic release (*Toxic Release*) over the prior two years, the number of EPA enforcement cases (*Past Enforcement Count*) over the prior five years, and the dollar amount of penalties settled in EPA enforcement cases (*Past Settlement Amount*) in which plants of the firm are defendants over the prior five years. *t*-statistics are presented in parentheses with robust standard errors clustered at the firm level. Variables are defined in Appendix B. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively

benchmarks. Under the greenwashing explanation, firms with high E ratings should pollute more in general, as indicated by worse environmental track records.⁴⁰ Our results are consistent, however, with a more nuanced view: superior environmental track records create pollution slack that reduces the costs of increased pollution to meet

⁴⁰ The results reported in Column (2) of Table 4, Panel B are also inconsistent with the greenwashing explanation: both *TR-E Score (Concern)* and *TR-E Score (Strength)* have a significant moderating effect on the link between *Suspect* and *Toxic Release*. As discussed in footnote 23, greenwashing is easier for strengths than concerns, because strengths (concerns) are based mainly on inputs (outcomes). If so, the greenwashing explanation predicts significant (insignificant) interactions for strengths (concerns).

Table 6 The moderating effect of plant-level environmental track record**Panel A: Environmental track record based on past pollution levels**

	(1)	(2)
<i>Suspect</i> × <i>High Past Pollution</i>	-0.086** (-2.17)	-0.092* (-1.89)
<i>High Past Pollution</i>	0.835*** (16.17)	0.918*** (16.32)
<i>Suspect</i>	0.113*** (4.19)	
<i>Controls</i>	YES	N/A
<i>Plant FE</i>	YES	YES
<i>Industry-Year FE</i>	YES	
<i>Firm-Year FE</i>		YES
<i>Observations</i>	34,020	33,069
<i>Adjusted R-squared</i>	0.910	0.917

Panel B: Environmental track record based on past enforcement cases

<i>Past Enforcement</i> =	<i>Past Enforcement Count</i>		<i>Ln(1 + Past Settlement Amount)</i>	
	(1)	(2)	(3)	(4)
<i>Suspect</i> × <i>Past Enforcement</i>	-0.117* (-1.96)	-0.120** (-2.46)	-0.026* (-1.66)	-0.029** (-1.98)
<i>Past Enforcement</i>	0.043* (1.78)	0.036* (1.69)	0.009** (2.19)	0.008** (2.04)
<i>Suspect</i>	0.068*** (3.65)		0.068*** (3.69)	
<i>Controls</i>	YES	N/A	YES	N/A
<i>Plant FE</i>	YES	YES	YES	YES
<i>Industry-Year FE</i>	YES		YES	
<i>Firm-Year FE</i>		YES		YES
<i>Observations</i>	38,442	37,314	38,442	37,314
<i>Adjusted R-squared</i>	0.892	0.899	0.892	0.899

This table presents results for the moderating effect of a plant's environmental track record on the relation between meeting benchmarks and toxic release. The dependent variable is $\ln(1 + \text{Toxic Release})$. The independent variable of interest is the interaction term between *Suspect* and a measure of the plant's environmental track record. *Suspect* equals one if a firm meets or beats the latest annual consensus analyst forecast by two cents or less. In Panel A, we measure track record by an indicator variable (*High Past Pollution*) that equals one if a plant's toxic release over the previous year is above the sample median for that year. In Panel B, we measure track record by the number of EPA enforcement cases (*Past Enforcement Count*) and the dollar amount of penalties settled in EPA enforcement cases (*Past Settlement Amount*) in which the plant is a defendant, both measured over the past five years. *Controls* include $\ln(\text{Assets})$, *Leverage*, PPE/Assets , $\ln(\text{Turnover})$, $\text{CapEx}/\text{Assets}$, *M/B*, and *Nonroutine Release*. *t*-statistics are presented with robust standard errors clustered at the firm level. Variables are defined in Appendix B. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively

earnings benchmarks.⁴¹ The results also suggest that the E ratings and their component indicators obtained from MSCI are reliable measures of firm efforts to control pollution related to TRI toxins.

6 Additional analyses

6.1 The moderating effect of environmentally sustainable investing

As a potential driver of the increasing demand for corporate sustainability, the practice of sustainable investing (SI), in which investors select stocks rated high on ESG, has experienced unprecedented growth in the US and Europe (Grewal and Serafeim 2020; Gillan et al. 2021). How then does SI affect the relation between meeting benchmarks and toxic release? On the one hand, SI takes a long-term view and discourages managerial short-termism by deemphasizing quarterly earnings (Starks et al. 2020; Cao et al. 2021). The view that SI curbs short-termism predicts a weaker link between pollution and meeting benchmarks for firms with higher institutional ownership by funds that are focused on E-S (E-S IO).

On the other hand, institutions that are focused on E-S likely screen and select stocks based on E ratings and environmental track records (Amel-Zadeh and Serafeim 2018; Curtis et al. 2021; Heath et al. 2021). And we find that firms with high E ratings and strong environmental track records are more likely to pollute more to boost earnings when meeting earnings benchmarks. If so, we expect to find a *stronger* link between meeting benchmarks and pollution for high E-S IO firms. For convenience, we label this opposing view the “pollution slack view.”

To test these competing views, we measure E-S IO following the methodology of Cao et al. (2021) and Brandon et al. (2021). We first identify E-S institutional investors among 13F institutions based on the E-S footprint of the firms they invest in. We then calculate, for each sample firm, the percentage of institutional ownership held by these E-S investors (see Appendix B for details). To estimate the moderating effect of E-S IO, we replace *E Score (Net)* in regression (2) with *E-S IO*. We estimate this modified regression at the firm level, as *E-S IO* is defined at the firm-year level. The results presented in column (1) of Table 7 indicate a significantly positive coefficient on the interaction term *Suspect* × *E-S IO*, consistent with the pollution slack view.

Similar to Table 4, we confirm this inference with a placebo test that is based on a measure of sustainable institutional ownership derived from social & governance indicators (*S&G-S IO*) and calculated similarly to *E-S IO*. We expect weaker results for *S&G-S IO* because it reflects other priorities, separate from E-S. Column (2) presents results for the regression with *S&G-S IO* and *Suspect* × *S&G-S IO* as two additional explanatory variables. Again, consistent with the pollution slack view and with E ratings being more relevant than S&G ratings, we find an insignificant coefficient on *Suspect* × *S&G-S IO*.

⁴¹ In untabulated results, we fail to find a significant relation between E ratings and the tendency to meet or just beat earnings benchmarks or between E ratings and the negative market reaction to just missing earnings benchmarks. This suggests that high E rating firms and low E rating firms face similar pressures to meet earnings benchmarks.

Table 7 The moderating effect of environmentally sustainable institutional holding

	(1)	(2)
<i>Suspect</i> × <i>E-S IO</i>	0.906** (2.36)	0.905** (2.27)
<i>E-S IO</i>	-0.934*** (-2.71)	-0.905*** (-2.65)
<i>Suspect</i> × <i>S&G-S IO</i>		-0.030 (-0.07)
<i>S&G-S IO</i>		-0.164 (-0.55)
<i>Suspect</i>	0.006 (0.09)	0.011 (0.14)
<i>Controls</i>	YES	YES
<i>Firm FE</i>	YES	YES
<i>Industry-Year FE</i>	YES	YES
<i>Observations</i>	4,785	4,785
<i>Adjusted R-squared</i>	0.911	0.910

This table presents results for the moderating effect of environmentally sustainable institutional ownership (*E-S IO*) on the relation between meeting benchmarks and toxic releases. The dependent variable is $\ln(1 + \text{Toxic Release})$. The independent variable of interest is the interaction between *Suspect* and *E-S IO*. *Suspect* equals one if a firm meets or beats the latest annual consensus analyst forecast by two cents or less. *E-S IO* is the percentage of institutional ownership held by environmentally sustainable institutions. We follow Cao et al. (2021) to identify environmentally sustainable 13F institutions based on the E (of ESG) footprint of investee firms. For comparison, we also examine the interaction between *Suspect* and *S&G-S IO*, which is defined similarly to *E-S IO* but is based on the S&G (of ESG) footprint of investee firms. Control variables (*Controls*) include *Ln(Assets)*, *Leverage*, *PPE/Assets*, *Ln(Turnover)*, *CapEx/Assets*, *M/B*, and *Nonroutine Release*. *t*-statistics are presented in parentheses with robust standard errors clustered at the firm level. Variables are defined in Appendix B. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively

6.2 Relation between meeting benchmarks and SO₂ emission

Liu et al. (2021) show that meeting analyst forecasts is associated with higher SO₂ emission for Chinese SOEs controlled by the central government but not for locally controlled SOEs or non-SOE firms. Given that our sample of U.S. firms are closest to non-SOE firms in China, we investigate whether the firms in our sample, which increase their toxic releases to meet earnings benchmarks, also increase their SO₂ emissions. As SO₂ is one of the six criteria air pollutants highlighted by the EPA, we do not expect our firms to increase SO₂ emissions, because SO₂ is considerably more hazardous and more closely monitored than TRI toxins. Evidence consistent with this prediction has already been shown in Table 2 for ozone and lead, two other criteria pollutants.

To test our prediction, we collect annual plant-level data on SO₂ emissions from the EPA's Clean Air Markets Division (CAMD) and National Emissions Inventory (NEI) databases (Shive and Forster 2020). Panel A of Table 8 presents descriptive statistics. The mean (median) SO₂ emission at the plant level is 591.17 (0.36) thousand pounds in our sample, indicating considerable right skewness in SO₂ emission. We re-estimate

Table 8 The effect of meeting benchmarks on sulfur dioxide emissions

Panel A: Summary statistics						
	Mean	STD	P25	P50	P75	N
Plant-level	591.17	2,552.95	0.04	0.36	23.94	5452
Firm-level	3,056.14	13,951.77	0.14	3.79	199.75	989

Panel B: Effect of meeting benchmarks on SO₂ emissions		
	Firm-level	Plant-level
	(1)	(2)
<i>Suspect</i>	0.080 (0.76)	0.007 (0.13)
<i>Ln(Assets)</i>	-0.047 (-0.19)	-0.032 (-0.21)
<i>Leverage</i>	-0.335 (-0.60)	-0.174 (-0.56)
<i>PPE/Assets</i>	-0.835 (-0.71)	-0.297 (-0.48)
<i>Ln(Turnover)</i>	-0.791* (-1.84)	-0.037 (-0.21)
<i>CapEx/Assets</i>	-0.825 (-0.37)	1.059 (0.78)
<i>M/B</i>	0.201 (1.52)	0.012 (0.18)
<i>Firm FE</i>	YES	
<i>Plant FE</i>		YES
<i>Industry-Year FE</i>	YES	YES
<i>Observations</i>	784	4,746
<i>Adjusted R-squared</i>	0.921	0.925

This table examines the relation between meeting earnings benchmarks and sulfur dioxide (SO₂) emissions. Panel A presents summary statistics for SO₂ emissions at the firm- and plant-year level. Panel B presents regression results for the effect of meeting benchmarks on SO₂ emissions, similar to the regressions in Table 2, Panel A for emissions of TRI toxins, except the sample period begins in 2008 (availability of SO₂ emission data). The dependent variable is $\ln(1 + SO_2 \text{ Release})$, the natural logarithm of one plus $SO_2 \text{ Release}$. The independent variable of interest is *Suspect*, an indicator variable equal to one if a firm meets or beats the latest annual consensus EPS forecast by two cents or less. Columns (1) and (2) present the results for regressions estimated on firm-year observations and plant-year observations, respectively. *t*-statistics are presented in parentheses with robust standard errors clustered at the firm level. All variables are defined in Appendix B. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively

regression (1) for SO₂ emission and report the results in Panel B of Table 8. Consistent with our prediction, and consistent with the results for non-SOE Chinese firms in Liu et al. (2021), *Suspect* is not associated with SO₂ emissions. This finding is also consistent with the inference in Liu et al. (2021) that the positive relation they find between meeting benchmarks and SO₂ emission might be unique to Chinese central SOEs because they face low regulatory costs, given that they are effectively exempt from environmental regulation and enforcement.

6.3 Other ways to meet earnings benchmarks by managing forecast errors

We investigate whether firms that cut pollution abatement costs also engage in other ways to manage the forecast errors underlying earnings benchmarks. Forecast errors can be altered by managing earnings and guiding forecasts (Dechow et al. 2010). Earnings management can be achieved by altering activities (REM) and by altering accruals (accruals-based earnings management or AEM). Our results are provided below.

First, we examine whether our sample firms use four other REM methods to boost earnings to meet benchmarks: overproduction (Roychowdhury 2006), cutting SG&A expense (Roychowdhury 2006), cutting R&D expense (Baber et al. 1991; Bushee 1998), and cutting advertising expense (Cohen et al. 2010). To do so, we re-estimate regression (1) by replacing the dependent variable *Toxic Release* with two measures of production (*Prod Ratio* and *Production*), SG&A expense (*SG&A/Sales*), R&D expense (*R&D/Sales*), and advertising expense (*AD/Sales*).⁴² *Prod Ratio* is based on actual plant-level production data from the EPA's Pollution Prevention (P2) database. The P2 database provides the ratio of this year's output or outcome of processes in which chemicals are used (Akey and Appel 2021) to the corresponding value for last year.⁴³ We measure *Prod Ratio* at the plant level as the average of the EPA's production ratios across all chemicals released that year from that plant. *Production* is a firm-level estimate of production levels obtained from financial reports, defined as the sum of cost of goods sold and change in inventory scaled by lagged total assets (Dechow et al. 1998; Roychowdhury 2006). We believe that *Prod Ratio* is more reliable than *Production* because it directly measures production. *SG&A/Sales*, *R&D/Sales*, and *AD/Sales* are also obtained from firm-level financial reports.

The results reported in Table 9 indicate that the coefficient on *Suspect* is insignificant across columns (1)–(5), suggesting that our sample firms do not use other types of REM to meet earnings benchmarks. The insignificant coefficient estimates on both production variables are important because the overproduction explanation (e.g., Roychowdhury 2006) is particularly relevant here: overproduction increases pollution and product costs are substantial for our sample of manufacturing firms.

To confirm further that the increased pollution comes from cutting variable pollution abatement costs, we examine two implications. One implication is that the cost of goods sold should decline for *Suspect* firm-years. To test this prediction, we replace the dependent variable in regression (1) with sales-deflated cost of goods sold (*COGS/Sales*). The regression results, in column (6) of Table 9, confirm that *COGS/Sales* is significantly lower for *Suspect* firm-years.⁴⁴ Declines in *COGS/Sales* without

⁴² To avoid the potential bias documented by Chen et al. (2018), we estimate a single-step regression for REM instead of a two-step procedure: we first estimate abnormal operating activities, then regress it on *Suspect*.

⁴³ For example, if a chemical is used in the production of refrigerators, the production ratio is defined as the number of refrigerators produced this year divided by the number produced last year. If a chemical is used in a capacity not directly related to the product, the EPA requires facilities to report the ratio reflecting changes in this activity. For example, if a chemical is used to clean molds, the activity ratio is defined as the number of molds cleaned this year divided by the number cleaned last year. If a particular chemical is used in more than one process or activity, plants are required to calculate the production ratio by weighting the different production or activity variables for the different processes involved.

⁴⁴ A significantly negative coefficient on *Suspect* is also observed when we use the natural logarithm of *COGS/Sales* or the natural logarithm of *COGS* as the dependent variable.

Table 9 The effect of meeting benchmarks on other types of real earnings management

	Plant-level	Firm-level				
	<i>Prod Ratio</i>	<i>Production</i>	<i>SG&A/Sales</i>	<i>R&D/Sales</i>	<i>AD/Sales</i>	<i>COGS/Sales (for comparison)</i>
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Suspect</i>	0.014 (1.25)	-0.005 (-0.61)	0.000 (0.16)	0.001 (1.20)	-0.000 (-0.11)	-0.007*** (-2.88)
<i>Ln(Assets)</i>	0.031* (1.66)	-0.036** (-2.12)	-0.016*** (-3.93)	-0.000 (-0.31)	-0.000 (-0.84)	-0.001 (-0.17)
<i>Leverage</i>	0.023 (0.37)	-0.055 (-1.00)	-0.003 (-0.34)	-0.004 (-0.84)	0.001 (1.14)	-0.019 (-1.21)
<i>PPE/Assets</i>	0.212*** (2.82)	-0.140* (-1.81)	-0.017 (-1.06)	0.004 (0.61)	-0.003 (-1.35)	0.026 (0.95)
<i>Ln(Turnover)</i>	0.058** (2.12)	0.857*** (24.39)	-0.037*** (-6.47)	-0.007*** (-3.38)	-0.000 (-0.87)	0.032** (2.58)
<i>CapEx/Assets</i>	-0.358* (-1.68)	-0.140 (-0.78)	0.060* (1.65)	0.005 (0.34)	-0.002 (-0.44)	-0.037 (-0.62)
<i>M/B</i>	0.026*** (2.77)	-0.060*** (-4.68)	-0.002 (-0.97)	0.001 (1.19)	0.000 (1.30)	-0.034*** (-8.45)
<i>Nonroutine Release</i>	-0.010 (-0.77)	-0.003 (-0.70)	0.001 (1.03)	0.000 (0.05)	-0.000 (-0.08)	0.000 (0.16)
<i>Plant or Firm FE</i>	Plant	Firm	Firm	Firm	Firm	Firm
<i>Industry-Year FE</i>	YES	YES	YES	YES	YES	YES
<i>Observations</i>	36,333	4,763	4,763	4,763	4,763	4,763
<i>Adjusted R-squared</i>	0.036	0.934	0.947	0.929	0.921	0.927

This table examines the relation between meeting benchmarks and other types of real earnings management. The dependent variable in columns (1)–(5) is production ratio at the plant level (*Prod Ratio*) using data from the EPA's Pollution Prevention (P2) database, firm-level production inferred from financial statements (*Production*), selling, general, and administrative expense (*SG&A/Sales*), research and development expense (*R&D/Sales*), and advertising expense (*AD/Sales*). For comparison purposes, column (6) provides results for cost of goods sold (*COGS/Sales*), which declines when pollution abatement costs are cut. The independent variable of interest is *Suspect*, an indicator that equals one if a firm meets or beats the latest annual consensus EPS forecast by two cents or less. *t*-statistics are presented in parentheses with robust standard errors clustered at the firm level. Variables are defined in Appendix B. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively

corresponding declines in production are consistent with suspect firms cutting back on pollution abatement costs. Finding that our sample firms cut pollution control costs without engaging in other forms of REM might be partially due to pollution abatement costs being substantial relative to other costs, such that cutting those costs has the potential to alter per share earnings sufficiently to meet benchmarks. Regardless, we acknowledge that our findings might be sample-specific and not generalize to other firms.

The other implication is that cutting pollution abatement costs to boost short-term earnings is more feasible when those costs are mostly variable. To proxy for the variable/fixed nature of pollution abatement costs, we use the approach taken in the management accounting literature to estimate the fixed/variable nature of firm-level costs (e.g., Banker et al. 2014). Consistent with our prediction, the results in Table IA.6 of the Internet Appendix suggest that the positive link between toxic release and meeting earnings benchmarks is concentrated (absent) in firms with cost variability higher (lower) than the industry-year median.

Turning to accruals-based earnings management, we expect a stronger link between pollution and meeting the earnings benchmark for firms with limited opportunity to manage accruals, as indicated by balance sheets that are bloated by accumulated prior efforts to boost earnings via accruals (Barton and Simko 2002). Consistent with this prediction, we show in Table IA.7 of the Internet Appendix that the positive association between *Suspect* and *Toxic Release* is stronger among firms with higher net operating assets, a proxy for reduced capacity to use accruals management.

Regarding expectation management, we fail to find a moderating effect of downward expectation management—measured by downward revision of management forecasts and downward guidance of the consensus analyst forecast over the year—on the positive relation between *Suspect* and *Toxic Release* (see Table IA.8 of the Internet Appendix).

7 Conclusion

We examine two key pillars supporting the increasing demand for corporate environmental sustainability as a market-based solution to environmental externalities. First, managerial short-termism exacerbates negative externalities. Second, environmental sustainability curbs managerial short-termism and the resulting negative environmental impact.

We test the first proposition by using earnings management to meet earnings benchmarks as a measure of managerial short-termism, and environmental pollution as a measure of its negative externality. Consistent with this proposition, we find that firms attempting to meet earnings benchmarks increase significantly their release of toxic chemicals. This finding suggests that firms cut pollution abatement costs to boost earnings, as the short-term benefits of meeting earnings expectations exceed the long-term costs of higher pollution.

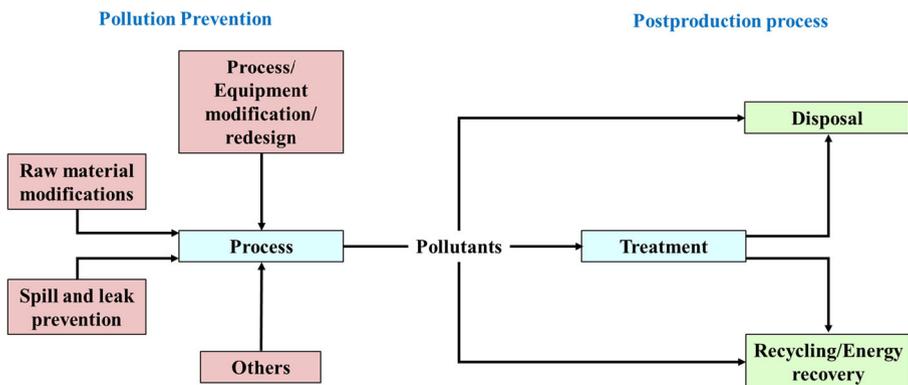
We test the second proposition by examining if the positive relation between pollution and meeting earnings benchmarks is attenuated for firms with higher environmental ratings. To our surprise, our evidence contradicts this proposition: firms that are viewed as more environmentally sustainable are in fact more, not less, likely to cut pollution abatement costs to meet earnings benchmarks. We show, however, that firms with higher E ratings have better environmental track records. This suggests an alternative explanation: better environmental track records create slack that lowers the costs of occasionally increasing pollution to boost earnings when needed to meet earnings benchmarks.

Our first set of results suggest that earnings management not only adversely affects stakeholders within the boundaries of the firm but also impacts society at large. This

finding contributes to the accounting literature on real earnings management as well as the economics literature on the determinants of environmental pollution. Our second set of results contribute to the literature on environmental sustainability and sustainable investing by questioning the view that environmental sustainability curbs short-termism. Firms with better environmental track records and firms held by more institutional investors that are focused on environmental sustainability are more likely to increase pollution to achieve short-term earnings benchmarks.

As is typical, some caveats are in order. First, data availability limits our ability to tie the pollution–meeting earnings benchmark link to cuts in pollution abatement activities. To make that connection, we need access to data on the relative costs and benefits of abatement technologies, which include how easy it is to switch on and off abatement processes, cost savings from cutting pollution abatement, and costs from increased pollution. This data limitation prevents us from conducting cross-sectional tests based on abatement technologies for our first research question. Second, while we find evidence consistent with the pollution slack explanation for our second research question, we acknowledge that there could be other explanations.

Appendix A: Pollution abatement



A plant's pollution level is jointly determined by its production processes and pollution abatement activities. Pollution abatement activities fall into two broad categories: pollution prevention and postproduction processing. Pollution prevention refers to practices that reduce the amount of hazardous substances entering a waste stream or being released to the environment before postproduction processing. Postproduction processing includes recycling, energy recovery, and treatment of hazardous substances.

Appendix B: Variable definitions

Sources for EPA data are provided in the definitions below. Definitions for indicators provided by MSCI ESG KLD STATS (shortened to MSCI) to build ESG-related variables are provided in Table IA.1 of the Internet Appendix. Forecast and actual EPS are taken from I/B/E/S, and all financial variables are taken from Compustat.

Variables	Definitions
Dependent Variables of Interest	
<i>Toxic Release</i>	The total amount (000's of lbs.) of all chemicals, administered under the EPA's Toxic Release Inventory (TRI) program, released to air, water, or ground by a plant or firm over a year.
<i>Toxicity-Weighted Release</i>	The toxicity-weighted amount of all chemicals, administered under the TRI program, released to air, water, or ground by a plant or firm over a year. Following the calculation of RSEI Hazard under the EPA's RSEI program (https://www.epa.gov/rsei), we first calculate the product of release quantity (000's of lbs.) and toxicity index (ranging from 0.02 to 1,400,000,000) for each chemical. We then aggregate the product across all chemicals at the plant or firm level.
<i>Toxic Release/Sales</i>	<i>Toxic Release</i> divided by sales of the firm for that year.
Independent Variables of Interest	
<i>Suspect</i>	An indicator variable equal to one if a firm meets or beats the latest consensus of analysts' annual EPS forecasts by two cents or less, where latest consensus is calculated as the average of each analyst's latest annual forecast issued within [-90, -4] days relative to the earnings announcement date.
<i>Suspect (180)</i>	<i>Suspect</i> based on the average of analysts' latest annual forecasts issued within [-180, -4] days relative to the earnings announcement date.
<i>Suspect (alternative consensus)</i>	<i>Suspect</i> based on the consensus forecast of annual EPS constructed by I/B/E/S in its summary history files at different dates: the last month before earnings announcement and at the ends of the four fiscal quarters for that year.
<i>Large Miss</i>	An indicator variable equal to one if a firm misses the latest annual consensus analyst forecast by more than three cents.
<i>Beat 2-5cents</i>	An indicator variable equal to one if a firm beats the latest annual consensus analyst forecast by more than two cents but no more than five cents.
<i>Large Beat</i>	An indicator variable equal to one if a firm beats the latest annual consensus analyst forecast by more than five cents.
<i>E Score (Net)</i>	Net environmental score defined as the number of strengths minus the number of concerns under the environment category of MSCI. We calculate <i>E Score (Net)</i> in two ways: using all environmental indicators covered in MSCI, and using only those indicators with at least ten years of data during our sample period.
<i>S&G Score (Net)</i>	Net composite social and governance score calculated as the sum of net scores, defined similarly to <i>E Score (Net)</i> , across community, diversity, employee relations, human rights, product, and corporate governance categories in MSCI.
<i>TR-E Score (Net)</i>	Net score for the subset of environmental strength/concern indicators that are directly related to toxic release generated by production processes. Specifically, it equals the number of strengths among indicators Env-Str-B, Env-Str-C, Env-Str-D, and Env-Str-X minus the number of concerns among indicators Env-Con-A, Env-Con-B, Env-Con-D, Env-Con-F, and Env-Con-X.
<i>TR-E Score (Strength)</i>	Strength score for the subset of environmental strength indicators that are directly related to toxic release generated by production processes. Specifically, it equals the number of strengths among indicators Env-Str-B, Env-Str-C, Env-Str-D, and Env-Str-X.
<i>TR-E Score (Concern)</i>	Concern score for the subset of environmental concern indicators that are directly related to toxic release generated by production processes. Specifically, it equals the number of concerns among indicators Env-Con-A, Env-Con-B, Env-Con-D, Env-Con-F, and Env-Con-X.
<i>NTR-E Score (Net)</i>	Net score for the subset of environmental strength/concern indicators that are not directly related to toxic release generated by production processes. Specifically, it equals the number of strengths among indicators Env-Str-A, Env-Str-E, and

	Env-Str-G minus the number of concerns among indicators Env-Con-C and Env-Con-E.
<i>NTR-E Score (Strength)</i>	Strength score for the subset of environmental strength indicators that are not directly related to toxic release generated by production processes. Specifically, it equals the number of strengths among indicators Env-Str-A, Env-Str-E, and Env-Str-G.
<i>NTR-E Score (Concern)</i>	Concern score for the subset of environmental concern indicators that are not directly related to toxic release generated by production processes. Specifically, it equals the number of concerns among indicators Env-Con-C and Env-Con-E.
<i>E-S IO</i>	Environmentally sustainable institutional ownership, defined as the average percentage of institutional ownership held by environmentally sustainable investment funds, which we identify from 13F institutions based on the environmental footprint of their portfolio firms following the methodology of Cao et al. (2021). Specifically, we calculate <i>E Score (Net)</i> for each firm in each year. We size-adjust the scores by sorting the stocks into ten deciles based on size, calculate the average <i>E Score (Net)</i> for each size decile, and subtract that average <i>E Score (Net)</i> from the raw <i>E Score (Net)</i> of each stock. We measure the environmental preference of an institutional investor by taking the value-weighted average of the size-adjusted <i>E Score (Net)</i> of all stocks in its portfolio at the end of each quarter. We then sort all institutions into three groups based on the value-weighted average size-adjusted <i>E Score (Net)</i> of their holding portfolios each quarter and define those in the top group as environmentally sustainable institutions. Finally, <i>E-S IO</i> at the stock level is calculated as the ratio of shares held by environmentally sustainable institutions to total shares held by all institutions.
<i>S&G-S IO</i>	Society & governance sustainable institutional ownership, defined as the average percentage of institutional ownership held by sustainable investment funds focused on society & governance. We construct this variable similar to <i>E-S IO</i> following the methodology of Cao et al. (2021) based on the S&G components of ESG ratings.
Other Variables	
<i>Assets</i>	Total assets at the beginning of the year.
<i>Leverage</i>	Total debt divided by beginning-of-year total assets.
<i>PPE/Assets</i>	Net property, plant, and equipment divided by beginning-of-year total assets.
<i>Turnover</i>	Total sales over the year divided by beginning-of-year total assets.
<i>CapEx/Assets</i>	Capital expenditures over the year divided by beginning-of-year total assets.
<i>M/B</i>	Ratio of market value to book value of total assets, as of the beginning of the year. Market value of total assets equals the sum of market value of equity, book value of total liabilities, and liquidation value of preferred stock minus deferred tax liabilities.
<i>Nonroutine Release</i>	The amount (000's of lbs.) of all chemicals administered under the TRI program that are released to air, water, or ground by a plant or firm over a year due to catastrophic or other one-time pollution events (e.g., accidents).
<i>Past Enforcement Count</i>	The number of EPA enforcement cases in which a plant (or plants of a firm) is (are) named as the defendant over the past five years. Enforcement cases are collected from the EPA's Federal enforcement and compliance (FE&C) database in the Enforcement and Compliance Online System (ECHO).
<i>Past Settlement Amount</i>	The dollar amount of settlement for all EPA enforcement cases in which a plant (or plants of a firm) is (are) named as the defendant over the past five years.
<i>High Past Pollution</i>	An indicator variable equal to one if the amount of toxic release from a plant over the previous year is above the sample median of that year.

<i>SO₂ Release</i>	The total amount (000's of lbs.) of SO ₂ emission from a plant, or plants of a firm, over a year reported to EPA's Clean Air Markets Division (CAMD) and National Emissions Inventory (NEI). SO ₂ data are only available after 2008.
<i>Prod Ratio</i>	The average of chemical-level production ratios for a plant-year. For each chemical administered under the TRI program, the production ratio is defined as the ratio of the reporting year's value reflecting the output or outcome of the process(es) in which the chemical is involved to its prior year's value. Production ratio is obtained from the EPA's Pollution Prevention (P2) database.
<i>Production</i>	The level of production inferred from financial statements equals the sum of cost of goods sold and the change in inventory, scaled by lagged total assets.
<i>SG&A/Sales</i>	Selling, general, and administrative expenses scaled by revenue for the firm-year.
<i>R&D/Sales</i>	Research and development expenses scaled by revenue for the firm-year.
<i>AD/Sales</i>	Advertising expenses scaled by revenue for the firm-year.
<i>COGS/Sales</i>	Cost of goods sold scaled by revenue for that firm-year.

Acknowledgments We thank Patricia Dechow (editor), two anonymous referees, Nerissa Brown, Judson Caskey (discussant), Vincent Castellani, Clara Chen, Rita Gunn, Naim Bugra Ozel, Shivaram Rajgopal, Aneesh Raghunandan, Theodore Sougiannis, Oktay Urcan, Edward Watts, Aaron Yoon, and participants at the 2021 Review of Accounting Studies Conference, 2021 AAA annual meeting, Penn State University, University of Illinois at Urbana-Champaign, University of Miami, University of New South Wales, Xiamen University, and Yale University for helpful comments and suggestions. We appreciate the comments and suggestions of the following EPA staff: Wes Austin, Ann Ferris, Al McGartland, and William Wheeler. Jake Thomas and Frank Zhang acknowledge financial support from the Yale School of Management. Wei Zhu acknowledges financial support from the Department of Accountancy, Gies College of Business. Wentao Yao acknowledges financial support from the Fundamental Research Funds for the Central Universities (No. 20720221048).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akey, P., and I. Appel. 2021. The limits of limited liability: Evidence from industrial pollution. *Journal of Finance* 76: 5–55. <https://doi.org/10.1111/jofi.12978>.
- Amel-Zadeh, A., and G. Serafeim. 2018. Why and how investors use ESG information: Evidence from a global survey. *Financial Analysts Journal* 74: 87–103. <https://doi.org/10.2469/faj.v74.n3.2>.
- Baber, W.R., P.M. Fairfield, and J.A. Haggard. 1991. The effect of concern about reported income on discretionary spending decisions: The case of research and development. *The Accounting Review* 66: 818–829.
- Banker, R.D., D. Byzalov, and J.M. Plehn-Dujowich. 2014. Demand uncertainty and cost behavior. *The Accounting Review* 89: 839–865.
- Baron, D.P. 2001. Private politics, corporate social responsibility, and integrated strategy. *Journal of Economics & Management Strategy* 10: 7–45. <https://doi.org/10.1162/105864001300122548>.
- Barton, J., and P. Simko. 2002. The balance sheet as an earnings management constraint. *The Accounting Review* 77: 1–27.

- Bénabou, R., and J. Tirole. 2010. Individual and corporate social responsibility. *Economica* 77: 1–19. <https://doi.org/10.1111/j.1468-0335.2009.00843.x>.
- Berg, F., J. Kölbel, and R. Rigobon. 2020. Aggregate confusion: The divergence of ESG ratings. Working paper. <https://doi.org/10.2139/ssrn.3438533>
- Bhojraj, S., P. Hribar, M. Picconi, and J. McInnis. 2009. Making sense of cents: An examination of firms that marginally miss or beat analyst forecasts. *Journal of Finance* 64: 2361–2388. <https://doi.org/10.1111/j.1540-6261.2009.01503.x>.
- Brandon, R. G., P. Krueger, and S. F. Mitali. 2021. The sustainability footprint of institutional investors: ESG driven price pressure and performance. Working paper. <https://ssrn.com/abstract=2918926>. Accessed 20 Jan 2022.
- Bui, L.T., and C.J. Mayer. 2003. Regulation and capitalization of environmental amenities: Evidence from the toxic release inventory in Massachusetts. *Review of Economics and Statistics* 85: 693–708. <https://doi.org/10.1162/003465303322369821>.
- Bushee, B.J. 1998. The influence of institutional investors on myopic R&D investment behavior. *The Accounting Review* 73: 19–45.
- Cao, J., S. Titman, X. Zhan, and W. Zhang. 2021. ESG preference, institutional trading, and stock return patterns. Working paper. <https://doi.org/10.2139/ssrn.3353623>.
- Caskey, J., and N.B. Ozel. 2017. Earnings expectations and employee safety. *Journal of Accounting and Economics* 63: 121–141. <https://doi.org/10.1016/j.jacceco.2016.12.002>.
- Chatterji, A.K., D.I. Levine, and M.W. Toffel. 2009. How well do social ratings actually measure corporate social responsibility? *Journal of Economics & Management Strategy* 18: 125–169. <https://doi.org/10.1111/j.1530-9134.2009.00210.x>.
- Chen, W., P. Hribar, and S. Melessa. 2018. Incorrect inferences when using residuals as dependent variables. *Journal of Accounting Research* 56: 751–796. <https://doi.org/10.1111/1475-679X.12195>.
- Cheong, F.S., and J. Thomas. 2018. Management of reported and forecast EPS, investor responses, and research implications. *Management Science* 64: 4277–4301. <https://doi.org/10.1287/mnsc.2017.2832>.
- Christensen, D.M., G. Serafeim, and S. Sikochi. 2021. Why is corporate virtue in the eye of the beholder? The case of ESG ratings. *The Accounting Review* 97: 147–175. <https://doi.org/10.2308/TAR-2019-0506>.
- Clement, M.B., and S.Y. Tse. 2003. Do investors respond to analysts' forecast revisions as if forecast accuracy is all that matters? *The Accounting Review* 78: 227–249. <https://doi.org/10.2308/accr.2003.78.1.227>.
- Cohen, D.A., R.N. Hann, and M. Ogneva. 2007. Another look at GAAP versus the street: An empirical assessment of measurement error bias. *Review of Accounting Studies* 12: 271–303. <https://doi.org/10.1007/s11142-007-9029-0>.
- Cohen, D.A., R. Mashruwala, and T. Zach. 2010. The use of advertising activities to meet earnings benchmarks: Evidence from monthly data. *Review of Accounting Studies* 15: 808–832. <https://doi.org/10.1016/j.jacceco.2010.01.002>.
- Cohen, D.A., and P. Zarowin. 2010. Accrual-based and real earnings management activities around seasoned equity offerings. *Journal of Accounting and Economics* 50: 2–19. <https://doi.org/10.1007/s11142-009-9105-8>.
- Cohn, J., and T. Deryugina. 2018. Firm-level financial resources and environmental spills. Working paper. <https://doi.org/10.3386/w24516>
- Currie, J., J.G. Zivin, J. Mullins, and M. Neidell. 2014. What do we know about short- and long-term effects of early-life exposure to pollution? *Annual Review of Resource Economics* 6: 217–247. <https://doi.org/10.1146/annurev-resource-100913-012610>.
- Curtis, Q., J.E. Fisch, and A. Robertson. 2021. Do ESG mutual funds deliver on their promises? *Michigan Law Review*, forthcoming. <https://ssrn.com/abstract=3839785>. Accessed 20 Jan 2022.
- Dechow, P.M., W. Ge, and C. Schrand. 2010. Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics* 50: 344–401. <https://doi.org/10.1016/j.jacceco.2010.09.001>.
- Dechow, P.M., S.P. Kothari, and R.L. Watts. 1998. The relation between earnings and cash flows. *Journal of Accounting and Economics* 25: 133–168. [https://doi.org/10.1016/S0165-4101\(98\)00020-2](https://doi.org/10.1016/S0165-4101(98)00020-2).
- Dechow, P.M., and R.G. Sloan. 1991. Executive incentives and the horizon problem: An empirical investigation. *Journal of Accounting and Economics* 14: 51–89. [https://doi.org/10.1016/0167-7187\(91\)90058-S](https://doi.org/10.1016/0167-7187(91)90058-S).
- Degeorge, F., J. Patel, and R. Zeckhauser. 1999. Earnings management to exceed thresholds. *Journal of Business* 72: 1–33. <https://doi.org/10.1086/209601>.
- Di Giuli, A., and L. Kostovetsky. 2014. Are red or blue companies more likely to go green? Politics and corporate social responsibility. *Journal of Financial Economics* 111: 158–180. <https://doi.org/10.1016/j.jfineco.2013.10.002>.

- Dowell, G., S. Hart, and B. Yeung. 2000. Do corporate global environmental standards create or destroy market value? *Management Science* 46: 1059–1074. <https://doi.org/10.1287/mnsc.46.8.1059.12030>.
- Duflo, E., M. Greenstone, R. Pande, and N. Ryan. 2013. Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India. *Quarterly Journal of Economics* 128: 1499–1545. <https://doi.org/10.1093/qje/qjt024>.
- Eaton, S., and G. Kostka. 2017. Central protectionism in China: The “central SOE problem” in environmental governance. *The China Quarterly* 231: 685–704.
- Eccles, R.G., I. Ioannou, and G. Serafeim. 2014. The impact of corporate sustainability on organizational processes and performance. *Management Science* 60: 2835–2857. <https://doi.org/10.1287/mnsc.2014.1984>.
- EPA. 1998. 1994 and 1995 Toxic Release Inventory: Data quality report. <https://nepis.epa.gov/Exe/ZyPDF.cgi/20009LE0.PDF?DockKey=20009LE0.PDF>. Accessed 20 Jan 2022.
- Ferrell, A., H. Liang, and L. Renneboog. 2016. Socially responsible firms. *Journal of Financial Economics* 122: 585–606.
- Fink, L. 2019. Larry Fink’s 2019 letter to CEOs. <https://www.blackrock.com/americas-offshore/en/2019-larry-fink-ceo-letter>. Accessed 20 Jan 2022.
- Friede, G., T. Busch, and A. Bassen. 2015. ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment* 5: 210–233. <https://doi.org/10.1080/20430795.2015.1118917>.
- Gibson, R., S. Glossner, P. Krueger, P. Matos, and T. Steffen. 2021. Responsible institutional investing around the world. Working paper. <https://doi.org/10.2139/ssrn.3525530>
- Gillan, S.L., A. Koch, and L.T. Starks. 2021. Firms and social responsibility: A review of ESG and CSR research in corporate finance. *Journal of Corporate Finance* 66: 101889.
- Goetz, M. 2018. Financial constraints and corporate environmental responsibility. Working paper. <https://doi.org/10.2139/ssrn.3230344>
- Graham, J.R., C.R. Harvey, and S. Rajgopal. 2005. The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40: 3–73. <https://doi.org/10.1016/j.jacceco.2005.01.002>.
- Greening, D.W., and D.B. Turban. 2000. Corporate social performance as a competitive advantage in attracting a quality workforce. *Business & Society* 39: 254–280. <https://doi.org/10.1177/000765030003900302>.
- Greenstone, M. 2002. The impacts of environmental regulations on industrial activity: Evidence from the 1970 and 1977 clean air act amendments and the census of manufactures. *Journal of Political Economy* 111: 1175–1219. <https://doi.org/10.1086/342808>.
- Greenstone, M. 2003. Estimating regulation-induced substitution: The effect of the clean air act on water and ground pollution. *American Economic Review* 93: 442–448. <https://doi.org/10.1257/000282803321947498>.
- Grewal, J., and G. Serafeim. 2020. Research on corporate sustainability: Review and directions for future research. *Foundations and Trends® in Accounting* 14: 73–127. <https://doi.org/10.1561/14000000061>.
- Heath, D., D. Macciocchi, R. Michaely, and M. C. Ringgenberg. 2021. Does socially responsible investing change firm behavior? Working paper. <https://doi.org/10.2139/ssrn.3837706>.
- Hoepner, A. G., I. Oikonomou, Z. Sautner, L. T. Starks, and X. Zhou. 2019. ESG shareholder engagement and downside risk. Working paper. <https://doi.org/10.2139/ssrn.2874252>.
- Joshi, S., R. Krishnan, and L. Lave. 2001. Estimating the hidden costs of environmental regulation. *The Accounting Review* 76: 171–198. <https://doi.org/10.2308/accr.2001.76.2.171>.
- Khan, M., G. Serafeim, and A. Yoon. 2016. Corporate sustainability: First evidence on materiality. *The Accounting Review* 91: 1697–1724. <https://doi.org/10.2308/accr-51383>.
- Kim, T., and Q. Xu. 2022. Financial constraints and corporate environmental policies. *Review of Financial Studies* 35: 576–635. <https://doi.org/10.1093/rfs/hhab056>.
- Kirk, M.P., D.A. Reppenhagen, and J.W. Tucker. 2014. Meeting individual analyst expectations. *The Accounting Review* 89: 2203–2231. <https://doi.org/10.2308/accr-50828>.
- Konar, S., and M.A. Cohen. 2001. Does the market value environmental performance? *Review of Economics and Statistics* 83: 281–289. <https://doi.org/10.1162/00346530151143815>.
- Krüger, P. 2015. Corporate goodness and shareholder wealth. *Journal of Financial Economics* 115: 304–329.
- Lev, B., and S. Sunder. 1979. Methodological issues in the use of financial ratios. *Journal of Accounting and Economics* 1: 187–210. [https://doi.org/10.1016/0165-4101\(79\)90007-7](https://doi.org/10.1016/0165-4101(79)90007-7).
- Levinthal, D., and C. Rerup. 2021. The plural of goal: Learning in a world of ambiguity. *Organization Science* 32: 527–543. <https://doi.org/10.1287/orsc.2020.1383>.

- Lins, K.V., H. Servaes, and A. Tamayo. 2017. Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *Journal of Finance* 72: 1785–1824. <https://doi.org/10.1111/jofi.12505>.
- Liu, Z., H. Shen, M. Welker, N. Zhang, and Y. Zhao. 2021. Gone with the wind: An externality of earnings pressure. *Journal of Accounting and Economics* 72: 101403. <https://doi.org/10.1016/j.jacceco.2021.101403>.
- Lopez-de-Silanes, F., J.A. McCahery, and P.C. Pudschedl. 2020. ESG performance and disclosure: A cross-country analysis. *Singapore Journal of Legal Studies* 3: 217–241. <https://doi.org/10.2139/ssrn.3506084>.
- Monin, B., and D.T. Miller. 2001. Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology* 81: 33. <https://doi.org/10.1037/0022-3514.81.1.33>.
- Navarro, P. 1988. Why do corporations give to charity? *Journal of Business* 61: 65–93. <https://doi.org/10.1086/296420>.
- Raghunandan, A. 2021. Financial misconduct and employee mistreatment: Evidence from wage theft. *Review of Accounting Studies* 26: 867–905.
- Raghunandan, A., and S. Rajgopal. 2021. Do socially responsible firms walk the talk? Working paper. <https://doi.org/10.2139/ssrn.3609056>.
- Roychowdhury, S. 2006. Earnings management through real activities manipulation. *Journal of Accounting and Economics* 42: 335–370. <https://doi.org/10.1016/j.jacceco.2006.01.002>.
- Shive, S.A., and M.M. Forster. 2020. Corporate governance and pollution externalities of public and private firms. *Review of Financial Studies* 33: 1296–1330. <https://doi.org/10.1093/rfs/hhz079>.
- Skinner, D.J., and R.G. Sloan. 2002. Earnings surprises, growth expectations, and stock returns or don't let an earnings torpedo sink your portfolio. *Review of Accounting Studies* 7: 289–312.
- Starks, L. T., P. Venkat, and Q. Zhu. 2020. Corporate ESG profiles and investor horizons. Working paper. <https://doi.org/10.2139/ssrn.3049943>
- Tomar, S. 2019. CSR disclosure and benchmarking-learning: Emissions responses to mandatory greenhouse gas disclosure. Working paper. <https://doi.org/10.2139/ssrn.3448904>
- White, A., D. Savage, J. Brody, D. Cavander, and L. Lach. 1995. Environmental cost accounting for capital budgeting: A benchmark survey of management accountants. Tellus Institute. <https://p2infohouse.org/ref/03/02262.pdf>. Accessed 20 Jan 2022.
- Xiong, X., and I. Png. 2019. Location of U.S. manufacturing, 1987–2014: A new dataset. Working paper. <https://doi.org/10.2139/ssrn.3401582>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.