Truth-Telling by Third-Party Auditors and the Response of Polluting Firms: Experimental Evidence from India

Esther Duflo
Michael Greenstone
Rohini Pande
Nicholas Ryan

Truth-telling by Third-party Auditors and the Response of Polluting Firms: Experimental Evidence from India[*]

Esther Duflo[†]
Michael Greenstone[‡]
Rohini Pande[§]
Nicholas Ryan[¶]

May 2, 2013

Abstract

In many regulated markets, private, third-party auditors are chosen and paid by the firms that they audit, potentially creating a conflict of interest. This paper reports on a two-year field experiment in the Indian state of Gujarat that sought to curb such a conflict by altering the market structure for environmental audits of industrial plants to incentivize accurate reporting. There are three main results. First, the status quo system was largely corrupted, with auditors systematically reporting plant emissions just below the standard, although true emissions were typically higher. Second, the treatment caused auditors to report more truthfully and very significantly lowered the fraction of plants that were falsely reported as compliant with pollution standards. Third, treatment plants, in turn, reduced their pollution emissions. The results suggest reformed incentives for third-party auditors can improve their reporting and make regulation more effective.

[†]MIT, eduflo@mit.edu

[‡]MIT, mgreenst@mit.edu

[§]Corresponding author. Harvard Kennedy School, Harvard University, Mailbox 46, 79 JFK St., Cambridge, MA 02138, rohini_pande@harvard.edu

[¶]Harvard, nickryan@fas.harvard.edu

I Introduction

The use of third-party auditing to monitor the compliance of firms with regulation is ubiquitous. Third-party audits are the norm in financial accounting, and in many countries credit ratings from third-party agencies serve an important regulatory role (White, 2010). Consumer and commodity markets use third-party auditors to monitor standards, including those for food safety, healthcare, flowers, timber and many durable goods (Hatanaka et al., 2005; Raynolds et al., 2007; Dranove and Jin, 2010). With respect to environmental regulation, the focus of this paper, several countries use third-party auditors to verify firm compliance with national laws and regulations (Kunreuther et al., 2002; Paliwal, 2006). Third-party auditing is also used to enforce international environmental standards, including ISO 14001 certification and verification of carbon abatement in the carbon offset market (Potoski and Prakash, 2005; Bhattacharyya, 2011).

These markets share a common characteristic—the auditor is chosen by, paid by and reports to the audited firm. This feature creates a conflict of interest between reporting the truth and reporting what is beneficial for the client. To maintain business, third-party auditors have incentives to shade or falsify their reports, which may corrupt information provision and, in turn, undermine regulation. Events brought to light by the recent financial crisis suggest this is a real concern.[1] Yet, despite periodic calls for reform to increase the independence of third-party auditors, we are unaware of a single instance of an enacted reform that fundamentally alters the incentives of third-party auditors.[2]

This paper reports on a two-year field experiment conducted in collaboration with the environmental regulatory body in Gujarat, India. Since 1996, the state has had a third-

---

[1]For overviews of problems in the U.S. corporate audit and credit ratings markets see Ronen (2010) and White (2010), respectively. Biased reporting appears to be a key issue for credit rating agencies: for a single credit agency, Griffin and Tang (2011) show higher accuracy of the internal surveillance team's judgments on CDO ratings than the business-oriented ratings team's, and that the accuracy difference predicts future downgrades. Strobl and Xia (2011) compared ratings for the same companies provided by two credit rating agencies, where one agency uses a issuer-pay model and the other an investor-pay model. The difference in ratings is more pronounced when the issuer-pay rating agency plausibly has more business at stake.

[2]In 2002, the Sarbanes-Oxley Act made auditors of public companies subject to oversight by a private-sector, nonprofit corporation. This corporation determines who can perform audits, conducts investigations and sets fines. Three former SEC Chairmen testified in favor mandatory auditor rotation, which was not adopted. (The Act also required SEC to report to Congress on credit rating agencies but did not reform this sector.) In 2003 the Securities and Exchange Commission adopted rules on auditor independence that focused on restrictions on and disclosure of non-audit activities. In 2008 New York State Attorney General Andrew Cuomo reached an agreement with credit rating agencies which required upfront payment for their ratings. The Dodd-Frank financial reform bill and the Sarbanes-Oxley Act restrict the services that auditors or credit rating agencies can offer plants that they audit.

party audit system for plants with high pollution potential, wherein certified auditors annually submit pollution readings and suggested pollution control measures for the audited plants to the Gujarat Pollution Control Board (Gujarat High Court, 1996). Although the Gujarat High Court put in place several safeguards to limit conflicts of interest, the basic financial arrangement underlying these audits is typical of the practice the world over—plants hire and pay auditors directly, and the work of those auditors is subject to very little oversight. In conversations we had before beginning this study, the regulators, auditors, and polluting plants all agreed that the status quo audit system produced unreliable information. As evidence of this common opinion, the reported market price for an audit was often lower than the cost of collecting pollution readings, suggesting that at least some readings were not even taken.

Our experiment altered the market structure in several complementary ways in order to incentivize accurate reporting. All 473 audit-eligible plants in two populous and heavily polluted industrial regions of Gujarat entered the experimental sample. In each region, half the plants were randomized into a treatment with four parts. First, treatment plants were randomly assigned an auditor that they were required to use. Second, auditors were paid from a central pool, rather than by the plant, and their fee was set in advance at a flat rate, high enough to cover pollution measurement and leave the auditor a modest profit margin. Third, a random sample of each auditor's pollution readings were verified with follow-up visits to the audited plants by an independent technical agency that collected readings for the same pollutants at the same places as the auditor, usually within a couple weeks of the auditor readings. We refer to the follow-up visits as backchecks for the remainder of the paper. While the 20% probability of a backcheck was public knowledge, actual backcheck visits were unannounced. Fourth and finally, at the start of the second year, treatment auditors were informed that their pay would be linked to their reporting accuracy, as measured by the backchecks. (During the first year, we did not specify any explicit consequence of good or poor performance. Auditors, however, may have anticipated lower chances of staying included in the scheme if found to be systematically biased.)

We collate data from several sources. We collected all audit reports for years one and two filed with the regulator. We directly obtained backcheck readings from the agencies conducting backchecks. Towards the end of the second year, we hired the same technical agencies to do identical backchecks in a random sample of control plants; these backchecks were unannounced

and not used to monitor or reward auditor behavior. The availability of auditor and backcheck readings from the same plants and at nearly the same times offers a unique opportunity to compare true pollution levels with the auditors' reports of pollution in both the treatment and control plants. Finally, roughly six months after the last audit visit in the experiment, we ran an independent endline survey of pollution outcomes in all treatment and control plants.

We have three main findings. First, status quo audit reporting is corrupted, as auditors systematically report plant pollution readings just below the regulatory standard. The average difference between audit and backcheck pollution readings across all reported pollutants is -0.30 standard deviations in the control group. A comparison of audit and backcheck readings in the control indicates that 29% of audit reports falsely report readings as below the relevant regulatory standard. Further, much of this false reporting comes in the form of extra reports just below the standard, which are presumably less likely to attract regulatory attention than would be reports showing compliance by a wide margin.

Second, the treatment caused auditors to report more truthfully and reduced the fraction of plants that were falsely reported as compliant with pollution standards. Relative to backcheck readings, auditors for treatment plants report pollution readings that are 0.15 to 0.21 standard deviations—or 50% to 70% higher—than control auditors. This result is robust to the inclusion of auditor fixed effects, which allows us to compare the behavior of the same auditors simultaneously working in *both* treatment and control plants. This, in turn, suggests that the results are not due to a selection of different auditors in treatment versus control plants. Further, auditors working in treatment plants are 23 percentage points or 80% less likely to falsely report a pollution reading as in compliance with the relevant regulatory standard.

Third, treatment plants reduced emissions, presumably because they understood that the regulatory authority would receive more reliable audit reports. Average pollution in the treatment group fell by 0.21 standard deviations, with reductions concentrated among plants with the highest readings. We document that in practice, the regulator reserves the harshest penalties for plants with readings that significantly exceed the standard, so it is not surprising that that the dirtiest plants responded by reducing emissions the most.

The treatment, which included multiple parts, was implemented as a single package. Hence, we cannot separately identify the effects of the treatment components—auditor assignments,

fixed pay from a central pool, backchecks and incentive pay—using experimental variation. It is clear that the treatment may have changed auditor behavior through several plausible channels. First, the assignment of auditors to plants at a fixed price means that plants cannot dismiss their auditor to obtain a better report and/or a lower price. Further, auditors cannot hold up plants by extracting their entire willingness to pay to avoid a bad report once they are assigned. Second, the regulator can use backchecks to monitor auditor quality and, though no sanctions for low quality were specified in the experiment in the first year, auditors may have anticipated a higher return to accurate reporting. Third, although auditors working in the treatment could have decided to just pocket the extra pay and not report differently, the above-market rate may have increased auditors' expected return to accurate reporting through an efficiency wage channel. Specifically, auditors for treatment plants may have decided to report more accurately because they had more to lose if decertified. We provide non-experimental evidence that financial incentives for accuracy in the second year independently contributed to improved reporting.

The paper contributes to several literatures. We provide empirical evidence on economic incentives in third-party audit markets, a literature that has been mainly theoretical so far (Dranove and Jin, 2010; Bolton et al., 2012). More broadly, we contribute to the empirical literature on corruption and development (for an overview, see Olken and Pande, 2012). We analyze data from multiple sources to measure the incidence of corruption by third-party auditors, adding to the set of papers that measure corruption by comparing outcomes reported by a potentially corrupt provider with independent estimates (see, for instance, Olken (2007) and Niehaus and Sukhtankar (2012)). Arguably, a unique strength of our paper is that we trace the impacts of a reform intended to reduce corruption all the way through to the key welfare outcome of interest—industrial pollution levels. Such pollution has been shown to be harmful to labor productivity and health (Hanna and Oliva, 2011; Graff-Zivin and Neidell, 2012; Greenstone and Hanna, 2013; Chay and Greenstone, 2003; Chen et al., 2013).

Turning to the specific channels underlying our experimental treatment, on assignments, Bazerman et al. (1997) and Bazerman et al. (2002) suggest that auditors find it psychologically impossible to remain impartial when deeply involved in their clients' interests. In this vein, the treatment provides a test of whether such ties are impervious to changes in market structure that incentivize truthful reporting. On monitoring, Becker and Stigler (1974) argued that

agents with higher wages will be less corrupt if they are under supervision. Rahman (2012) argues theoretically for a contract where the principal randomly changes a state variable and measures whether the agent reports on this change; backchecks are a good substitute for such manipulation in this context given that pollution itself varies with some randomness. Absent supervision or monitoring, evidence remains mixed on whether higher pay matters in and of itself (Rauch and Evans, 2000). Speaking to the monitoring and wage interaction, Di Tella and Schargrodsky (2003) show that higher wages reduce corruption in hospital procurement in Buenos Aires only when the probability of audit is reasonably high, which is consistent with our findings.

Our findings on auditor and plant behavior are valid in the specific context of the reform evaluated and, of course, may not apply to other sectors or to environmental regulation in other countries. That caution notwithstanding, this paper presents clear evidence that altering economic incentives can cause third-party auditors to switch from biased reporting towards truth-telling, causing regulated plants to respond in turn.

The remainder of the paper is organized as follows. In Section II, we describe the background and experimental design. In Section III we discuss our data collection and present summary statistics. Section IV presents the econometric approach and results and Section V concludes.

II Background and Experimental Design

A Study Context

Our study was conducted in Gujarat, one of India's fastest growing industrial states (Chakravorty, 2003). Since 1991-92, the peak of industrial licensing reform, net state domestic product in Gujarat has grown at an average of 8% per year. Today, the state accounts for about 5% of the Indian population, but 9% of India's registered manufacturing employment and 19% of output (Authors' calculation, Annual Survey of Industries, 2004-05). Rapid industrial growth has, however, been accompanied by a severe degradation of air and water quality. Gujarat contains the two most polluted industrial clusters in India, and three of India's five most polluted rivers (Central Pollution Control Board, 2007, 2009b). Essentially all large cities in the state, as well as some industrial areas, violate the National Ambient Air Quality standards for

Respirable Suspended Particulate Matter (RSPM) (Central Pollution Control Board, 2009a), an air pollutant dangerous for human health.

High levels of industrial pollution persist despite a stringent regulatory framework for pollution control (Greenstone and Hanna, 2013).[3] National laws set minimum levels of stringency for pollution standards, but basically all enforcement of environmental regulations occurs at the state level. State Pollution Control Boards, such as the Gujarat Pollution Control Board (GPCB) are responsible for enforcing the provisions of the Water Act (1974) and the subsequent Air (1981) and Environmental Protection (1986) Acts and their attendant command-and-control pollution regulations. GPCB is responsible for the monitoring and regulation of approximately 20,000 plants.

B   Environmental Audit Regulation

The main instruments that GPCB uses to limit industrial pollution are plant-level inspections and third party environmental audits. This paper focuses on the environmental audit system.

In 1996, in order to remedy the perceived failure of inspections in enforcing pollution standards, the High Court of Gujarat introduced the first third-party environmental audit system in India (Gujarat High Court, 1996). Under the scheme, plants with high pollution potential must submit a yearly environmental audit, conducted by an audit firm hired and paid for by the plant. Audit-eligible plants are classified as Schedule I (most polluting) or Schedule II (highly, but somewhat less polluting) on a basis of three dimensions: what the plant produces, where it sends its effluent (i.e., wastewater), and the volume of that effluent.[4] Schedule I plants must be audited by Schedule I auditors, usually an engineering college or similar institution. Schedule II plants must be audited by a private audit firm, called a Schedule II auditor. This study concerns reporting of Schedule II auditors and henceforth we refer to plants in Schedule II as "audit eligible." We also refer to regulated industrial plants as "plants" throughout and reserve the word "firms" for audit firms.

Auditors visit each plant for about one day in each of three seasons of the year to observe environmental management practices and measure pollution. Auditors compile their findings

---

[3]The Water (Prevention and Control of Pollution) Act of 1974 created the Central Pollution Control Board as a coordinating body to set pollution standards and the state boards as enforcement agencies.

[4]For example, plants that produce certain types of dyes and dye intermediates are classified in Schedule II, roughly, if their effluent is between 25 and 100 thousand liters per day, with variations around this classification based on whether the effluent discharged by the plant goes on to further treatment in a common effluent treatment plant (CETP). A plant with effluent below 25,000 liters would be exempt from the audit requirement.

from the three visits in a standardized format, fixed by the audit regulation, and submit the audit report to the plant and GPCB by February 15th of the following year. The final audit report describes the production process and physical state of the plant, including the measures the plant has taken for pollution control and the results of pollution sampling during each of the visits. Finally, auditors provide recommendations on pollution control to the plant.

On paper, the audit system also includes several safeguards and severe penalties for auditors found cheating. Each four-member audit team must meet technical standards and be re-certified by the regulator every two years.[5] Audit teams can audit at most 15 plants per year, and an audit firm, which may employ several teams, can audit a plant at most three years in a row. Auditors with reports found to be inaccurate are liable to be decertified and their reports on behalf of other plants declared void.

On the other side of the market, for an eligible plant, failure to submit an audit is punishable by closure and disconnection of water and electricity.[6] A report showing noncompliance with the terms of a plant's environmental consent can also be punished by closure or fine (Gujarat High Court, 1996). As we demonstrate below, the GPCB issues penalties to plants with a surprisingly high frequency when they have evidence of violations.

Nevertheless, all sides consider the audit system, as currently implemented, to function poorly. Industry recently litigated against the scheme, somewhat ironically and without success, to get the High Court of Gujarat to throw out the audit requirement on account of GPCB not following up on audit reports (Gujarat High Court, 2010). The regulator, for its part, believes that inaccurate reporting renders audits useless for enforcement, so review of submitted audits by GPCB is mostly *pro forma*.

Consistent with auditor shopping, we observe strong price competition in the environmental audit market. In interviews conducted prior to the experiment, both auditors and plants claimed that an audit report could be purchased for as low as INR 10,000-15,000 (roughly $200-$300). Our data on actual prices paid by control plants indicate that, conditional on reporting any payment, plants reported a mean payment of roughly INR 24,000. It is highly relevant that these measures of the price for an audit are significantly less than our best estimates of the true costs of conducting an audit for most of the plants in the sample, which

---

[5]Team members are required to have degrees in environmental engineering, chemical engineering, chemistry and biology, and minimum of two members must have at least one year's experience in environmental management.

[6]In practice, some plants do not submit reports, usually claiming they are not eligible.

includes sending audit teams to plants three times per year, taking the requisite pollution readings, and having the results analyzed at a certified laboratory (more on this below); the implication is that auditors were frequently not taking complete pollution measurements and presumably reporting readings based on other factors.

C Experimental Sample and Design

In collaboration with GPCB, we designed and evaluated a modified audit system which sought to improve the accuracy of auditor reporting.[7]

Our sample is the population of audit-eligible plants in the GPCB regions of Ahmedabad and Surat, the two largest cities of Gujarat. We obtained from GPCB a list of all 2,771 red category (i.e. high pollution potential) plants with reported capital investment less than INR 100 million (about USD 2 million), which are designated small or medium scale. Based on available data and in accord with the eligibility criteria, we selected 633 plants as the provisional sample of audit-eligible plants.[8] Just before the 2009 audit season, we randomly assigned half of the plants within this provisional sample, stratified by region, to the audit treatment group. After the randomization, we collected the detailed sectoral information needed to determine eligibility and, using exactly the same criteria, eliminated plants found to be ineligible from both the treatment and control groups, reaching the study sample of 473 plants, 49.2% of which belong to the treatment group.

Treatment plants were assigned to the audit treatment once, in 2009, for the audit years 2009 (hereafter year one) and 2010 (year two). Treatment plants were formally notified of the changes in the audit regulation that would apply to them by a letter from GPCB.[9] Relative to the status quo, the treatment altered three components of the audit system during year one: an auditor was randomly assigned to the plant, paid from a central pool at a fixed rate and its reports were backchecked for accuracy. In year two only, direct incentive pay for auditor accuracy was added. These components were implemented as follows.

Assignment and Fixed Pay. Auditors were randomly assigned to treatment plants by

---

[7]This experiment was designed and undertaken concurrently with the evaluation of another intervention, an increase in inspection frequency for some plants, which was conducted stratified on the audit treatment and which we study in a separate paper.

[8]GPCB did not have a definitive list of audit-eligible firms at the time of sample selection.

[9]The text of the letter sent to treatment plants is in the Online Appendix. Plants commonly receive regulatory notices and we do not believe the letter itself was a treatment channel, since it did not induce treatment plants to submit audits at a higher rate than control plants.

the research team and paid from a central pool of funds raised for the study. The payment was fixed at INR 45,000 in the first year. This rate was estimated by applying GPCB's sampling charges to the average plant characteristics in the audit sample and adding a small margin. Variation in actual cost arises due to a plant's sector and other characteristics; textile plants which dominate our sample are at the high end of this range with an estimated audit cost of roughly INR 40,000. Payment for auditors working in treatment plants thus should cover the average cost of completing an audit. That said, the treatment payment was significantly above the average price of INR 24,000 that control plants report paying for an audit. As we noted above, this status quo price appears to be well below what would be required for auditors to collect the required pollution samples.

For auditor recruiting, at the start of each year all GPCB-certified Schedule II auditors were solicited for their interest to participate in the treatment. In both years interest was oversubscribed, relative to the number of treatment plants. (Auditor interest in the program likely reflected the fact that it offered better working terms, including payments which were in the high range of the market.) Consequently, at the beginning of each year, auditors were randomly allocated a number of plants in proportion to their capacity, measured by number of certified audit teams, which was predetermined.

Auditors could use auditing capacity not allocated in the treatment to conduct audits in the control group. Thus, some of the audit firms were working under two very different sets of incentives at the same time, and we exploit this variation. In the first year, out of 42 audit firms, 24 worked in control only, 9 in treatment only, and 9 in both. In 2010, out of 34 audit firms, 7 worked in control only, 12 in treatment only, and 15 in both.[10]

Backchecks. A randomly selected 20% of auditor plant readings in the treatment were backchecked in the field by technical staff of independent engineering colleges. (These colleges were certified as Schedule I auditors and hence would never directly audit the Schedule II plants in our sample.) Backchecks measure the same pollutants and in the same manner as audits; since they are Schedule I auditors, backcheckers have much experience collecting and analyzing pollution samples themselves. The median backcheck occurred ten days after an audit visit in the treatment. Auditors were aware of the possibility of being backchecked, and

---

[10]The increase of auditors working in treatment only or both in the second year comes from the fact that, in the first year, some auditors were not able to participate because they had already reached their capacity when the program was announced.

knew that results would be used for quality control, although in year one no sanctions for poor performance were specified in advance and all auditors were paid the same regardless of accuracy.[11]

Incentive Pay. In year two, an explicit incentive pay for auditor accuracy, as measured by backchecks in the treatment, was added to the basic set of reforms. Incentive payments used a formula that was first applied to auditor readings in year one to demonstrate to each auditor how accuracy was measured. The pay formula first calculated the difference δ_p between audit report pollution concentration readings and backcheck readings, normalized by the standard deviation of backcheck readings, for each pollutant p. It then averaged the scores for six water and three air pollutants into indices for each media and created the overall measure of auditor quality as the average of these two:

$$\Delta_{Water} = \sum_{p \in Water} \delta_p, \qquad \Delta_{Air} = \sum_{p \in Air} \delta_p, \qquad \Delta_{All} = (\Delta_{Water} + \Delta_{Air})/2.$$

Auditor readings that matched backchecks exactly would thus mean an index value of $\Delta_{All} = 0$, while a value of one means that the weighted average auditor reading exceeded the weighted average backcheck reading by one standard deviation.

Auditors were grouped into three payment categories based on this summary of the difference between their reported readings and the backcheck readings. The least accurate quartile of auditors was paid INR 35,000 per audit. The next least accurate quartile received INR 40,000 per audit, and the most accurate half was paid INR 52,500 per audit. The bonus scheme therefore maintained the average pay of INR 45,000 from year one.

D How Does the Treatment Change Auditor and Plant Incentives?

Under the status quo market structure, several factors likely contributed to auditor misreporting. First, plants could shop for an auditor who would provide a favorable report and condition payments on the contents of that report. In such a market, an audit firm has an incentive to build a reputation for leniency to make it more likely that it would be hired by the same or another plant in the future. This market structure also rewarded cost-cutting by

---

[11]At the end of year one, GPCB received aggregated reports that summarized the accuracy of auditors and the ranges that determined the bonuses in year two. This format did not give them the information necessary to levy sanctions or penalties against specific auditors.

auditors, since they could offer plants reports indicating compliance with the standards at the lowest price if they skimped on data collection, which is anyway unnecessary if the contents of the report are agreed on beforehand. Indeed, this possibility is consistent with data that showed the equilibrium audit price in the control or status quo is below the estimated cost of collecting samples and conducting laboratory tests. Working against these incentives for auditors to misreport, the GPCB has the power to decertify auditors found to file false reports.

The treatment increased auditors' incentives to file accurate reports for several reasons. First, assignment to a plant by an external authority and a fixed pay structure meant that auditor compensation was independent of what it reported. This independence reduced auditor incentives to report a plant as compliant in order to maximize current and future payments from the plant. Second, the introduction of backchecks increased auditor monitoring and likely raised the perceived likelihood that, although not explicitly part of the experiment, the regulator would disbar auditors who submitted false reports (or at least not assign them to treatment plants in year two). In equilibrium, this also raises the bribe a plant would have to pay the auditor to induce false reporting. The introduction of incentive pay in the second year, which explicitly rewarded accuracy, enhanced this effect.

Third, the level of pay was fixed to be high enough for auditors to cover the costs of conducting an audit, including collecting and analyzing the air and water samples. This increase in pay relative to the status quo market price was arguably necessary to induce auditors to do the work. However, the higher payments, relative to the control market pay, may have interacted with increased monitoring to further incentivize accurate reporting in the treatment, if treatment auditors decided to report more accurately because they had more to lose if disbarred from auditing. This efficiency wage channel might have been weakened by two factors, however. First, the experiment was limited to a two-year horizon, reducing the incentive to behave well to stay included. Second, the increase in profits for an auditor who switched from not even measuring pollution, in the status quo, to collecting and analyzing pollution samples, in the treatment, was likely small.[12]

The different treatment elements are complementary. In particular, without the threat

---

[12]We estimate that doing the work in the treatment would cost INR 6,000 in travel and around INR 20,000 in sample collection and analysis charges for a total INR 26,000 increase in costs. This is larger than the increase in INR 21,000 in average payment to an audit firm in the treatment, relative to the control. Of course, there was heterogeneity in these changes in costs and payments, so for some auditors, profits were higher if they chose to report accurately in the treatment if they chose not to in the control.

of monitoring via backchecks, just setting a higher level of pay could be easily undermined by side payments. The ease of making side payments is greater without random assignment. In addition, absent random assignment, backchecks alone provide weak incentives since a reputation for leniency would remain valuable for any auditor seeking to be rehired. And to close the loop, random assignment alone is insufficient if the price is not regulated to ensure that it covers the costs of performing an audit.

Finally, improvement in auditor incentives to truthfully report plant pollution should have increased plants' incentives to reduce pollution levels. Specifically, treatment plants will respond to more accurate reporting by increasing abatement activities if the expected cost of GPCB penalties is sufficiently high relative to the cost of abatement.

III Data and Summary Statistics

In this section we describe the multiple data sets used in the analysis, and provide summary statistics on our plant sample.

A Data Sources

The key outcomes of interest are accuracy of auditor reporting and the pollution response of plants. Two data sources are used to measure accuracy. First, audit reports filed with GPCB in 2009 and 2010. These reports cover a mandated set of water pollutants (i.e., biochemical oxygen demand, chemical oxygen demand, total dissolved solids, total suspended solids, ammoniacal nitrogen) and air pollutants (sulfur dioxide, nitrogen oxides, and suspended particulate matter) described in Table A1.[13] Both treatment and control plants followed the same practice of scheduling audit dates in advance of actual auditor visit to collect air and water samples. This opens the possibility that plants alter short-run behavior to ensure low pollution readings, for example by running air pollution control equipment or reducing boiler load when pollution is to be measured. We expect that, if anything, this should be more likely in the treatment, reducing estimates of how much the treatment increased pollution readings in audits.

---

[13]Auditors record these pollutants at various stages in the treatment process and with respect to different systems in the plant. We use pollutant concentrations at the final outlet from the plant for water samples, as these are the readings with a direct impact on the environment and are therefore most closely attended by both auditors and GPCB. For air, we focus on boiler-stack samples for the widest comparability across the sample, as most plants have boilers.

The second source of data for auditor accuracy is the backchecks, which were conducted in a sample of treatment plants throughout 2009 and 2010. These backchecks were conducted on the same pollutants and locations as the audits, and were scheduled to occur close to the audit visits. Auditors and backcheckers use the same technology and standardized procedures to measure pollution. Treatment backcheck data were complemented by "midline" backchecks after the third season of audit visits in year two in both treatment and control groups, using the same process and agencies. Auditors were not informed about the probability of backchecks in the control plants, and these backchecks were neither transmitted to GPCB nor used to compute any specific auditor payment.[14] The midline data allow for a direct measurement of the comparative accuracy of auditors working in treatment and control plants, measured as the difference between auditor and backcheck readings.

The third source is data on plant pollution emissions which comes from an endline survey conducted from April through July of 2011, approximately six months after the last audit visits in the treatment group. Pollution sampling in the survey was mostly conducted by the agencies that did backchecks and included the same pollutants as discussed above. The endline analysis includes all plants in the audit sample (treatment and control). Overall, we collected 2,953 pollution samples from 408 plants in the study sample, an average of 7.2 pollutants per plant.[15] Attrition in the endline survey was balanced across treatment and control groups.[16]

Finally, we utilize GPCB administrative data to better understand plant and auditor incentives in the status quo and their response to the treatments. These data cover GPCB's plant inspections for plants in the audit sample between 2008 and 2011. We link 8,627 GPCB inspections, with their accompanying pollutant readings, to 4,269 subsequent actions or penalties as documented by the regulatory files on each plant. The actions range in severity from

---

[14]The midline sample was drawn from treatment and control groups in order to maximize the number of plants covered by auditors working simultaneously in both the treatment and control groups, and to use information on the dates of audit visits to conduct backchecks that were as close as possible to the date of the initial visit. In the treatment group, the sample plants were randomly selected stratified by auditor. In the control group, the sample plants were drawn non-randomly in order to ensure coverage of auditors working in the treatment simultaneously. Priority for the survey was first given to plants that previously submitted an audit report by an auditor working in the treatment group. The control sample was completed by adding those plants for which auditors submitted a date for the audit visit and finally by adding randomly selected plants for which auditors had not submitted a date.

[15]The audit intervention was conducted concurrently with another treatment, an increased frequency of regulatory inspection, which directly affected plants, but not auditors. For ease of interpretation of our pollution results, we restrict the pollution sample to the subset of audit sample plants not subject to the other experimental intervention.

[16]While a somewhat greater share of plants were surveyed in the treatment group, 88.8% versus 83.8% of control plants, the difference of 5.09 percentage points (standard error 3.16 percentage points) in these rates is not statistically significant. Most attrition was attributable to plant closure.

letters of warning up to orders that the plant be disconnected from electricity supply.

B   Summary Statistics

Despite being *prima facie* audit-eligible, some plants in the sample failed to submit audits in year one and/or year two. There were several reasons for these instance of non-submission: GPCB had judged them not-audit-eligible in the past, they changed products since data used to determine eligibility was compiled, they protested their audit-eligible status, they chose not to submit and incur the risk of a penalty, or they closed. Table I shows that treatment and control plants were about as likely to submit audit reports. Treatment plants were slightly less likely to report in the first year (70% versus 74%) slightly more likely in the second year (70% versus 64%), though neither difference is statistically significant. These rates of submission are comparable to those in 2008, the year prior to the experiment (72% in treatment plants and 69% in control plants). The treatment, therefore, does not appear to have induced more plants to submit audit reports.

Though balanced in aggregate, there may be heterogeneity in which plants submit across the treatment and control groups. We use GPCB administrative data from prior the experiment to estimate probits for audit submission.[17] The results are presented in Online Appendix Table 1. We fail to reject the null hypothesis that the effect of the plant characteristics on submitting an audit during the experiment is equal in the treatment and control groups; the exceptions are that treatment plants are relatively more likely to submit in the second year of the experiment and that, relative to the treatment group, textile plants in the control group are relatively more likely to submit audit reports (though, across the board, treatment plants have higher submission rates).

We take multiple approaches to address the possibility that selection bias influences the estimates of the treatment on audit reporting. First, several specifications control for true pollution levels measured in the backchecks; in these specifications, the source of any selection bias would have to be on a dimension other than pollution emissions to influence the results. Second, we implement DiNardo et al.'s (hereafter DFL) reweighting scheme, using administrative data from before the experiment, so that the distribution of audit submitters' observables

---

[17]We use the following variables to predict submission: whether a plant is located in an industrial estate, whether it is in the textiles sector, whether its effluent flows to a common treatment plant, the amount of wastewater it generates, whether it submitted an audit before the experiment, whether it was cited by the regulator for a violation before the experiment, and a dummy for the second year of the experiment.

resembles that of all plants (DiNardo et al., 1996). The reweighting is based on the results of the estimation of a probit for submission as a function of baseline plant characteristics (without the treatment indicator or its interaction with the characteristics), shown in column 2 of Online Appendix Table 1. The distributions of predicted submission from the model have a broad support that is common across plants that do and do not submit. For completeness, we also use the standard selection correction approach (Heckman, 1979).

Table II presents summary statistics and a randomization check using baseline characteristics of plants in the study sample, among plants that did submit a report in either year. Plants submitting audits are similar across treatment and control. In Panel A we consider plant characteristics. Most sample plants are textile factories eligible for environmental audit due to high effluent volume. Textiles is the largest registered manufacturing sector by employment in India and second largest in Gujarat (Authors' calculation, ASI (2005)). Both treatment and control plants have similar pollution potential as measured by effluent quantity and type of fuel used. Treatment plants are 10 percentage points less likely to have a bag filter, a type of air pollution control equipment, installed, but are similar to control plants with respect to other air pollution control equipment such as cyclones and scrubbers.[18]

Table II, Panel B reports on the interactions of sample plants with GPCB in the year prior to the study by treatment status. A little over 80% of this group submitted an audit report in the year prior to the study's initiation. Roughly the same fraction was inspected and over 40% of sample plants were mandated to install equipment.[19] Based on the GPCB records, a significant number of sample plants have been subject to costly regulatory actions: around a quarter were cited for any type of violation and fully 10% of plants, in both treatment and control, had their utilities disconnected at least once. About three percent were required to post a bank guarantee (i.e., bond) against future environmental performance. These variables are balanced across treatment and control plant. Consistent with being less likely to have a bag filter, treatment plants were more likely (at the ten percent level) to have received a citation for an air pollutant violation than control but equally likely for water pollutants and all citations together.

---

[18]In the same comparison of covariate balance for the full study sample, unconditional on submission (not shown), bag filter installation remains the only difference between treatment and control plants significant at the five-percent level.

[19]This 40% is atypically high; during the prior year GPCB had conducted an air pollution control equipment installation campaign that affected many sample plants.

In summary, treatment and control plants that submitted an audit had similar interactions with the GPCB in the year prior to the experiment. Furthermore, it is evident the regulator has a meaningful track record of action so the information reported by auditors had the potential to change plant behavior.

Finally, both the midline sample (the subset of plants that were backchecked during the midline) and the sample of plants audited by auditors working in both treatment and control group (relevant for fixed effect specifications) remain well-balanced along the observables shown in Table II.[20]

IV   Econometric Approach and Results

The results are divided into three parts. First, we use data on control plants to examine the status quo audit market. Second, we estimate the effect of the treatment on auditor reporting behavior as measured by pollution levels in audit reports and corresponding backchecks. Third, we measure how two years of altered auditor incentives influenced plant polluting behavior.

A   Auditor Reporting in the Control Group

A unique feature of this paper's setting is that we observe both auditor reports and an independent measure of the underlying pollution—backchecks of the same pollutants. This allows us to assess whether the market was producing reliable information for the regulator.

Figure 1, Panel A plots the distributions of concentrations of Suspended Particulate Matter (SPM), an important air pollutant, from audit reports and backchecks for the control group. In each distribution a vertical line marks the SPM regulatory standard of 150 mg/m^3, and gray shading shows the share of the probability mass that falls between 75% of the standard and the standard, the zone where we might expect measurements to fall if auditors are trying to show firms as compliant without being too conspicuous.

We observe several striking facts. The top half of Panel A shows that auditors report the vast majority of plants (93%) as compliant with the SPM standard, and there is a high concentration of readings just below the standard; 73% of plants have auditor-reported SPM concentrations in the narrow range from 75% to 100% of the pollution standard. Such bunching below the limit is consistent with targeted misreporting by auditors but also with two other

---

[20]The one additional imbalance in both samples is that the treatment plants generate more wastewater.

explanations: plants may minimize abatement costs subject to the constraint of not exceeding the standard, although emissions includes some randomness which makes it difficult to hit the standard exactly, or regulatory capture may cause the standard to be deliberately set at a level that allows plants to narrowly comply.

The second half of Figure 1, Panel A uses backcheck data to test these possibilities. Our second finding is significant dispersion in the backcheck distribution for SPM. Only 19% of the plants have readings in the range that covers 75% to 100% of the standard, which is 54 percentage points less than in the audit distribution. Further, substantial probability mass exceeds the standard: 59% of backcheck readings exceed the standard compared to just 7% in the audit distribution. There are also more very low backcheck readings. This increase in the left tail of backchecks, relative to audits, may be explained by the cost of measuring air pollution concentrations, making it cheaper for auditors to report narrow compliance by default than to properly sample and document a very low reading. In summary, the audit and backcheck distributions together provide striking evidence that, at least in the case of SPM, auditors fabricate data in order to falsely report plants as narrowly comply with the regulatory standard. Backchecks do not similarly cluster beneath the standard, belying the two alternative explanations for the pattern of auditor reporting.

Next, we use regression analysis to check whether the difference in pollution readings between audit reports and backchecks holds across the full range of pollutants, continuing to use the control sample only. Table III reports results from OLS regressions on the stacked data, including both backchecks and audit readings, of the form:

$$\mathbf{1}\{Compliant\}_{ij} = \beta_1 \mathbf{1}\{AuditReport\} + \alpha_r + \epsilon_{ij}. \tag{1}$$

In Panel A, $\mathbf{1}\{Compliant\}_{ij}$ equals 1 for readings of pollutant i from plant j that are between 75% and 100% of the regulatory standard, and in Panel B, $\mathbf{1}\{Compliant\}_{ij}$ equals 1 if the reading is below the standard. The coefficient of interest is β_1 on the dummy $\mathbf{1}\{AuditReport\}$; each plant appears twice, as the data is pooled across matched pairs of audits and backchecks, so β_1 indicates how likely a pollutant report is to be compliant in audits relative to the omitted category of backcheck readings. The specification includes fixed effects α_r for the regions $r \in \{Ahmedabad, Surat\}$, on which treatment assignment was stratified, and we cluster stan-

dard errors at the plant level to account for correlation in the errors for different pollution samples taken at the same plant (on average we have seven pollutants per plant).

Table III, Panel A, column (1) shows that across all pollutants, pollution levels in audit reports are 27 percentage points more likely to show narrow compliance than backchecks. This large increase is against a baseline of just 10% of backcheck readings that fall in the 75%-100% range. As to whether the reading is compliant, Panel B, column (1) shows that across all pollutants 55.7% of backchecks are below the relevant regulatory standard. The coefficient for audit reports indicates that an additional 28.8% of readings from audits are falsely reported as below the standard. This finding is evident for both air and water pollutants.

Taken together, Figure 1 and Table III suggest that neither plant abatement behavior nor regulatory capture underlies the clustering observed just beneath the standard. Rather, the auditors frequently failed to truthfully report pollution readings that would provide regulators with the information necessary to act against the more than 40% of the plants in violation of the regulatory standard.

B   The Effect of the Treatment on Auditor Behavior

B.1   Truth-telling about Regulatory Compliance

Next, we examine the impact of the treatment on auditor reporting. Figure 1, Panel B shows the distributions of SPM concentrations as reported by audits and backchecks for treatment plants during the midline.

The top half of Panel B reveals that in the audit treatment group 39% of readings are within 75% to 100% of the standard. This is far below the 73% that were in this range in the control group audit reports. Further, the support of the audit and backcheck distributions is much more similar among the treatment plants than it was among the control. The similarity is especially evident for readings above the standard, which were very sparse among control audits. While the treatment increased truth-telling by auditors, it did not end false reporting. The 39% share of audit readings in the shaded area still exceeds the 14% share in the backcheck distribution.

To examine the impact across pollutants we pool samples of *all* pollutant readings from audits and backchecks for plant j, both collected in the final season of year two. We estimate

ordinary least-squares regressions of a difference-in-difference form:

$$\mathbf{1}\{Compliant\}_{ij} = \beta_1\mathbf{1}\{AuditReport\} \times T_j + \beta_2\mathbf{1}\{AuditReport\} + \beta_3 T_j + \alpha_r + \epsilon_{ij}, \qquad (2)$$

where $1\{Compliant\}_{ij}$ and α_r are defined as before, and standard errors are clustered at plant level. β_3 controls for any difference in true compliance, as measured by backchecks, across treatment and control. There are two coefficients of interest. The first is β_2, which measures how much more likely an audit report is than a backcheck to be compliant in the control group. The second, β_1, measures how treatment changes this difference between pollution levels in audit reports and backchecks or the frequency of false compliance reports.

The results are in Table IV. The treatment increased truth-telling about compliance with the regulatory standard across the full set of pollutants. In Panel A, column (1), audit reports are 19 percentage points less likely to falsely report a reading in the narrow range of 75-100% of the standard in the treatment than in the control. This is a reduction of 69%, relative to the control mean. Similarly Panel B, column (1) reveals that audits in the treatment are 81% (-0.234/0.288) less likely to falsely report compliance with the standard. These effects are evident separately for both water pollutants, in column (2), and air pollutants, in column (3), with the air effects larger in magnitude.[21]

The effect of the treatment on reporting of compliance and narrow compliance is basically unchanged after applying the DFL reweighting for selection.[22] This approach reweights observations in the sample of plants that submit by the odds ratio of non-submission, to approximate the observable characteristics of all plants. As the results are qualitatively unchanged with this reweighting, the difference in compliant pollution levels in audit reports across the treatment and control groups does not appear to reflect selection bias in audit reporting. It is also possible that the backcheck compliance measure in these regressions already controls for selection.

Figure 2 summarizes how the audit treatment changes the density of the audit report pollutant distribution. To normalize pollution readings relative to the regulatory standard,

---

[21]For comparability, we omit pH from all panels as both high and low readings can be harmful, unlike for the other pollutants. In Panel A the results are unchanged if we include pH, for which the standard is a range rather than a maximum limit.

[22]For example, the coefficients of interest (standard error) on Audit report $\times$ Treatment corresponding to the first row of Table IV Panel B, where compliance is the dependent variable, are -0.190 (0.041), -0.158 (0.049) and -0.241 (0.067).

we subtract the standard for each pollutant, and divide by the pollutant standard deviation in backchecks. The horizontal axis therefore marks the number of standard deviations above or below the regulatory limit. We then fit 40 separate regressions for indicator variables for a pollutant reading belonging to a particular 0.05-standard-deviation-width bin on an indicator for an audit report being from the audit treatment. Weakly negative (positive) values indicate that treatment auditors are less (more) likely to report readings in that bin. The treatment dramatically reduces the amount of mass just beneath the standard. The treatment auditors instead report significantly more readings in the bins more than 0.1 standard deviations below the standard and especially in the bins stretching up to 0.5 standard deviations above the standard. Note that for all pollutants together, unlike for SPM, we do not see systematic increases in pollution reports at levels well below the standard.[23]

The treatment, then, succeeded in greatly increasing the frequency with which auditors truthfully report pollution readings above the relevant standard. The problem in the status quo was systematic distortion and the treatment created the conditions necessary for auditors to choose to report true pollution readings with much greater frequency.

B.2   Truth-telling about Pollution Levels

Though the compliance threshold is discrete, the regulator and policy makers are also interested in the level of continuous pollution emissions, since this ultimately affects public health. We therefore examine next the reported concentrations of pollutants in audit reports. We standardize pollutants by subtracting the mean and dividing by the standard deviation of the same pollutant reading among backcheck samples. Thus, a one-unit change in the pollution measures can be interpreted as a change of one standard deviation. We estimate:

$$y_{ijt} = \beta T_j + \alpha_r + \alpha_t + \epsilon_{ijt}, \tag{3}$$

where y_{ijt} is the standardized audit report reading of pollutant i in plant j taken in year t. The set-up is similar to equation 2, except the data are drawn only from audit reports and cover multiple years. The sample includes all plants that submitted a report in at least one

---

[23]Across all pollutants, a significant 4 percentage points more audit readings in the treatment than in the control are below 75% of the regulatory standard. This difference shrinks to 2 percentage points, and is no longer significant, after conditioning on backchecks below 75% of the standard.

year. (The advantage of using only audit data and not backchecks is data availability for both years of experiment.) To account for annual variation in pollution, α_t are fixed effects for the year $y \in \{2009, 2010\}$. We cluster standard errors at plant level.

The parameter of interest, β, measures whether auditors' reported concentrations differ in treatment and control groups. In specifications with auditor fixed effects, α_a, β is identified from cases where the same auditor works under both treatment and control market structures. This specification is noteworthy, because the resulting within-auditor estimates are freed from concerns that the estimates are due to better auditors selecting into the treatment group.

Table V measures the effect of the treatment on auditor reporting, where odd-numbered columns are the base specification and even-numbered columns include auditor fixed effects. The three panels are for three different measures of auditor reporting. Panel A presents estimates from specifications where the outcome is a dummy for the audit report being below the regulatory standard (compliance). The column (1) results show that treatment audit reports were 15 percentage points less likely to report that a pollutant reading was in compliance. The column (2) specification includes auditor fixed effects, so that the treatment effect is estimated from variation in reporting across the treatment and control groups from audit firms working in both groups; the point estimate is slightly more negative than before, with treatment audit reports 18 percentage points less likely to report compliance. The remaining columns suggest a greater decline in reported compliance for water relative to air pollution readings, although the differences are not large in the context of the standard errors.

Panel B examines continuous measures of reporting. In column (1), the mean audit report reading for all pollutants in treatment plants is a significant 0.103 standard deviations higher than the mean report in the control. The addition of auditor fixed effects, as shown in column (2), increases the point estimate of the audit treatment to 0.131 standard deviations, but the estimates with and without auditor fixed effects remain statistically indistinguishable. It is noteworthy that the effect of the treatment remains constant, even when the identification comes from within-auditor variation, showing that the same audit firms report differently under the two regimes. The coefficients on the audit treatment are similar for both water and air pollutants considered separately, as shown in columns (3) through (6).

The magnitude of these effects is substantial. Consider the estimate in column (2) of 0.131 standard deviations for all pollutants, which is roughly of the same size as the effect

for Biochemical Oxygen Demand (BOD) estimated alone (not shown). For plants where the effluent did not go on for further treatment, the standard deviation of BOD in backchecks in final-outlet samples was 203 mg/l and the mean 191 mg/l, as against a concentration standard of 30 mg/l. An effect of size 0.131 standard deviations thus represents 26.7 mg/l for BOD, or 89% of the standard. The mean and standard deviation of SO_2 readings in backchecks were 64 and 108 parts per million (ppm), respectively, as against a standard of 40 ppm. A 0.131 standard deviation movement is thus 35% of the standard. Consistent with our earlier discussion of the distributions, these specifications show that the change in pollutant reports shifts several plants from compliance to non-compliance and is economically significant, in representing a meaningful increase in reported pollution.

One concern with using auditor reports of pollution concentrations as an outcome is that they combine true pollutant emission, measurement error and potential auditor manipulation of the results. If the treatment caused plants to reduce emissions, then auditor reports conflate changes in auditor reporting and pollution emissions such that the estimate of β will understate the effect of the treatment on auditor reporting. We investigate this possibility by returning to the midline data, using the difference between audit and backcheck readings of the same pollutant,

$$y_{ij}^D = y_{ij}^{Audit} - y_{ij}^{Backcheck},$$

as an outcome. This difference controls at the plant level for any possible effect of the intervention on actual pollution levels, but is only available on a comparable basis for the midline sample. Readings are matched on pollutant i, plant j, sampling location (boiler stack or final outlet) and date. When backchecks are treated as the truth, negative values indicate underreporting as they show auditors reporting lower readings than the true emissions concentrations.

The results are reported in Table V, Panel C. The difference between audit and backcheck readings is -0.304 standard deviations in the control, reflecting the observed negative bias in auditor reporting. The treatment coefficient of 0.210 standard deviations (standard error of 0.073 standard deviations) indicates that auditors working in treatment plants report substantially higher readings, after accounting for any changes in actual emissions across the treatment and control plants. Indeed, this finding implies that the treatment erases nearly 70% of the underreporting observed in the control group. The treatment coefficient in audit differences

is smaller in the column (2) specification with auditor fixed effects and has a larger standard error; however, the null of zero can be rejected with a p-value of 0.12 and the point estimate in column (2) is within one standard error of the column (1) estimate, from the specification without fixed effects.

The Table V results are qualitatively unchanged by the application of the DFL correction for the selection of plants into submitting audit reports. In Panel A, the DFL correction leaves the point estimate for all pollutants basically unchanged. In the Panel B specifications for pollutant levels, the DFL correction increases the point estimates for audit treatment for the all and water pollutants specifications. The point estimates in the difference specification of Panel C are about 20% smaller, which is within one standard error of the Table V estimates. Further, the DFL correction increases the standard errors in panels B and C such that the 95% confidence intervals of all of the Table V and corresponding DFL estimates easily overlap.[24] Note that the changes in point estimates are larger in the levels specifications, where selection may be a greater concern since backchecks are not used as a control for contemporaneous pollution.[25] We conclude that correcting for selection into audit submission on observables does not change our findings with respect to auditor reporting.

B.3   Evidence on Possible Treatment Channels

Our treatment spanned two calendar years and included several components—random assignment of auditors to plants, fixed payments from a central pool, and backchecks—each of which may have independently influenced auditor reporting. Additionally, the auditors received bonus pay for accurate reporting in the second year.

All the channels described in Section II.D may have been at play. The sample size was too small to randomly assign each component separately and obtain meaningful estimates of their effects. Moreover, some of the features are not likely to be effective in isolation (e.g., fixed pay from a central pool without random assignment would still leave plants with an incentive to shop for favorable auditors).

---

[24]With the DFL correction, the treatment coefficients (standard errors) corresponding to the specifications in odd columns of Table V, Panel A are -0.150 (0.027), -0.202 (0.035) and -0.035 (0.026). The same coefficients in the level regression of Panel B are 0.230 (0.114), 0.314 (0.168) and 0.064 (0.025), and in the difference regression of Panel C are 0.165 (0.119), 0.117 (0.152) and 0.241 (0.165).

[25]Heckman (1979) gives an alternative approach to selection correction. Using that approach, and the same variables as used to predict submission in the DFL approach, all coefficients are nearly identical to the main results in both Panels B and C. The Heckman selection correction approach cannot be applied to discrete outcomes (e.g., compliance in Panel A), because it assumes a normally distributed error.

Since financial incentives were introduced only in year two, there is a possibility to evaluate them. The key challenge is that this variation is non-experimental. Over time, arguably because of our experiment, treatment plants were reducing their emissions, relative to control plants (on this, see Section C below). The impact of the reform on reporting behavior may also have had a time effect. Consequently, we need to accurately specify the time trends and assume no other factors that influence audit reports changed discretely between the two years in the treatment relative to the control. To measure changes in reporting over time, our sample is the full sample of audit reports in both years, and our outcome variable, as we lack backcheck reports for control plants except during the midline, is the reported pollution reading from audits only.

Figure 3 reports the trends in audit reporting over time. The figure plots the mean standardized pollution reading in audit reports for all pollutants, by year and each of the three audit seasons, separately for plants in the treatment and control groups. Pollution in audit reports in the control group is roughly flat at -0.3 standard deviations across the two years. In the treatment group, pollution reports are generally well above those in the control. Over the course of the first year, pollution levels in treatment audits decline with each season, and over the second year they also decline, somewhat less sharply. The most striking feature of Figure 3 is the sharp increase in reported pollution readings in the treatment at the beginning of year two when financial incentives were initiated. In other words, the figure is consistent with the financial incentives having an independent effect on the accuracy of auditor reports.

This finding is documented in Table VI. Column (1) duplicates the base specification of Panel B, Table V for reference, showing the coefficient on the year two (2010) indicator. Column (2) includes as additional covariates a linear trend in fractional years since January 1, 2009, the interaction of the linear trend with a treatment indicator, and the interaction of the year two dummy with the treatment indicator. The parameter of interest is associated with this last variable, and it tests for a discrete relative change at the beginning of year two in auditor-reported pollution readings among treatments.

Three findings in Table VI echo those of Figure 3. First, as already seen in Panel A of Table V, the year-one treatment features cause higher auditor reported pollution levels (indeed, the figure indicates that this effect is evident from the very beginning of the experiment). Second, the interaction of the treatment indicator and the time trend, which is fractional

25

in the date of visits and therefore runs continuously from zero to two, shows a significant downward trend in reported pollution among treatment plants. This finding foreshadows the finding in the next section that the treatment caused plants to reduce their emissions. In contrast, there is no declining trend in the control group.

Third and most importantly, there is a discrete increase in pollution levels in audit reports in the treatment group, relative to the control, at the beginning of year two that interrupts the relative decline of reports in the treatment. Specifically, treatment pollution levels in audit reports are a statistically significant 0.257 standard deviations higher at the beginning of year two and this exactly coincides with the introduction of the incentive pay scheme. This discrete increase is robust to alternative functional form assumptions about the differential time trends.[26] These results suggest that the incentive pay component of the treatment led auditors to report higher, more accurate audit readings.

Table 3 in the Online Appendix provides tentative additional evidence that the risk of being backchecked also had an independent effect on auditor behavior. Specifically, we ask whether an audit firm that had a client plant recently backchecked reports differently, using data from the midline survey (conducted over the last season of audit visits in the experiment). When this midline began, pollution readings for treatment plants under an auditor had been regularly backchecked for nearly two years, but no backchecks had occurred in the control. The table shows that having recently been backchecked during the midline increases the accuracy of reporting in control plants but not in treatment plants. We find this consistent with auditors updating their beliefs to recognize that backchecks were possible in control plants, and improving reporting in response.

Finally, our results allow us to rule out some channels. The auditor reporting results in regressions with auditor fixed effects, showing that the same audit firms report differently in the treatment and control groups of plants, suggest that higher pay did not directly improve reporting through an income effect, as extra income would have given auditors more resources in both groups. Selection of better auditors into treatment participation and Hawthorne effects at the auditor level are also inconsistent with the within-auditor estimates of a significant treatment effect.

---

[26]The addition of quadratic and cubic trends and their interaction with the treatment indicators causes the discrete jump in reporting in year two to increase to 0.284 standard deviations (standard error 0.116 standard deviations).

C The Effect of the Treatment on Plant Emissions

Given the increase in truth-telling, a natural next question is whether plant abatement behavior responded to more accurate auditor reporting. We expect plants to respond only if truthful reports on high pollution will cost them more, through regulatory sanctions, than pollution abatement.

Our analysis utilizes plant-level pollution emissions data from the endline survey. To avoid sample selection issues, and because the audit treatment may have affected the emissions of all plants, not just those that chose to submit audit reports (and the worst polluters may chose not to submit), our sample is all plants that entered our sample, not just those which filed audit reports.

To measure the effect of the treatment on pollution emissions, we report the results of fitting OLS regressions for pollution outcomes on region fixed effects and a treatment indicator with the cross-sectional endline survey data. The outcome variables are both the continuous pollution outcome and a compliance dummy. Further, we also estimate quantile regressions of the form

$$Q_{y_{ij}|X_j}(\tau) = \beta T_j + \alpha_r + \epsilon_{ij},$$

where $Q(\tau)$ is the τ-quantile of the pollutant concentration conditional on treatment status and regional indicator variables.

Table VII reports the results. In Panel A, the outcome variable is a standardized measure of pollution emissions calculated in the same manner as in the auditor accuracy tables above. On average, the treatment plants reduced pollution by a statistically significant 0.211 standard deviations.[27] This effect is driven by a large decrease of 0.300 standard deviations in water pollutant concentrations, shown in column (2). The estimated effect on air pollution in column (3) is smaller and insignificant. Since the volume of effluent emitted did not change in response to the experimental treatments, these reductions in concentrations represent reductions in the total emitted effluent load—i.e., less water pollution—among treatment plants.

---

[27]The coefficient for all pollutants is due in part to several control plants with very high pollution readings. It decreases in magnitude to -0.143 standard deviations (standard error 0.068 standard deviations) and -0.114 (standard error 0.058 standard deviations) when the readings are top-coded above the 99.5 and 99 percentiles of the pollutant distribution, respectively. There are, however, several pieces of corroborating evidence that these readings are genuine and should not be top-coded. The plants in question have track records of noncompliance, and other pollutant samples collected at the same plants, both before the high endline pollutant readings and at the endline, also show pollution levels far above average.

In Panel B, the outcome variable is whether the pollutant reading is compliant, that is below the regulatory standard. In column (1), compliance is estimated to have increased by a small and insignificant 2.68 percentage points. Columns (2) and (3) indicate that the effects are similar for water and air pollution. The fact that reductions in pollution did not increase compliance significantly suggests that these reductions were concentrated among plants with pollution levels far from the regulatory threshold.

To explore the source of estimated mean pollution reductions, Figure 4 plots treatment effects from quantile regressions of standardized endline pollutant levels on audit treatment and region fixed effects. Quantile effects are estimated from the 0.05-quantile to the 0.95-quantile at 0.05-quantile intervals. The shaded area is the 95% confidence interval. While no individual quantile coefficient is significant at conventional levels, the point estimates show a clear pattern wherein the treatment reduced pollution more at higher quantiles of the pollution distribution. Up to the 0.75-quantile, the point estimates are very close to zero, but from the 0.80-quantile onwards the point estimates sharply decrease to less than -0.5 standard deviations at the 0.95-quantile. It is evident that the mean reduction in pollution is largely a consequence of reductions in the right tail of the pollutant distribution.

This pattern of pollution reductions appears to be related to the GPCB penalty structure. Figure 5 investigates how GPCB's regulatory actions correspond to the degree of observed pollution violations. We consider GPCB's regulatory follow-up to its own inspections, as observed during the three years beginning the year before the study and running through its end. The figure plots how often GPCB takes different types of regulatory actions in response to observing pollution at different levels above the pollution standard, measured during GPCB inspections of audit sample plants. We classify actions in four categories of increasing severity from the bottom (dark grey bars) to the top (light grey bars).

The severity of GPCB's actions increases monotonically in the amount by which the standard is exceeded and sharply only for very high violations, suggesting that GPCB does not treat all exceedances of the standard equally. For example, plants polluting at above the standard but less than 1.5 times the standard receive the most severe actions (i.e., closure warning and disconnection) in less than 10% of the cases, and even plants between two and five times the standard less than 40% of the time. In contrast, plants with readings higher than ten times the standard receive these actions more than 70% of the time.

Thus the most costly sanctions are, in practice, reserved for the right tail of the pollutant distribution. This relationship between high pollution concentrations and likely penalties is a logical explanation for why the treatment, which broadly improved the quality of auditor reporting, induced only high-polluting plants to clean up. All else equal, these firms have a much higher expected fine once the auditors' reports are accurate.[28]

V Conclusion

This paper reports on a large-scale reform to the third-party environmental audit system in the Indian state of Gujarat, conducted in collaboration with the environmental regulator. The goal was to change the structure of the audit market to incentivize accurate reporting and ultimately pollution abatement. The treatment consisted of random assignment of auditors to plants, payments to auditors at a fixed rate from a central pool (rather than the plant), random backchecks of auditors and, in the second year, a bonus for accurate reporting.

There are three primary findings. First, the status quo audit system appears corrupted, with auditors systematically underreporting the pollution emissions of control plants at levels just below the regulatory standard. Second, the treatment greatly increased the accuracy of auditor reporting, viewed in terms of compliance, levels, or differences with backchecks. This finding is robust to the inclusion of auditor fixed effects which provides estimates based on a comparison of the behavior of the same auditor working under both the treatment and control/status quo market structures simultaneously. Third, treatment plants reduced their pollution emissions. This decrease largely comes from water pollutants which were the original spur for the development of the audit scheme and which remain regulatory priorities. The reductions are concentrated entirely in highest polluting plants, which face the greatest risk of regulatory sanction.

We attribute these results to the package of reforms that altered the market structure for audits in the treatment group. Although we cannot separately identify the influence of each treatment component, we make several observations on what underlies the treatment effects. First, where we have some variation, the non-experimental evidence suggests that incentive

---

[28]Compared to the results in Panel A of Table VII, the effect of the treatment on pollution emissions is smaller and insignificant in the selected sample of plants that submitted audits. This reduction in the treatment effect supports the idea, introduced in the audit reporting results above, that especially dirty plants do not submit audit reports in the control group. Indeed, the GPCB's penalty structure that reserves the harshest penalties for plants that greatly exceed the regulatory standard sets incentives that induce this selective audit reporting.

pay increased the propensity of auditors to accurately report. Second, fixed effects estimates show that the treatment causes within-auditor changes in reporting. These estimates suggest that auditor-level effects, such as income or Hawthorne effects, do not drive the treatment impact. Third, economic intuition and existing evidence from the corruption and monitoring literature suggest that higher auditor pay without the other treatment components would not have changed auditor behavior much, given the absence of monitoring in the status quo. Without some monitoring, rational auditors would pocket any increase in payments or kick it back to their clients rather than improve audit quality. In combination with backchecks, however, efficiency-type wages may have contributed to strengthen auditors' incentive to report accurately if they expected inaccurate reporting would lead to disbarment from the treatment or outright decertification, and hence a loss of future auditing income.

These results are encouraging in the context of environmental regulation in India. A critical regulatory challenge is gathering accurate information in the face of agency problems either in third-party reporting or within the regulator itself. Like the environmental audit system in Gujarat, the national system for Environmental Impact Assessments (EIA) has foundered on this problem because, as put by a former Minister of Environment, "[T]he person who is putting up the project will be preparing the report" (The Hindu, 2011). Our findings demonstrate that judicious reforms can enable the regulator to collect accurate information on pollution emissions. Further, at least in cases where the regulator is known to assign strict penalties, the provision of this information to the regulator causes the plants most likely to be penalized to reduce their emissions. Of course, we only measure the impact of these reforms over two years. It is possible that, in the longer run, results may be more muted, for example if the backcheckers began colluding with auditors.

Was the reform worth the extra cost? A determination of whether the reform package brought about a net social gain requires estimates of the costs of increased monitoring through backchecks, auditor effort and pollution abatement and of the benefits of lower pollution. The Online Appendix makes a tentative effort at such an analysis and suggests net benefits. We estimate gross costs to be around USD 1,300 per plant and, tentatively, gross benefits to be USD 7,300 per plant, for a net social gain of USD 6,000 per treatment plant. That is, reductions in pollution more than offset additional costs incurred by plants, auditors and the regulator. These particular numbers are far from definitive and rely on several strong

assumptions; in particular the numbers for marginal damages may not be right in the context of this experiment and our estimate of abatement costs is incomplete, including only abatement capital.

We believe that the core problem—that auditors face a conflict of interest, or, at least, poor incentives to tell the truth—exists in all third-party audit markets. Indeed, we are unaware of a single market wherein the audited party does not directly hire the auditor, and while there are often provisions for some monitoring, those often appear to be fairly weak. At least in the case of the Gujarat environmental audit market, we have documented both that this market structure produces very unreliable audit reports and that a politically and logistically feasible reform can greatly improve market outcomes.

Our findings are likely of broader relevance, though the exact nature of any audit reform will reflect a particular market's status quo functioning and institutional details. For example, in complex settings with a fixed cost of establishing an auditor-client relationship, such as the cost of a financial auditor learning a client's books, assignment or rotation of auditors to firms for periods longer than one year may be preferable. The strength of reputation effects is another salient difference, with concentrated market structures giving better incentives for high-quality audits even when auditors are hired by the firms on which they report.[29] That said, recent history in the United States suggests that in the absence of regulatory oversight even very concentrated markets do not suffice to keep the quality of reporting reliably high.

All of this underscores the value of designing third-party audit markets so the first-order incentive is for auditors to tell the truth. Our study shows that, at least in one instance, it was possible to do so.

References

ASI (2005). Annual Survey of Industries. *Ministry of Statistics and Programme Implementation, Govt. of India.*

Bazerman, M., Loewenstein, G., and Moore, D. (2002). Why good accountants do bad audits. *Harvard Business Review*, 80(11):96–103.

---

[29]Duflo et al. (2013) finds some evidence of a reputation for more accurate reporting yielding higher audit prices even in the Gujarat environmental audit market studied here.

Bazerman, M. H., Morgan, K. P., and Loewenstein, G. (1997). The impossibility of auditor independence. *Sloan Management Review*, 38(4):89–94.

Becker, G. S. and Stigler, G. J. (1974). Law enforcement, malfeasance, and compensation of enforces. *J. Legal Stud.*, 3:1.

Bhattacharyya, S. (2011). The clean development mechanism. *Energy Economics*, 5:623–645.

Bolton, P., Freixas, X., and Shapiro, J. (2012). The credit ratings game. *The Journal of Finance*, 67(1):85–111.

Boyd, C. E. (2000). *Water Quality: An Introduction.* Klumer Academic Publishers.

Central Pollution Control Board (2007). Annual report. Technical report, http://cpcb.nic.in/upload/AnnualReports/AnnualReport_34_Annual-Report-06-07.pdf.

Central Pollution Control Board (2009a). Annual report. Technical report, http://cpcb.nic.in/upload/AnnualReports/AnnualReport_37_ANNUAL_REPORT-08-09.pdf.

Central Pollution Control Board (2009b). Comprehensive environmental assessment of industrial clusters. Technical report, http://www.cpcb.nic.in/upload/NewItems/NewItem_152_Final-Book_2.pdf.

Chakravorty, S. (2003). Industrial location in post-reform India: patterns of inter-regional divergence and intra-regional convergence. *The Journal of Development Studies*, 40(2):120–152.

Chay, K. and Greenstone, M. (2003). The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. *Quarterly Journal of Economics*, 118(3).

Chen, Y., Ebenstein, A., Greenstone, M., and Li, H. (2013). The long-run impact of air pollution on life expectancy: Evidence from China's Huai River policy. *Mimeo.*

Currie, J. and Walker, R. (2011). Traffic congestion and infant health: Evidence from E-ZPass. *American Economic Journal: Applied Economics*, 3(1).

Di Tella, R. and Schargrodsky, E. (2003). The role of wages and auditing during a crackdown on corruption in the city of Buenos Aires. *Journal of Law and Economics*, 46(1):269–292.

DiNardo, J., Fortin, N., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5):1001–1044.

Dranove, D. and Jin, G. Z. (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, 48(4).

Duflo, E., Greenstone, M., Pande, R., and Ryan, N. (2013). What does reputation buy? differentiation in a market for third-party auditors. *American Economic Review, Papers & Proceedings*, 103(3).

Graff-Zivin, J. and Neidell, M. (2012). The impact of pollution on worker productivity. *American Economic Review*, 102(7).

Greenstone, M. and Hanna, R. (2013). Environmental regulations, air and water pollution and infant mortality in India. *Mimeo, MIT*.

Griffin, J. and Tang, D. (2011). Did credit rating agencies make unbiased assumptions on CDOs? *American Economic Review, Papers & Proceedings*, 101(3):125–30.

Gujarat High Court (1996). Environmental Audit Scheme for Industries Manufacturing Specified Products. *Special Civil Application*, 770/1995.

Gujarat High Court (2010). Gujarat Dyestuff Manufacturers Association vs. State of Gujarat. *Special Civil Application*, 22609/2005.

Hanna, R. and Oliva, P. (2011). The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City. *National Bureau of Economic Research Working Paper*, (17302).

Hatanaka, M., Bain, C., and Busch, L. (2005). Third-party certification in the global agrifood system. *Food Policy*, 30(3):354–369.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–162.

IHD-WHO Working Group (1978). *Water Quality Surveys*. World Health Organization.

Kunreuther, H., McNulty, P., and Kang, Y. (2002). Third-party inspection as an alternative to command and control regulation. *Risk Analysis*, 22(2):309–318.

Niehaus, P. and Sukhtankar, S. (2012). The marginal rate of corruption in public programs: Evidence from India. *Mimeo, Dartmouth.*

Olken, B. and Pande, R. (2012). Corruption in developing countries. *Annual Review of Economics*, 4.

Olken, B. A. (2007). Monitoring corruption: Evidence from a field experiment in Indonesia. *Journal of Political Economy*, 115(2).

Paliwal, R. (2006). EIA practice in India and its evaluation using SWOT analysis. *Environmental impact assessment review*, 26(5):492–510.

Potoski, M. and Prakash, A. (2005). Green clubs and voluntary governance: ISO 14001 and firms' regulatory compliance. *American Journal of Political Science*, 49(2):235–248.

Rahman, D. (2012). But who will monitor the monitor? *American Economic Review*, 102(6):2267–2297.

Rauch, J. E. and Evans, P. B. (2000). Bureaucratic structure and bureaucratic performance in less developed countries. *Journal of Public Economics*, 75(1):49–71.

Raynolds, L., Murray, D., and Heller, A. (2007). Regulating sustainability in the coffee sector: A comparative analysis of third-party environmental and social certification initiatives. *Agriculture and Human Values*, 24(2):147–163.

Ronen, J. (2010). Corporate audits and how to fix them. *The Journal of Economic Perspectives*, 24(2):189–210.

Strobl, G. and Xia, H. (2011). The issuer-pays rating model and ratings inflation: Evidence from corporate credit ratings. *Mimeo, University of North Carolina.*

The Hindu (2011). Present system of environmental impact assessment has flaws, says Ramesh. March 19, 2011.

U.S. Environmental Protection Agency (2010). National ambient air quality standards.

Waite, T. D. (1984). *Principles of Water Quality*. Academic Press.

White, L. (2010). Markets: The credit rating agencies. *The Journal of Economic Perspectives*, 24(2):211–226.

World Health Organization (2006). WHO air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide. *World Health Organization*.

VI   Figures

Figure 1: Readings for Suspended Particulate Matter (SPM, mg/Nm^3), Midline
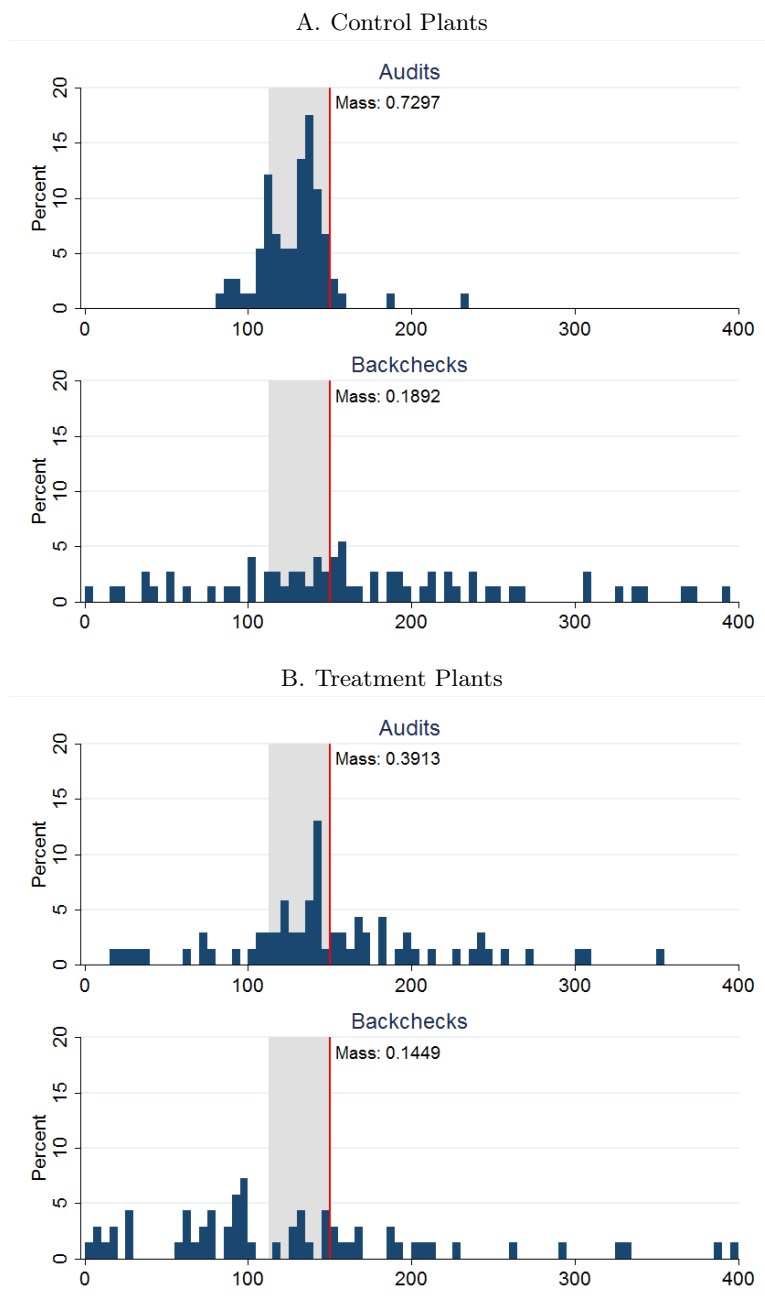
The figure shows distributions of pollutant concentrations for Suspended Particulate Matter (SPM) in boiler-stack samples taken during the midline survey. Panel A shows the distributions of readings at control plants from audits and backchecks, respectively, and Panel B readings at treatment plants from the same two sources. The regulatory maximum concentration limit of 150 mg/Nm^3 for SPM is marked with a vertical line and the area between 75% and 100% of the limit is shaded in gray.

Figure 2: Audit Treatment Effect in Density Bins, All Pollutants
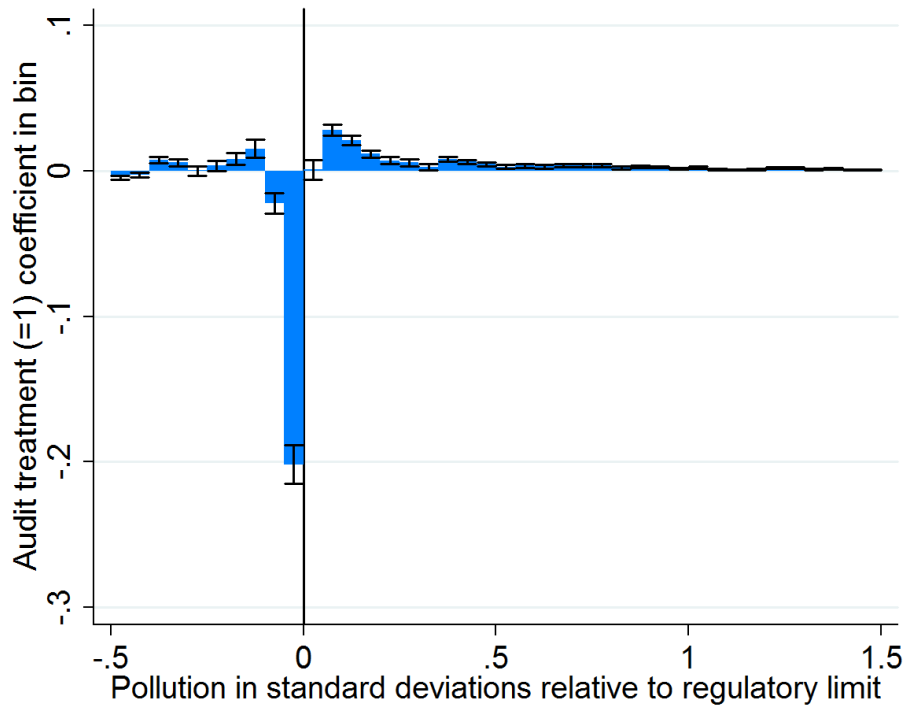
The figure reports point estimates and standard errors from 40 OLS regressions where the dependent variables are indicators for a pollutant reading being within a given density bin and region fixed effects and the independent variable is audit treatment. All pollutants are included with $Water = \{$NH$_3$-N, BOD, COD, TDS, TSS$\}$ and $Air = \{$SO$_2$, NO$_x$, SPM$\}$ and $All = Water \cup Air$. Pollutants are standardized by subtracting the regulatory standard for each pollutant and dividing by the standard deviation in backchecks of that pollutant. Density bins are 0.05 standard deviations wide.

Figure 3: Time Series of Audit Reports by Treatment Status, All Pollutants
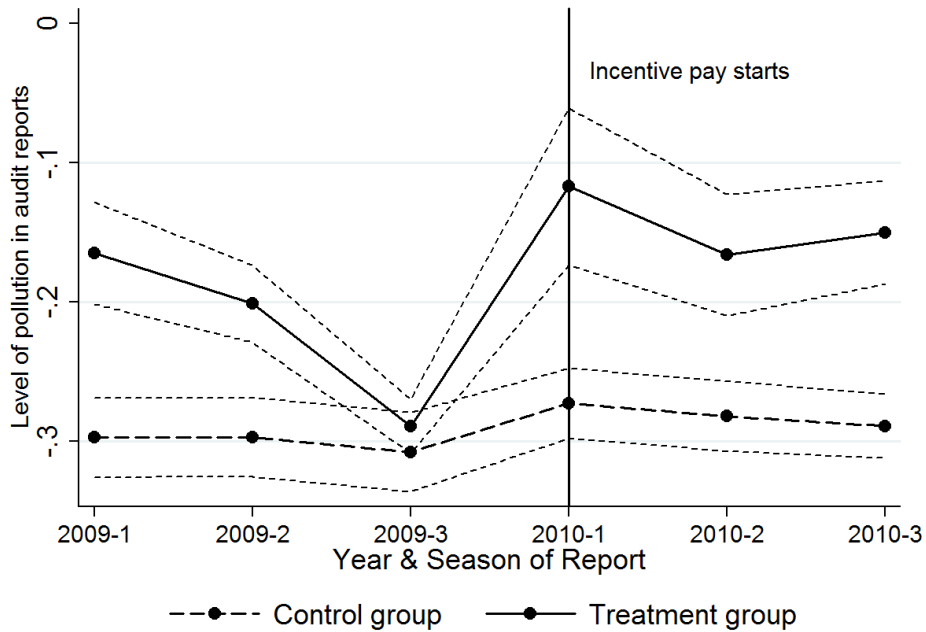
The figure reports the mean standardized pollution level reported in audits by time and treatment status. Time is divided into two years and the three seasons of the year in which auditors are required to monitor pollution. All pollutants are included with $Water = \{$NH$_3$-N, BOD, COD, TDS, TSS$\}$ and $Air = \{$SO$_2$, NO$_x$, SPM$\}$ and $All = Water \cup Air$. Pollutants are standardized by subtracting the mean for each pollutant and dividing by the standard deviation, where both statistics are calculated from backchecks of that pollutant. The dotted lines around the mean reports in each group give 95% confidence intervals for the mean using standard errors clustered at the plant level.

Figure 4: Quantile Treatment Effects of Audit on Endline Pollution
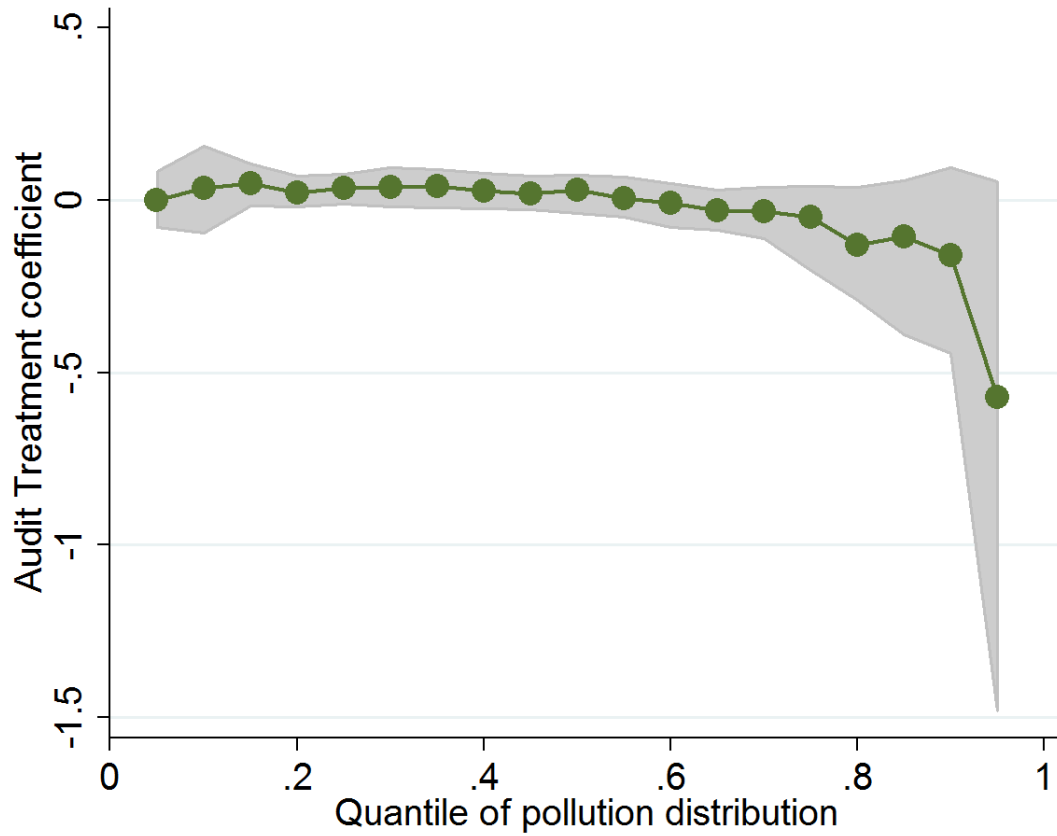
The figure reports estimates from quantile regressions of standardized endline pollution for all pollutants on a dummy for audit treatment assignment and region fixed effects in the audit sample of plants not subject to the cross-cut experimental treatment, analogous to the OLS specifications in Table VII. The quantiles are from 0.05-quantile to the 0.95-quantile at 0.05-quantile intervals. The gray area shows 95% confidence intervals for the treatment coefficient at each quantile from a cluster-bootstrap with replacement at the plant level with 200 replications.

Figure 5: Regulatory Actions by Degree of Violation
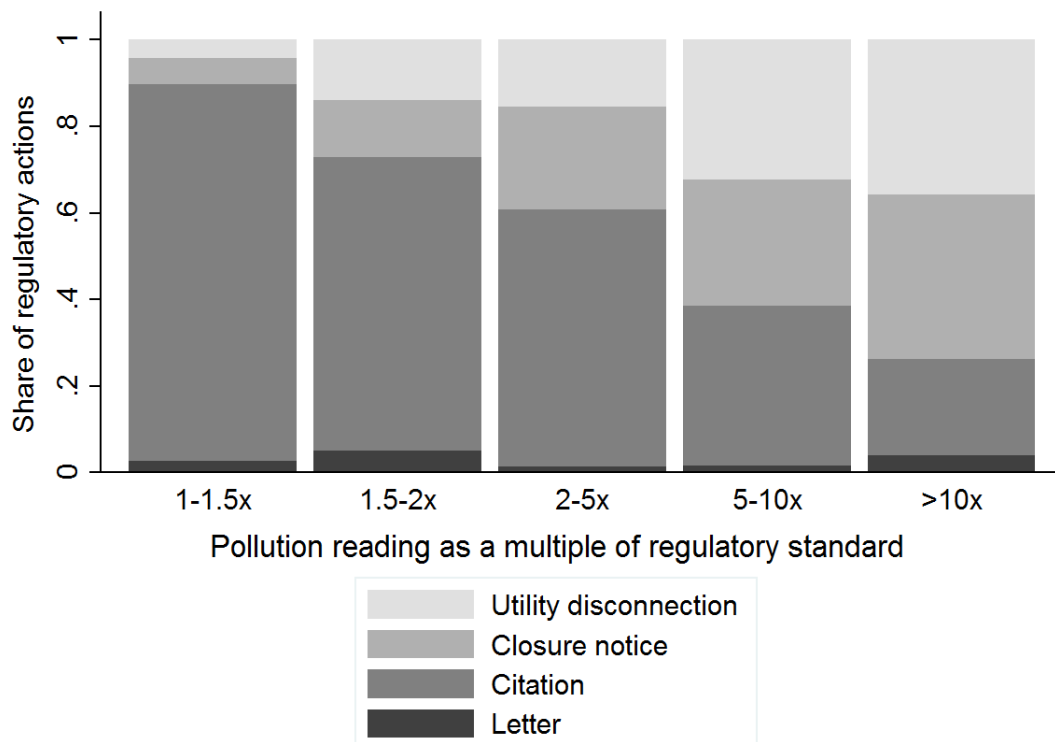
The figure reports the regulatory responses to pollution readings measured at different levels of noncompliance during regulatory inspections for audit sample plants over the three years beginning one year prior to the study. Pollutant readings are shown in bins of readings at specified multiples above the regulatory standard. The bars indicate the type of regulatory action taken in response to a given reading. Actions increase in severity from bottom (dark bars) to top (light bars): a letter is official but not legal correspondence to the firm noting the violation and possibly threatening action, a citation is a legal regulatory notice requiring a response from the firm, a closure warning is a warning that the plant will be closed unless a violation is remedied, and a disconnection is an order to the utility that a plant's power be disconnected. All of these actions were coded based on complete administrative records of plant interactions with the regulator. Going left to right across the bars, the number of violating plants (actions) used to calculate the action shares at each degree of violation are 153 (305), 102 (159), 141 (178), 70 (120) and 72 (126).

VII   Tables

TABLE I

SUBMISSION OF AUDIT REPORTS

	Treatment (1)	Control (2)	Difference (3)
Panel A. 2009			
Audit submitted	163	177	
Total plants	233	240	
Share submitted	0.70	0.74	-0.038
			(0.041)
Panel B. 2010			
Audit submitted	164	153	
Total plants	233	240	
Share submitted	0.70	0.64	0.066
			(0.043)

The table reports on the number of audit reports submitted to the regulator for plants in the audit sample over the two years of the experiment. Column (3) shows differences between treatment and control group submission rates with standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE II
Audit Treatment Covariate Balance

	Treatment (1)	Control (2)	Difference (3)
Panel A. Plant Characteristics			
Capital investment INR 50m to 100m (=1)	0.092	0.14	-0.051
	[0.29]	[0.35]	(0.033)
Located in industrial estate (=1)	0.57	0.53	0.042
	[0.50]	[0.50]	(0.051)
Textiles (=1)	0.88	0.93	-0.030
	[0.33]	[0.26]	(0.025)
Effluent to common treatment (=1)	0.41	0.35	0.078
	[0.49]	[0.48]	(0.049)
Wastewater generated (kl/day)	420.5	394.6	35.4
	[315.9]	[323.4]	(31.6)
Lignite used as fuel (=1)	0.71	0.77	-0.024
	[0.45]	[0.42]	(0.029)
Diesel used as fuel (=1)	0.29	0.25	0.038
	[0.45]	[0.43]	(0.046)
Air emissions from flue gas (=1)	0.85	0.87	-0.0095
	[0.35]	[0.33]	(0.016)
Air emissions from boiler (=1)	0.93	0.92	0.026
	[0.26]	[0.27]	(0.027)
Bag filter installed (=1)	0.24	0.34	-0.10**
	[0.43]	[0.47]	(0.046)
Cyclone installed (=1)	0.087	0.079	0.0010
	[0.28]	[0.27]	(0.027)
Scrubber installed (=1)	0.41	0.41	-0.018
	[0.49]	[0.49]	(0.050)
Panel B. Regulatory Interactions in Year Prior to Study			
Whether audit submitted (=1)	0.82	0.81	0.022
	[0.38]	[0.39]	(0.038)
Any equipment mandated (=1)	0.42	0.49	-0.047
	[0.50]	[0.50]	(0.047)
Any inspection conducted (=1)	0.79	0.78	0.016
	[0.41]	[0.42]	(0.042)
Any citation issued (=1)	0.28	0.24	0.035
	[0.45]	[0.43]	(0.045)
Any water citation issued (=1)	0.12	0.12	-0.0031
	[0.33]	[0.33]	(0.034)
Any air citation issued (=1)	0.027	0.0052	0.021*
	[0.16]	[0.072]	(0.013)
Any utility disconnection (=1)	0.098	0.094	0.0029
	[0.30]	[0.29]	(0.031)
Any bank guarantee posted (=1)	0.033	0.026	0.0045
	[0.18]	[0.16]	(0.017)
Observations	184	191	

The sample includes firms in the audit sample that submitted an audit report in either year; the balance for all audit sample firms, discussed in the text, is similar. Columns (1) and (2) show means with standard deviations in brackets. Column (3) shows the coefficient on treatment from regressions of each characteristic on treatment and region fixed effects. 50 INR ≈ 1 USD. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE III

COMPLIANCE IN AUDITS RELATIVE TO BACKCHECKS, CONTROL GROUP ONLY

	All pollutants (1)	Water pollutants (2)	Air pollutants (3)
Panel A. Dependent variable: Narrow compliance (Dummy for pollutant between 75% and 100% of regulatory standard)			
Audit report (=1)	0.270***	0.297***	0.230***
	(0.025)	(0.034)	(0.033)
Control mean in backchecks	0.097	0.110	0.077
Panel B. Dependent variable: Compliance (Dummy for pollutant at or below regulatory standard)			
Audit report (=1)	0.288***	0.273***	0.311***
	(0.023)	(0.033)	(0.032)
Control mean in backchecks	0.557	0.538	0.586
Observations	1132	688	444

Regressions include region fixed effects. "Audit report" is a dummy for a pollutant reading reported in an audit, as opposed to reported in a backcheck, which is the omitted category. Sample of matched pollutant pairs from audit reports submitted to the regulator in the control group only and corresponding backchecks from the midline survey. Pollution samples from final-stage effluent outlet for water and boiler stack for air. Pollutants included are $Water = \{NH_3\text{-}N, BOD, COD, TDS, TSS\}$ and $Air = \{SO_2, NO_x, SPM\}$ with $All = Water \cup Air$. Standard errors clustered at the plant level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE IV
Compliance in Audits Relative to Backchecks by Treatment Status

	All pollutants (1)	Water pollutants (2)	Air pollutants (3)
Panel A. Dependent variable: Narrow compliance *(Dummy for pollutant between 75% and 100% of regulatory standard)*			
Audit report × Treatment group	-0.185***	-0.212***	-0.143***
	(0.034)	(0.044)	(0.046)
Audit report (=1)	0.270***	0.297***	0.230***
	(0.025)	(0.034)	(0.033)
Treatment group (=1)	-0.0034	-0.013	0.011
	(0.0176)	(0.025)	(0.024)
Control mean in backchecks	0.097	0.110	0.077
Panel B. Dependent variable: Compliance *(Dummy for pollutant at or below regulatory standard)*			
Audit report × Treatment group	-0.234***	-0.166***	-0.345***
	(0.039)	(0.050)	(0.056)
Audit report (=1)	0.288***	0.273***	0.311***
	(0.023)	(0.033)	(0.032)
Treatment group (=1)	0.058*	0.0075	0.145***
	(0.034)	(0.0477)	(0.041)
Control mean in backchecks	0.557	0.538	0.586
Observations	2236	1378	858

Regressions include region fixed effects. "Treatment group" is a dummy equal to one for plants where auditors were randomly assigned, paid a fixed rate from a common pool and subject to bakchecks. "Audit report" is a dummy for a pollutant reading reported in an audit, as opposed to reported in a backcheck, which is the omitted category. Sample of matched pollutant pairs from audit reports submitted to the regulator and corresponding backchecks from the midline survey, in both the treatment and control groups. Pollution samples from final-stage effluent outlet for water and boiler stack for air. Pollutants included are $Water = \{$NH$_3$-N, BOD, COD, TDS, TSS$\}$ and $Air = \{$SO$_2$, NO$_x$, SPM$\}$ with $All = Water \cup Air$. Standard errors clustered at the plant level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE V
Audit Treatment Effects on Auditor Reporting

	All pollutants		Water pollutants		Air pollutants	
	(1)	(2)	(3)	(4)	(5)	(6)

Panel A. Dependent variable: Compliance
(Dummy for pollutant in audit report at or below regulatory standard)

Treatment group (=1)	-0.151***	-0.180***	-0.172***	-0.188***	-0.108***	-0.149***
	(0.015)	(0.022)	(0.021)	(0.030)	(0.016)	(0.021)
Auditor fixed effects	No	Yes	No	Yes	No	Yes
Control mean	0.828	0.828	0.807	0.807	0.863	0.863
Observations	13170	13170	8373	8373	4797	4797

Panel B. Dependent variable: Level of pollutant in audit report

Treatment group (=1)	0.103***	0.131***	0.117**	0.131**	0.0853***	0.132***
	(0.035)	(0.038)	(0.053)	(0.059)	(0.0214)	(0.022)
Auditor fixed effects	No	Yes	No	Yes	No	Yes
Control mean	-0.291	-0.291	-0.350	-0.350	-0.194	-0.194
Observations	13170	13170	8373	8373	4797	4797

Panel C. Dependent variable: Level of pollutant in audit report
minus level of pollution in backcheck

Treatment group (=1)	0.210***	0.153	0.152	0.156	0.312***	0.166*
	(0.073)	(0.099)	(0.102)	(0.138)	(0.083)	(0.095)
Auditor fixed effects	No	Yes	No	Yes	No	Yes
Control mean	-0.304	-0.304	-0.354	-0.354	-0.225	-0.225
Observations	1118	1118	689	689	429	429

Regressions include region fixed effects in all panels and year fixed effects in Panels A and B only. Pollution samples are from the final-stage effluent outlet for water and boiler stack for air. Pollutants included are $Water = \{NH_3\text{-}N, BOD, COD, TDS, TSS\}$ and $Air = \{SO_2, NO_x, SPM\}$ with $All = Water \cup Air$. Panels A and B use the full sample of audit reports that reached the regulator over the two years of the experiment. The outcome variable is a dummy for whether a pollutant was reported compliant in audit reports, in Panel A, and the level of pollution reports, in Panel B. Panel C uses the midline survey sample with pollution in audit reports less pollution in backchecks as the outcome. Standard errors clustered at the plant level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE VI

INCENTIVE PAY FROM TREATMENT EFFECT OVER TIME

	Level of pollutant in audit report, all pollutants	
	(1)	(2)
Treatment group (=1)	0.103***	0.218***
	(0.035)	(0.063)
Incentive pay (year=2010)	0.051**	0.002
	(0.027)	(0.033)
Incentive pay $\times$ treatment		0.257***
		(0.092)
Years (fractional) from Jan 1, 2009		0.029
		(0.023)
Years (fractional) $\times$ treatment		-0.220***
		(0.068)
Region fixed effects	Yes	Yes
Observations	13166	13166

Regressions include region fixed effects. Pollution samples from final-stage effluent outlet for water and boiler stack for air. Sample of all audit reports to the regulator over the two years of the experiment. Pollutants included are $Water = \{$NH$_3$-N, BOD, COD, TDS, TSS$\}$ and $Air = \{$SO$_2$, NO$_x$,SPM$\}$ with $All = Water \cup Air$. Standard errors clustered at the plant level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE VII
ENDLINE POLLUTANT CONCENTRATIONS ON TREATMENT STATUS

	All pollutants (1)	Water pollutants (2)	Air pollutants (3)
Panel A. Dependent variable: Level of pollutant in endline survey			
Audit treatment assigned (=1)	-0.211**	-0.300*	-0.053
	(0.099)	(0.159)	(0.057)
Control mean	0.076	0.114	0.022
Observations	1439	860	579
Panel B. Dependent variable: Compliance			
(Dummy for pollutant in endline survey at or below regulatory standard)			
Audit treatment assigned (=1)	0.027	0.039	0.002
	(0.027)	(0.039)	(0.028)
Control mean	0.573	0.516	0.656
Observations	1439	860	579

Regressions include region fixed effects. Pollution samples from final-stage effluent outlet for water and boiler stack for air. Pollutants included are $Water = \{NH_3\text{-}N, BOD, COD, TDS, TSS\}$ and $Air = \{SO_2, NO_x, SPM\}$ with $All = Water \cup Air$. Endline survey data in the audit sample of plants not subject to the cross-cut experimental treatment. Standard errors clustered at the plant level in parentheses. $^*\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

A   Data Appendix

Pollutant	Description
Panel A: Water Pollutants	
Biochemical Oxygen Demand (BOD)	A measure of the amount of dissolved oxygen consumed by microscopic organisms in a confined sample of water. The BOD and volume of an effluent determine the oxygen demand that will be imposed on receiving waters (Boyd, 2000). The demand for oxygen from effluent may deplete available molecular oxygen, precluding other biological processes, such as marine plants or life, that require oxygen (Waite, 1984).
Chemical Oxygen Demand (COD)	A measure of the oxygen demand of the organic matter in a sample as determined by oxidation of the organic matter with potassium dichromate and sulfuric acid. Often used as a proxy for BOD in determining the oxygen demand of effluent.
Total Dissolved Solids (TDS)	Primarily inorganic substances dissolved in water, including calcium, magnesium, sodium, potassium, iron, zinc, copper, manganese, etc. Water with high dissolved solids is said to be mineralized and decreases the survival of plant and animal life, degrades the taste of water, corrodes plumbing and limits use of water for irrigation (Boyd, 2000; IHD-WHO Working Group, 1978). Depending on the composition of solids TDS may have adverse health effects on people with cardiac disease or high blood pressure.
Total Suspended Solids (TSS)	Organic and inorganic or mineral particles too large to be dissolved but small enough to remain suspended against gravity in an effluent (Boyd, 2000). Contribute to turbidity and color of water and proxy for adverse effects from individual solid components.
Ammoniacal Nitrogen (NH_3-N)	The nitrogen contained in unionized ammonia and ammonium. Though nitrogen is a vital nutrient, some forms of ammonia nitrogen are toxic to aquatic life (Boyd, 2000). The toxicity of ammonia nitrogen increases with decreasing dissolved oxygen concentrations.
Panel B: Air Pollutants	
Sulfur Dioxide (SO_2)	A reactive oxide of sulfur. Short-term exposure has been linked to adverse respiratory effects particularly damaging for asthmatics. SO_2 also contributes to formation of fine particles (World Health Organization, 2006).
Nitrogen Oxides (NO_x)	A group of reactive gases including nitrous acid, nitric acid and NO_2. Nitrogen oxides are toxic at high concentrations and contribute to formation of ozone and fine particles, which are detrimental to health (World Health Organization, 2006).
Suspended Particulate Matter (SPM)	A mixture of small particles and liquid droplets with a number of components, including acids, organic chemicals, metals and soil or dust (U.S. Environmental Protection Agency, 2010). Particulate matter affects respiratory and cardiovascular health and has been shown to increase infant mortality and shorten lifespans (World Health Organization, 2006; Currie and Walker, 2011; Chen et al., 2013).

The table describes the pollutants used throughout the analysis of both auditor reporting and plant pollution emissions. Auditor incentive pay during year two of the experiment was based on these pollutants and the water parameter *pH*. We omit *pH* from the analysis because environmental damages from *pH* are not monotonic in the reading—both high (alkanine) and low (acidic) readings can be bad—which makes results for *pH* incomparable to results for other pollutants. For this same reason, the standard for *pH* is a range, rather than a maximum.