Truth-Telling by Third-Party Auditors: Evidence from a Randomized Field Experiment in India*

Esther Duflo, Michael Greenstone, Rohini Pande and Nicholas Ryan

March 2, 2012

Abstract

In a wide range of markets, from environmental and corporate audits to credit ratings, private third-party entities are directly hired and paid by the firms on which they report. In this paper, we demonstrate that the conflict of interest faced by third-party auditors can corrupt the information they provide and undermine regulation. Our evidence comes from a two-year randomized field experiment on environmental audits of industrial plants conducted in collaboration with a state regulator in India. For a randomly selected set of plants, the relationship between auditors and client plants was severed, in that auditors were randomly assigned to plants and backchecked on their accuracy. Random backchecks by independent surveyors allow us to observe true pollution outcomes. We report three main findings. First, auditors who face the standard incentives, in the control group of plants, systematically underreport pollution readings relative to the truth, as revealed by independent backchecks of the same readings. Auditors working in the control falsify reports in a targeted fashion, reporting many plants as just meeting regulatory limits. Second, the reports of auditors for the treatment group of plants are statistically equal to the truth. Notably, many auditors worked simultaneously in the treatment and control groups of plants, and the improvement in reporting is evident when we compare the reporting of the same auditors across the two sets of economic incentives. Third, plants in the treatment group, assigned to independent auditors, reduced pollution emissions relative to the control group of plants. This response is seen only for water pollutants, which are subject to greater regulatory scrutiny.

*We thank Sanjiv Tyagi, R. G. Shah, and Hardik Shah for advice and support over the course of this project. We thank Pankaj Verma, Eric Dodge, Vipin Awatramani, Logan Clark, Yuanjian Li and Nick Hagerty for excellent research assistance. We thank the Sustainability Science Program (SSP), the Harvard Environmental Economics Program, the Center for Energy and Environmental Policy Research (CEEPR), the International Initiative for Impact Evaluation (3ie), the International Growth Centre (IGC) and the National Science Foundation (NSF Award #1066006) for financial support. Ryan thanks an AMID Early Stage Researcher Fellowship for financial support. All views and errors are solely ours.

I Introduction

In a wide range of settings, third-party auditing is used to enforce regulation or otherwise verify quality. For example, in the United States, the Securities and Exchange Commission (SEC) requires public companies to file independently audited annual financial statements. The U.S. credit ratings agencies also serve an important regulatory role (White, 2010).[1] Independent audits play a large role a range of consumer and commodity markets, including those for food safety, flowers, timber, healthcare and many durable goods (Hatanaka et al., 2005; Raynolds et al., 2007; Dranove and Jin, 2010). Third parties are important in the enforcement of various environmental standards including ISO 14001 certification and the verification of carbon abatement in the carbon offset market (Potoski and Prakash, 2005; Bhattacharyya, 2011).

A central concern with third-party auditing is that third-party auditors may face a conflict of interest between providing credible reports and maintaining business with their client firms. A growing body of evidence suggests that this problem may be acute. The financial crisis of 2007-2008 brought to light the poor quality of credit ratings during the subprime boom; 80% of collateralized debt obligations originally rated AAA by Standard & Poor's in 2005-2007 were downgraded to below investment grade by mid-2009 (International Monetary Fund, 2009). Using data for a single credit agency, Griffin and Tang (2011) find that the agency's internal surveillance team's judgements on CDO ratings were more accurate than the business-oriented ratings team's and that the difference between these measures is predictive of future downgrades. Strobl and Xia (2011) compare the 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. They find the difference in ratings across the two agencies is more pronounced when the the issuer-pay rating agency plausibly has more business at stake, such as when an issuer has more short-term debt.[2] While there have been many calls for reforming the third-party audit system to increase the independence of auditors, reforms to date have not addressed the core problem that firms, when they directly hire their auditor, can shop for a favorable report.[3]

[1]The SEC now designates credit rating agencies as nationally recognized statistical ratings organizations to qualify them for ensuring compliance with certain regulations, such as bank capital requirements.

[2]For broad overviews of the corporate audit and credit ratings markets see Ronen (2010) and White (2010), respectively.

[3]In 2002, the Sarbanes-Oxley Act overhauled the regulatory structure for corporate audits, and made audi-

In this paper, we report on a randomized field experiment that provides causal evidence on how conflict of interest undermines information provision by third-party auditors. The results show how a simple package of reforms to promote auditor independence can restore accuracy and improve firm compliance with regulation.

We designed and implemented this experiment in collaboration with a state pollution regulator in India, the Gujarat Pollution Control Board (GPCB). Since 1996, Gujarat has had a third-party audit system for plants with a high pollution potential. Auditors visit each plant thrice over the year and summarize their findings in an audit report submitted to both the plant and GPCB, which can form the basis of regulatory action (Gujarat High Court, 1996). The report covers pollution readings from each auditor visit and describes the production process and physical state of the plant, including measures taken for pollution control. The audit system includes several safeguards for ensuring audit quality: auditors must be accredited by an audit committee, auditors are forbidden from taking on additional consultancy work for plants they audit, a given audit team cannot undertake more than 15 audits in a year and auditors that are found to be inaccurate can in principle be decertified. These safeguards aside, the basic financial arrangement underlying these audits is the typical practice the world over — plants hire and pay auditors directly.

Our experiment was designed to make auditors financially independent of the firm audited and to measure the effect of this independence on auditor reporting. We varied auditor hiring and payment practices for a randomly chosen 233 of 473 audit-eligible plants in two populous and heavily polluted industrial regions of Gujarat. In the audit treatment, auditors were randomly assigned to plants, paid centrally and backchecked on a random sample of audit visits. Auditors were notified in advance that some of their work might be backchecked but the actual backcheck visits, conducted by an independent technical agency, were unannounced. During the backcheck, the survey company collected pollution readings for the same pollutants at the same sampling locations as were reported in audits. In the first year, auditor accuracy was measured but no consequence of poor performance was explicitly specified. In the second

tors of public companies subject to oversight by a private-sector, nonprofit corporation, the Public Company Accounting Oversight Board. The board determines who can perform audits, conducts investigations and sets fines. Three former SEC Chairmen testified in favor mandatory auditor rotation, which was not adopted. 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 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 firms that they audit.

year, auditors' pay was directly linked to accuracy, as measured by the backchecks, and this link was announced beforehand. Backchecks were of no regulatory consequence for audited plants.

We have three main findings. First, auditors that face standard economic incentives systematically underreport pollution readings, relative to the truth as measured by backchecks. The average difference between backcheck and audit pollution readings across all pollutants for which we gathered data is 0.30 standard deviations in the control group, indicating a significant negative bias in reports. In particular, in the control plants, auditors systematically and incorrectly report many pollution readings to be just below the regulatory standard. Across all pollutants, the control group shows an excess mass of 30% of pollutant reports in the area beneath the pollution standard, relative to the same area in the distribution of backcheck readings.

Our second finding is that our reform eliminates this bias. Auditors working in the treatment group improve the accuracy of their reports to the extent that by the end of the experiment they are statistically equal to the truth. This large improvement in auditor reporting is evident in specifications with auditor fixed effects that compare the behavior of the same auditors working in the treatment and control. Treatment auditors show an excess mass of only 7% in the area under the standard, a decrease of 76% relative to the control. Additional, non-experimental analysis suggests that the different components of the treatment package, backchecks and financial incentives for accuracy, independently contributed to this success.

Our final finding is that plants in the treatment group respond to independent audits by significantly reducing their pollution emissions, as measured by an independent endline survey we conducted. This reduction occurs mainly in water pollutants, which are a regulatory priority.

These results are consistent with a model of auditor and firm behavior where firms can induce auditors to provide false reports by offering to pay higher prices. Auditors balance these payments and the value of a reputation for leniency against the risk of being disbarred if they are caught misreporting. The experiment reduced the ability of firms to shop for an audit they like and the value of leniency, while increasing the probability of false reports being detected. All of these changes increased auditors' incentives to report truthfully. Firms, given the greater cost of filing a falsely low report, are induced to abate pollution rather than risk

a fine for filing a high report.

These results demonstrate the power of the incentives faced by third-party auditors. They 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 remarkably clear evidence that altering economic incentives can cause auditors to switch from biased reporting to truth-telling in a very short span of time, and that this change leads to improvement in real outcomes, here pollution. To our knowledge, there is no comparable evidence in the prior literature for any sector.

Our findings are consistent with some insights offered by models of information intermediaries. First, incentives for reputation building are typically weaker when information is more complex (Mathis et al., 2009). Pollution readings may be relatively complicated to verify as they require careful, integrated field sampling and laboratory analysis. Second, competition can lead to less accurate information. Dranove and Jin (2010) review the theoretical and empirical evidence on third-party certification from a broad range of markets and find that competition may motivate certifiers to relax their rating standards and sellers to shop for favorable ratings. In the context of credit rating agencies Bolton et al. (2011) show that competition among CRAs can lead to rating shopping and rating inflation such that efficiency is lower under competition than monopoly. Rating inflation is particularly likely when some part of payments are made after the rating is completed, as in our setting. One element of our experiment was to suppress the ability to shop for a report among auditors by directly assigning auditors to industrial plants, reducing the potential for strategic contracting between auditors and firms.

The remainder of the paper is organized as follows. In Section II, we describe the background and the experimental design. In Section III we model the behavior of polluting firms and environmental auditors and lay out expected responses to the experiment in terms of model parameters. In Section IV we discuss our data collection and in Section V present the results. We conclude and discuss some implications for policy in this context and others in Section VI.

II Background and Experimental Design

A Environmental Regulation in Gujarat

High levels of industrial pollution are common in India despite a stringent regulatory framework for pollution control, as set by national and state laws and court orders (Hanna and Greenstone, 2011). At the Federal level, 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. Most regulatory decisions and all enforcement power were devolved to the states. State Pollution Control Boards are responsible for enforcing the provisions of the Water Act and the subsequent Air (1981) and Environmental Protection (1986) Acts and their attendant regulations. In Gujarat, the Gujarat Pollution Control Board (GPCB) enforces these traditional command-and-control pollution regulations. GPCB is responsible for monitoring and regulation of approximately 20,000 plants.

Gujarat is one of India's fastest growing industrial states. The state attracted the largest share of investment after licensive reform of any state in India (Chakravorty, 2003). Since 1991-1992, the peak of industrial licensing reform, net state domestic product has grown an average of 8% per year. Gujarat has about 5 percent of the Indian population, but accounts for 9 percent of India's registered manufacturing employment and 19 percent of output (Authors' calculation, Annual Survey of Industries, 2004-05). This concentration of industry has been accompanied by a degradation of air and water quality: The Central Pollution Control Board categorized eight industrial clusters in Gujarat as severely or critically polluted, including the two most polluted in the country (Central Pollution Control Board, 2009b); Gujarat contains three of India's five most polluted rivers (Central Pollution Control Board, 2007). Essentially all of the large cities in the state, as well as some industrial areas, are classified as non-attainment with respect to the National Ambient Air Quality standards for Respirable Suspended Particulate Matter (RSPM) (Central Pollution Control Board, 2009a).

B Environmental Audit Regulation

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

In 1996, in order to remedy the perceived failure of regulatory 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, heavily polluting plants must receive a yearly third-party audit, conducted by a private auditing firm hired and paid for by the plant. Auditors visit each factory three times each year for about one day, observe the plant's environmental management practices and measure their pollution outputs. Auditors compile the results of these visits and submit their audit report to GPCB by February 15th of the following year. The final audit report, which is sent to the plant and GPCB, 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 to the plant based on this evidence.

Plants are required to submit an environmental audit if they have high pollution potential. Plants of the most polluting types are categorized as Schedule I, and must be audited by a university or similar institution, called a Schedule I auditor. Somewhat less polluting plants are categorized as Schedule II, and must be audited by a private audit firm, called a Schedule II auditor. Plants with lower pollution potential are exempted. The classification of pollution potential is done on three dimensions: what product the plant produces, where it sends its waste effluent and the volume of that effluent. 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). This study concerns only the reporting of Schedule II auditors and henceforth we refer to plants in Schedule II as "audit eligible."

C Environmental Audit Market

The audit system as originally implemented has several weaknesses and has been criticized by both industry and environmentalists. Plants directly select and pay the auditor and review of the auditors by the GPCB is mostly *pro forma*. GPCB officials suspect that the reports submitted by auditors are not reliable. Meanwhile, industry recently litigated against the scheme, somewhat ironically and without success, to get the 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).

These criticisms have arisen despite the fact that the basic structure of the environmental audit system includes various safeguards and strong penalties. Auditors are certified by GPCB and can audit a plant at most three years in a row. Each audit team is limited to auditing at most 15 plants per year. In principle, non-submission of an audit is punishable by closure and disconnection of water and electricity connections, and a report showing non-compliance with the terms of a plant's environmental consent by closure or payment of compensation (Gujarat High Court, 1996). Moreover, auditors with reports found to be inaccurate are liable to be de-certified and their reports on behalf of other plants declared void, an incentive to build a reputation for quality.

Despite these safeguards, the environmental audit market is characterized by strong price competition and low audit quality that varies widely across audit firms. Informally, before the experiment started, both auditors and firms claimed that an audit report could be purchased for as low as INR 10,000-15,000 rupees. We collected information on audit prices after the experiment from both auditors and plants. Conditional on reporting any payment, plants in the treatment and control report very similar average payments of around INR 24,000.[4] Auditors report a mean minimum plant willingness-to-pay of INR 25,000 and a mean maximum of INR 52,000.[5] Using GPCB pollution sampling and analysis charges, we estimated that the cost of performing a thorough audit should be between INR 20,000-40,000 depending on the plant's characteristics. As about 80% of our audited sample are textile plants, for which the cost basis of an audit is around INR 40,000, we take the sum of evidence on audit prices as showing that a significant portion of audit reports are being purchased at prices below the cost of an audit duly performed. Auditors rank price as by far the most important factor in plant selection of auditors, at a mean of 4.7 on a scale of 1 to 5. Audit quality is last amongst four factors at 2.5 on the same five-point scale. The audit market supports a range of quality but in general is very competitive, with price competition fostering very low quality at the bottom of the market. Indeed, many of the best-reputed auditors reported they had scaled back or closed their audit businesses to focus on environmental consulting.

[4]Fifty plants report making payment in the treatment group despite the fact that audits were paid for by the experiment, and our understanding is that they took the question to mean what their "typical payment" was, not the 2010 payment.

[5]The fact that the mean reported by auditors is higher than the mean reported by firms is consistent with the fact that low-cost auditors audit more firms. The minimum reporting here may be more reliable as auditors may have expected answers given at this conference to be accounted for in future audit reforms.

D Experimental Sample

Study plants were drawn from 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 100m (about USD 2.2m), which were designated small or medium scale, and selected the 633 audit-eligible firms as a provisional sample. We randomly assigned half of plants within this provisional sample, stratified on region, to the audit treatment group. After the randomization, we collected the required detailed sectoral information and eliminated the firms found to be ineligible from both treatment and control, reaching the final study sample of 473 plants, 49.2% of them belonging to the treatment group.

Table 1 presents summary statistics for this study sample using administrative baseline date from GPCB records. Plants assigned to the audit treatment and control look very similar. About 70% of the sample, which was selected for being audit-eligible, submitted an audit report in the prior year. Most sample firms are textile factories eligible for environmental audit due to having high volumes of waste effluent. Both the treatment and the control firms have similar pollution potential as measured by effluent quantity and type of fuel used. Treatment firms are 9 percentage points less likely to have a bag filter, a type of air pollution control equipment, installed, but are similar to control firms with respect to other control equipment such as cyclones and scrubbers.

E Experimental Design

We designed and evaluated a modification of the existing audit system in collaboration with the GPCB in order to increase the independence of auditors and the accuracy of auditor reporting.[6] This modification, the audit treatment, consisted of three changes to the existing audit system: random assignment of auditors to plants, payments to auditors from a central pool, and backchecks of auditor performance. Treatment plants were assigned to the audit treatment once, in 2009, for the audit years 2009 (hereafter year 1) and 2010 (year 2). In year 2, direct incentive pay for auditor accuracy was added to the basic set of three reforms. We now discuss the components of the audit treatment in turn.

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

In the treatment group, auditors were randomly assigned to plants by the research team rather than being selected by the audited plants. All Schedule II audit firms certified by GPCB were solicited for their interest to participate in the treatment separately in each year. In both years interest was oversubscribed, relative to the number of treatment plants, and so at the beginning of each year auditors were randomly assigned to a number of plants allocated *pro rata* on the basis of their capacity, measured in number of certified audit teams.

Auditors working in the treatment plants were paid from a central pool of funds raised for the study, rather than being hired and paid directly by the plants audited. The strong interest in the program may reflect both that treatment audit payments, discussed below, were in the high range of the market and that the audit treatment offered auditors better working terms, for example in assigning client plants.

A random set of auditor visits to plants in the treatment were backchecked in the field by technical staff of independent engineering colleges (which were certified as Schedule I auditors) soon after 20% of audit visits. Auditors were aware of the possibility of being backchecked and aware that backcheck results would be used for quality control, although in the first year the terms of any sanction for poor performance were left somewhat vague and auditors were paid INR 45,000 regardless of accuracy.[7] The results of backchecks held no regulatory consequence for audited plants and were only sent to the regulator as part of aggregated reports on the accuracy of auditors.

In year 2 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 year 1, after which auditors were shown their measured accuracy. The pay formula first calculated the difference between audit report pollution concentration readings and backcheck readings,standardized by the variability of backcheck readings, for each pollutant. 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. An index of zero thus means that an auditor always reported readings that matched backchecks exactly and of one that the average auditor and backcheck readings were one standard deviation apart. Auditors were grouped into three bins of the least accurate quartile, paid Rs. 35,000 per audit, the next least accurate quartile, paid Rs. 40,000 per audit, and the most accurate half, paid

[7]This rate was estimated by applying GPCB's own sampling charges to the average plant characteristics in the audit sample and adding a small margin.

Rs. 52,500 per audit, maintaining the average pay of the prior year.

Note the audit treatment was assigned at the plant level, and that to the extent they had capacity, auditors were allowed to work both with treatment and control plants. This aspect of the design allows comparisons of the effect of the audit treatment on auditor reporting within the same set of audit firms. Many firms did work in both treatment conditions. In the first year, out of 42 auditors, 24 worked in control only, 9 in treatment only, and 9 in both. In 2010, out of 34 auditors, 7 worked in control only, 12 in treatment only, and 15 in both.[8]

III Model

This section presents a model that clarifies the incentives faced by auditors and firms and helps to interpret the potential effects of the reform. We model the behavior of environmental auditors and polluting firms as a three-stage game with the environmental regulator. Auditors report to the environmental regulator and may risk their future business to falsely report low pollution given the payments offered by firms. Firms decide whether inducing false reporting is worthwhile given their own costs of abatement and the payment needed to compensate the auditor for reporting falsely. We discuss the setting, the solution and the relation of the model to our experimental treatments. The Model Appendix in Section XI derives the equilibrium studied.

A Set-up

We consider a three-stage game between a firm, an auditor and the regulator.

In the first stage, firms observe their abatement cost and set a payment schedule for their environmental auditor. Firms produce output with pollution by maximizing the objective

$$S - c - w - f,$$

where S is the value of production, c is the cost of abating pollution, w is the amount paid to environmental auditors and f is any fine imposed by the regulator for violations of pollution standards. All firms pollute at a level p_h, above the regulatory limit, with no abatement. By

[8] 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.

paying an abatement cost c firms can reduce their emissions to p_l, below the regulatory limit. Firms draw c from a cumulative distribution function G.

Firms make a joint decision about whether to abate pollution and what schedule of payments to offer auditors in order to minimize their overall cost of production. We assume that S is sufficiently large that environmental regulation does not cause firms to shut down. The payment $w(\hat{p}|p)$ offered auditors can be a function of both the firm's true pollution level p and the auditor's report $\hat{p}$ on the pollution level. Let the notation $w_{lh} = w(\hat{p} = p_l|p = p_h)$ and similarly for other conditional payments. The goal of the firm when offering payments is to minimize the sum of the cost of an audit and any potential fine.[9]

Auditors observe the pollution level of firms and report on firm pollution to the regulator. The payoff to an auditor is given by

$$w - a + \mathbb{E}V(\hat{p}|p)$$

where w is the payment for performing an audit this period, a is the cost of preparing an audit and the final term is the expected value of future audit business. The continuation value of future audit business depends on audit reports $\hat{p}$ through an exogenous function V, which captures the effect of auditor reputation *with firms*. We assume a reputation for leniency is valuable to auditors, so that $V_{lh} = V(p_l|p_h) > V_{hh} = V_{hl} = V_{ll} = V_0$ where the base continuation value V_0 is the same for all other conditional reports. Define $\Delta V = V_{lh} - V_0$ as the incremental reputational value of falsely-low reporting, i.e. the benefit of notoriety.[10] The expected value depends on the likelihood that an auditor will be disbarred if found to be inaccurate.

We model the audit market as competitive but assume that auditors nonetheless receive some surplus from participating in the market. The marginal cost of an audit may be interpreted as including some rent to being certified as an auditor or the opportunity cost to auditors of doing audits, which can employ their staff in environmental consultancy roles if

[9] We assume not filing an audit is very costly and so is not considered as an action. In practice not filing is similar to reporting $\hat{p} = p_h$.

[10] Bolton et al. (2011) and many other models of reputation also take continuation values as exogenous and, moreover, as wholly independent of current actions. In our set up the continuation values may have been affected by the treatment, as a reputation for leniency is less valuable in a system with random assignment of auditors. Since the experiment was conducted for only two years, changes in continuation values were likely to be small and we do not emphasize them in the model.

auditing is not profitable.

Finally, the regulator reviews audit reports $\hat{p}$ and may impose fines on plants and disbar inaccurate auditors. With probability q the regulator observes the true state of pollution at a plant, either through the reports of its own staff or backchecks of auditor performance.[11] The regulator decides whether to impose a fixed fine f based on the audit report and whether to disbar auditors that report falsely. The continuation value for a disbarred auditor is zero. As the focus of the paper is on auditor behavior we take a simple view of the regulator as seeking only to minimize pollution.

B Equilibrium, and response to Experimental Treatments in the Model

We restrict attention to equilibria with regulation, defined as equilibria in which the regulator imposes fines when p_h is observed and always disbars auditors reporting falsely. There are equilibria without regulation but we believe that equilibria with regulation are more likely to occur given that the regulator is relatively long-lived.[12]

Proposition 1. *There is a unique subgame-perfect equilibrium with regulation in which one of three outcomes obtains. Firms* clean up *if* $c \leq min\{f, qV_{lh} - \Delta V\}$, underreport *if* $qV_{lh} - \Delta V < min\{c, f\}$ *and* defy *the regulation if* $f < min\{c, qV_{lh} - \Delta V\}$.

The firm minimizes costs across one of three strategies. Most simply, if abatement cost is low the firm can abate to p_l and pay auditors at cost for giving a truthful report. If abatement cost is high, the firm can choose between buying a truthful, high report and incurring a fine and buying a falsely low report. In any equilibrium with regulation, inducing low reporting requires $w_{lh} \geq w_{hh} - \Delta V + qV_{lh}$, as the firm has to compensate the auditor for the risk of disbarment qV_{lh} that comes with a false report. The firm gets a discount on false reporting due to the reputational benefit ΔV of notoriety to auditors.

Baseline case. In the *status quo* audit system, at the beginning of the experiment and in the control group of the audit treatment, the probability q of audit reports being verifiable

[11] We assume the regulator does not use this direct observation to fine plants, as backchecks are not legally valid grounds to impose fines. Alternatively, plant abatement costs may incorporate the baseline likelihood of being penalized by the regulator directly.

[12] If the game were dynamic the regulator would be considered a long-lived player against short-lived auditors and firms. In the manner of Kreps and Wilson (1982), it would then select its most preferred equilibrium, in which some firms abate, despite moving last.

was very low. As $q \to 0$ the cost of buying falsely low reports $qV_{lh} - \Delta V$ will be negative and thus less than strictly positive abatement costs and fines. Auditor underreporting should therefore be pervasive in these conditions and, as $G(qV_{lh} - \Delta V) = 0$ for small q, no firms will abate in response to audits.

Backchecks. We model the backcheck intervention in the experiment as raising q, the probability an audit report is verifiable.

Proposition 2. *A raise in q from the baseline sufficient for $qV_{lh} - \Delta V > f$ induces a share $G(f)$ of firms to abate.*

The stated rise in q is enough that polluting firms would prefer incurring a fine to paying the auditor a premium for false reporting. Given such a rise, the relevant margin of decision for a firm becomes whether to abate or to incur a fine. The $G(f)$ share of firms with costs beneath the fine abate. The necessary rise in q will be smaller the larger is the continuation value of environmental auditing business for auditors, and larger the larger is the benefit of notoriety.

Incentive Pay. The incentive pay treatment augments the auditor continuation value for accurate reports so that inducing low reporting is more costly than a fine if $q(V_{lh}+B)-\Delta V > f$, where B is the incentive pay bonus.

Assignments. We model the assignment of auditors to firms as firms being required to offer payments unconditional on the contents of the audit report, so $w(\hat{p}|p) = w(p)$. This change captures that firms cannot refuse to pay an auditor to which they were assigned if the assigned auditor submits a high report.[13] Assignments may also reduce the value of notoriety, ΔV, to the extent that auditors expected that the modified audit system might persist beyond the experiment.

Proposition 3. *Under assignments, a raise in q from the baseline sufficient for $qV_{lh} \geq \Delta V$ induces a share $G(f)$ of firms to abate.*

With a flat fee auditors now consider only their reputations when deciding whether to report falsely. This makes the required raise in q smaller than in the case with conditional

[13] In the treatment, auditors were paid with funds raised for the study, not by firms. As a flat fee above cost will not enter into the choice between reports, the salient aspect of this change is the fact that auditors were paid for reports regardless of whether they reported violations. Firms could however still have offered to pay more below the table for a favorable report

payments. If assignments reduce V_{lh} and thus $\Delta V = V_{lh} - V_0$, then the q required for abatement is smaller than if ΔV does not change. Both channels through which assignments may work push in the same direction. The consequence for abatement still depends on the distribution of firm costs.

The response to the experiment could take several forms depending on how abatement costs compare to the value of auditor reputation and regulatory fines. Each of the audit treatments weakly induces abatement by raising the cost of inducing false reporting from auditors. If the full treatment of backchecks, assignments and audits is weak, auditors will falsely report and firms will not abate. If the treatment is strong enough, in that $q(V_{lh} + B) \geq \Delta V$, then the response to the treatment will be a combination of abatement, from a $G(f)$ share of firms, and higher reporting from the rest. With false reporting more costly, firms trade-off abatement and the risk of regulatory fines, and abate if their abatement costs are low. The composition of the response to the experiment, provided the treatment is strong, will then reveal the extent to which firms are able to abate at costs less than the fine for a detected violation.

IV Data and Empirical Strategy

A Data

The key outcomes of interest in the experiment are the accuracy of auditor reporting and the pollution response of client plants. Two data sources are used to measure accuracy. First, audit reports filed by auditors working in the treatment or control with the GPCB for both 2009 and 2010. These reports cover a mandated set of water and air pollutants, described in Table 9. Depending on the number of stages of effluent treatment, pollutant samples may be taken during audits prior to any treatment, after primary treatment, after secondary treatment, or at the final outlet from the plant. For most of the analysis, we focus on final-outlet concentrations for water samples, as these are the readings that have a direct impact on the environment and are therefore most closely attended by both auditors and GPCB. Similarly air samples may come from stacks (chimneys) attached to industrial boilers, thermic-fluid heaters or various industrial processes, such as chemical reaction vessels. We focus on boiler-stack samples for the widest comparability across the entire sample, as most

plants have boilers.

Data on audit reports is only available if a plant submitted an audit report in a given year. Not all plants in the sample submit audits, despite being prima facie audit-eligible, for various possible reasons.[14] As shown in Table 2, plants were about as likely to submit audit reports in the treatment as in the control. Control firms were slightly more likely to report in the first year and treatment in the second year. In the treatment, 70% of plants submitted audit reports in each year. In the control, 74% and 64% of plants submitted audit reports in years 1 and 2, respectively. The rates of submission during the experiment are comparable to what GPCB records indicate for submission in 2008. As shown in Table 1, 72% and 69% shares of plants submitted in treatment and control, respectively, before the experiment began. The treatment therefore does not appear to have induced plants to submit audit reports that would otherwise have not and so the samples of pollutant readings we consider below are not differentially selected.

The second source of data for accuracy comes from the backchecks, which were conducted in a sample of treatment plants throughout 2009 and 2010 and in a single wave, which we call the "midline" survey, after the third season of audit visits in 2010 in both the treatment and control. Backchecks were performed in the treatment throughout the experiment but the control only at the end of year 2, so analysis involving a difference between backcheck and actual readings will only use year 2 data. Backchecks were conducted on the same pollutants and locations as the audits, and were scheduled to come within no more than 90 days of the initial sampling. The median backcheck lag in the midline sample was 36 days.[15] This midline survey across treatment conditions enables direct measurement of the comparative accuracy of auditors working in the treatment and control groups of plants.

The midline survey sample was drawn from the 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

[14] GPCB had judged them not-audit-eligible in the past, they had changed products since the data used to determine eligibility was compiled, they protested their audit-eligible status, they closed or they simply chose not to submit and incur the risk of a penalty rather than hire an environmental auditor.

[15] It was 6 days in the rest of the study, when backchecks were used for monitoring and incentives.

in the treatment at the same time. 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 randomly selected plants, amongst plants for which auditors submitted a date for the audit visit, and finally randoml plants that had not submitted a date.

Despite the non-random sampling of observations for backchecks during the midline, the overall composition of this sample remains well-balanced along the fifteen observables that were shown in Table 1. The 130 plants in the midline treatment and control significantly differ at the 5% level in that midline treatment plants are less likely to have a bag filter installed, as are treatment plants in the overall sample, and consume more water and generate more waste water than midline control plants. The primary outcome studied in the midline sample is the accuracy of auditor reporting, which is the difference between audit and backcheck readings, and thus control for the true pollution levels of the plant. These few differences on observables at baseline would therefore need to differentially effect audit and backcheck readings to influence results drawn from the midline sample.

The source of data on plant pollution response is an endline survey conducted from April through July of 2011. The mean survey date, at the end of May, was approximately six months after the surveyed plant would have had their final audit visit under the audit treatment and three and a half months after the audit reports in year two were due to the regulator. The survey was conducted by mostly the same agencies that did backchecks and included the same pollution samples discussed above, in addition to economic and abatement cost measures that we do not use in this paper. The endline survey covered firms outside the audit sample that were subject to an additional experimental treatment. We use survey data only on those firms that were in the audit sample and in the control group for the other experimental treatment.

B Empirical Strategy

Given the randomized design, the empirical strategy is straightforward. To measure the impact of the audit treatment on reported pollutant concentrations, accuracy of reporting, or actual pollution, we estimate ordinary least-squares regressions of the form:

$$y_{ij} = \beta T_j + \alpha_r + \alpha_y + \alpha_a + \epsilon_{ij},$$

where y_{ij} is an outcome variable of interest i at plant j, α_r are fixed effects for the region $r \in \{Ahmedabad, Surat\}$, α_y are fixed effects for the year $y \in \{2009, 2010\}$.[16] Pollutant measures are standardized by subtracting the mean and dividing by the standard deviation of the same pollutant reading amongst backcheck samples at the same sampling location. In some specification, we introduce fixed effects α_a for the auditor a, to focus on the within-auditor difference in behavior across the two incentives systems. We pool standardized measures for different pollutants in most specifications and cluster standard errors are at the plant level.

Our first outcome is the recorded concentration of pollutants, at the plant level, in the audit reports. Audit reports are a composition of the "true" level of a pollutant and measurement error in the sampling, as well as the auditor's manipulation of the results. If the audit treatment affects the true level of pollution, by changing plant behavior, then changes in audit reports of concentrations will reflect changes in polllution as well as changes in reporting. The audit report may thus be a lower bound on the changes in reporting behavior. We still present this data, in addition to the difference between matched backchecks and audits, because it is available for a much larger sample.

Our second outcome of interest is the difference between back check and audit reading, which accounts for any possible effect of the intervention on the actual pollution levels. These differences are available on a comparable basis only in the midline sample, but are more direct measures of reporting.

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

where y_{ij}^{D} is the backcheck reading less the audit reading. Samples are matched on pollutant i, plant j, sampling location (boiler stack or final outlet) and date. Positive values indicate underreporting and negative overreporting. Our third outcome, from the endline survey, is actual pollution at the firm level.

We also use two approaches to model how reporting changes at different points in the pollutant distribution. First, we estimate OLS regressions as above where the dependent variable is an indicator for a pollutant reading being within a given density bin. Density bins are 0.05 standard deviations wide for each pollutant and centered around the regulatory standard, so that weakly negative values indicate compliance and positive values non-compliance. Second,

[16] In specifications including only one year of data, year fixed effects are omitted.

we also estimate quantile regressions of the form

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

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

V Results

A Audit Reports

The distributions of audit report pollutant readings suggest that auditors report pollution in treatment plants more accurately than in control plants. Figures 1 and 2 show the distributions of pollutant concentrations, as reported by three different sources, for two representative pollutants, Bio-chemical Oxygen Demand (BOD), a water pollutant, and Suspended Particulate Matter (SPM), an air pollutant. The top panel in each figure shows audit reports in the control, the middle panel shows audit reports in the treatment and the bottom panel shows the distribution of readings from independent backchecks of pollutant readings. In each figure the pollutant standard or standards are shown by a vertical line.[17] On each panel of each figure, we shade the area between 75% of the standard and the standard in grey. The first two panels show the excess mass of the respective control and treatment audit distributions, relative to the same region in the backcheck distribution for that pollutant.

The figures show a striking pattern, in which reports from control audits are clustered immediately below the standard for each pollutant, while reports from treatment auditors are not. Reports from auditors in the treatment show smooth distributions very similar to the backcheck distributions. For example, Figure 1 shows the distribution of Biochemical Oxygen Demand concentrations. In Panel A, there is a large spike of non-CETP readings, shown by the light bars, in the audit control immediately below the non-CETP standard, with excess mass of 0.534 (standard error 0.014) between 75% of the standard and the standard, relative to the backcheck distribution in Panel C. There is a broader rise in the control distribution beneath

[17] Some water pollutants have different standards depending on whether the effluent goes to a Common Effluent Treatment Plant (CETP) after leaving the plant or goes directly to a drain or receiving body (non-CETP). For such pollutants, the distribution of samples taken from CETP plants is shown by the dark portion of the histogram bars and the samples taken from non-CETP plants by the light portion of the bars

the non-CETP with excess mass of 0.122 (standard error 0.027) relative to the backcheck distribution, which exhibits no spike just below the standard.[18] In the audit treatment group, shown in Panel B, BOD readings show greatly reduced excess mass of 0.148 (standard error 0.014) in the area beneath the standard non-CETP standard and no excess mass beneath the CETP standard. The pollutant density is still greatest at the standard in the treatment, but the distribution of readings is much flatter and broader, like the distribution from backchecks, shown in Panel C. Figure 2 shows a similar patter for an important air pollutant, suspended particulate matter. There is an excess mass of nearly 50% of all observations beneath the standard, shifted from both the right tail of the distribution, where there are far fewer high readings than in the treatment or backcheck distributions, and from the left tail as well. We attribute the decrease in the left tail to the relatively high cost of sampling air pollutants. It may be less costly for auditors to report narrow compliance by default than to properly sample and document a very low reading.

Table 3 summarizes calculations of the excess mass in the audit treatment and control relative to the backcheck distribution, by year, for a range of water and air pollutants. Across pollutants, auditors in the control are far more likely to report that plants are just beneath the pollutant standards than auditors in the treatment. The excess mass beneath the standard for all pollutants together is 0.297 (standard error 0.005) in the control, and 0.0743 (standard error 0.0053) in the treatment (Panel A, columns 5 and 6).

Individually, the other water and air pollutants in the table show a pattern similar to that for BOD. The excess mass beneath the standard in the control, shown in Panel B, column 6 for both years together, ranges from a minimum of 0.154 to a maximum of 0.66. All of these are statistically greater than zero by wide margins. The excess mass beneath the standard in the treatment, shown in column 5, ranges from 0.0023 to 0.187. Most water pollutants show significant excess mass but of a relatively small magnitude. Air pollutants (SPM and SO_2) show substantial excess mass beneath the standards in the treatment, though still far less than in the control. For example, SPM shows excess mass of 0.187 in the treatment but 0.498 in the control.

Table 4, Panel A, shows ordinary least-squares regressions of standardized audit pollutant

[18] Such spike may have been observed without misreporting if firms were able to abate pollution just enough to be in compliance with the standards.

concentrations, pooled by pollutant type, on audit treatment status. [19] This table includes the full sample of all pollutant readings reported across the two years of the experiment by auditors in both treatment and control. As shown in column 1, the mean audit report reading for plants in the audit treatment is 0.103 standard deviations (with a standard error of 0.0354 standard deviations) higher than the mean report in the control for all pollutants. The coefficients on the audit treatment are similar for both water and air pollutants considered separately, as shown in columns 3 through 6. Adding auditor fixed effects, as shown in column 2, somewhat increases the point estimate of the audit treatment to 0.131 standard deviations (with a standard error of 0.0384 standard deviations).

The magnitude of these effects is quite large in terms of the units of the original pollutants. 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 Bio-chemical Oxygen Demand (BOD) estimated alone (not shown). The standard deviation of BOD in backchecks in final-outlet samples where the effluent did not go on to further treatment 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 would thus be 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 the discussion of the distributions above, then, these specifications show that the change in pollutant reports is large enough to shift the reports for many plants from compliance to non-compliance and is economically significant, in the sense of representing a meaningful increase in reported pollution.

Figure 3 summarizes the changes in the density of the reported pollutant distribution due to the audit treatment. For this figure, all pollutant readings are standardized by subtracting the regulatory standard and dividing by the pollutant standard deviation in backchecks, so that the horizontal axis indicates the number of standard deviations above or below the regulatory limit. We regress indicator dummies for a pollutant reading belonging to a particular 0.05-standard-deviation-width bin on an indicator for audit treatment. The figure clearly shows that the treatment dramatically reduces the amount of mass just beneath the regulatory standard. Density in the bins above the standard rises. Figure 4 presents quantile

[19] Pollutants included are $Water = \{NH_3\text{-N}, \text{BOD}, \text{COD}, \text{TDS}, \text{TSS}\}$ and $Air = \{SO_2, NO_x, \text{SPM}\}$ with $All = Water \cup Air$.

regressions showing where this mass goes. Quantile regressions are estimated at each quantile from 0.05 to 0.95 in 0.05 increments, to see how the audit treatment affects different parts of the reported pollutant distribution. Incentives for accurate reporting that implicitly or explicitly depend on the absolute difference between auditor readings and the actual pollutant value may have different reporting effects at different points in the pollutant distribution if the costs and benefits of misreporting vary across that distribution. The figure shows that the treatment effect for all pollutants is driven by large movements in the tails of the distribution. At around the 0.10-quantile and the 0.85-quantile, we estimate a treatment effect of 0.15, somewhat greater than the mean effects in Table 4. In the middle of the distribution effects are indistinguishable from zero. The standards for each pollutant fall at different precise quantiles in the standardized distribution but are mostly gathered in the left tail. The auditors working in the treatment thus appear more likely both to report readings just above the standard, as shown by the large treatment effects at low quantiles, and to report very high readings, as shown by the large treatment effects in high quantiles. In the middle of the distribution, where a report may already incriminate a plant but not very much so, the treatment effects are small.

These results suggest that there is a cost to the degree of misreporting and a value to firms to avoid reporting figures that are extremely high. If a modest amount of misreporting is sufficient to show that the firm is compliant, the auditor will offer a compliant report. In such a case any discrepancy found by the regulator might be due to measurement error or changes in the conditions at the plant. If the true pollution readings are so high that they would almost certainly attract the regulator's attention, auditors are also willing to shade sharply downwards. This shading, in the model, requires that auditor payment rise to compensate for the risk of being caught reporting falsely.

B Differences Between Audit Reports and Backchecks

While available for a smaller sample (midline survey only), the difference between backcheck and audit readings is a more robust measure of auditor reporting than audit report levels as this difference accounts for actual pollution. In particular, if plants under the audit treatment abated pollution, as under a strong treatment regime in the model, then differences in audit report levels would understate the treatment effect on reporting, as the treatment effect would

be a composition of lower pollution levels with higher, more truthful reporting.

Matched backcheck-audit differences in Table 4, Panel B, clearly show that auditors report more truthfully in the treatment group of plants. The absolute difference between backcheck and audit readings is 0.210 standard deviations (standard error 0.0726 standard deviations) smaller in the treatment for all pollutants.[20] As shown by the control group means in the table, backcheck readings are larger than audit readings in the control, indicating that audits tend to underreport pollution concentrations in that group. The negative treatment effects therefore mean that auditors in the treatment are getting closer to the true pollutant readings. That the estimates for differences are generally somewhat larger than the estimates in audit levels, especially for air pollutants, indicates that, while the audit readings are rising to meet the backchecks in the treatment, the backcheck readings are also falling at the same time. We discuss plant response to the treatments in Section V D below.

Unlike in the specifications in audit levels, the treatment coefficient in audit differences becomes somewhat smaller in specifications with auditor fixed effects. For example, in column (2), the treatment effect is estimated to be -0.153 standard deviations (standard error 0.0986 standard deviations) as opposed to the -0.210 baseline specification estimate. Columns (4) and (6) show this change to be driven entirely by air pollutants, as the coefficients for water pollutants are about the same in specifications with and without fixed effects.

To benchmark this improvement in auditor accuracy, Table 5 presents estimates of the backcheck - audit differences in the treatment group during different time periods. As shown by Panel A, audit reports are inaccurate in year 1 (2009) and improve over time, as shown by Panels B and C. The accuracy of treatment auditors for all pollutants improved by 79% (=(0.379-0.078)/0.379) over the experiment and by the midline survey, taken at the end of the experiment, auditor reports are statistically unbiased. The mean backcheck-audit difference, shown in Panel C, column 1, is 0.078 standard deviations (standard error 0.057 standard deviations).

There is a small but robust difference between results for water and air pollutant readings running through the results on auditor reporting. Most of the improvement shown in Table 5 occurs through a 66% (=(0.513-0.176)/0.513) improvement in the accuracy of water pollutant

[20] In estimates of the specifications of Table 4 including the number of days between the audit and backcheck as a control variable, not shown, this lag is uniformly not significant. The treatment effect estimates become somewhat smaller, with the treatment effect for all pollutants -0.181 standard deviations (standard error 0.0789 standard deviations) as opposed to the -0.206 reported when not controlling for the backcheck lag.

reports, which were far less accurate to begin with and remain less accurate in the treatment. Water pollutants show greater excess mass below pollutant standards in Table 3. We attribute water pollutants still displaying some bias to the perception amongst firms and auditors that GPCB cares more about water pollution. In the auditor survey described in Section II C, we asked auditors what pollution measurements were viewed as most important by plants and GPCB. Auditors responded that pH and COD, both measures of water quality, were the most important to both parties. This perception is justified. In GPCB's own files, firms are most frequently cited for violations of standards for COD and BOD, both water pollutants. In Figure 5 we plot the excess mass displayed in the treatment firms against the share of regulatory actions that cited each pollutant. We measure regulatory violations by entering all regulatory actions for plants in the study sample from 2007 through mid-2011, where each action includes a citation of reason. The plot shows that excess mass is larger in the control for precisely those pollutants that are most likely to be cited against firms. An increase of one percentage point in citation share leads to a one percentage point greater excess mass for a given pollutant, with a remarkably strong fit. Auditors are responsive to incentives for accuracy overall and also responsive to incentives to mute potentially incriminating reports at the level of the individual pollutant.

C Unbundling the Experiment: Role of Incentive Pay and Backcheck Treatment Components

The experimental treatment was composed of several components, each of which may independently have had an effect on auditor reporting. In this section we look within the audit report data to estimate, using non-experimental variation, whether the treatment components of direct incentive pay and backchecks of audit accuracy had independent effects of reporting.

Incentive pay. The treatment in year two included direct incentive pay for auditors based on the accuracy of their reports. Comparing the size of the treatment effects in the two years of the study may therefore help to decompose the treatment into the effect of auditor independence, in assignments and means of payment, and the effects of direct incentives for accuracy. Table 6 presents estimates of the treatment effects for all pollutants with controls for time to isolate the effect of incentive pay. The first column duplicates the base specification of Table 4 for reference showing the effect of the dummy for the year being equal to 2010,

when incentive pay was in effect. The second and third columns add an interaction of that dummy with the treatment and a linear trend in fractional years since January 1, 2009. We present these specifications for completeness but do not expect them to be reliable given that the results above suggest levels of pollution may be differentially decreasing in the treatment plants. Columns 4 and 5 therefore add a separate linear trend for the treatment group. In these specifications, without and with auditor fixed effects, there is a negative downward trend in reported pollutant concentrations in the treatment group, as shown by the coefficient Years (fractional) X treatment, but no trend in the control, as shown by the main coefficient on Years (fractional) from Jan 1, 2009, of 0.0290 standard deviations per year (standard error 0.0226 standard deviations per year).

Against this trend, there is a positive interaction between incentive pay, the dummy for the year being 2010, and treatment. In column 4, without auditor fixed effects, this interaction coefficient is 0.257 standard deviations (standard error 0.0920 standard deviations), larger than the overall estimate of the treatment effect. This result is robust in the fixed effect specification; it is also robust if we allow for quadratic or cubic trend (not shown). Auditor reports were therefore discretely higher in year two than in year one after controlling for pollutant trends in the treatment and control separately, which is notable given that, within years, pollutants in the treatment are trending downwards. This variation is not experimental so the interpretation of the dummy for 2010 rests on a correct specification of the time trend. Still, this result suggests that there was a direct positive impact of incentive pay.

Backchecks. Backchecks alone also appear to improve auditor reporting. In Table 7 we study the accuracy of auditor reporting within the midline survey. In this survey, treatment firms were backchecked, as they had been throughout the experiment, but at higher rates, while control firms were backchecked for the first time.

Auditors were repeatedly told that their work in the treatment firms was going to be backchecked. They were not told they could be backchecked for their work in the control firms, though we solicited approximate dates of audit visits during the midline from these auditors. We expect that during the midline survey having been backchecked once previously could have increased the subjective probability that backchecks were now happening both in treatment and in control. We therefore examine whether the performance of auditors changed with the exposure to backchecks that occurred prior to a given audit within this survey period.

The first column of Table 7 regresses audit accuracy on treatment status, an indicator for whether the auditor been subject to any prior recent backchecks (during the midline period) in any of the firms it audited, and an interaction of this term with audit treatment. Having had a prior backcheck in the midline is associated with a -0.170 standard deviation (standard error 0.0997 standard deviations) improvement in audit accuracy in the control, shown by the main effect in the third row of column (1). It is not associated with any improvement in the treatment, as there is a positive offsetting effect in the treatment interaction term in the second row. This is what we expected to see: recent backchecks are associated with greater accuracy in the control, but not in the treatment, where auditors already knew that they could be backchecked. In column (2), with auditor fixed effects, the point estimates for the coefficients on these terms are roughly the same, though far less precisely estimated. In columns (3) and (4) we separate the backcheck variable into a past backcheck indicator in an auditor's treatment plants and in an auditor's control plants separately.[21] We could have expected that a recent backcheck in a control firm causes auditors to expect control firms might be backchecked more in the future, and hence should have a larger effect, but the point estimates weawkly suggest the opposite. These estimates are sufficiently imprecise that the effect of past backchecks in the auditors' control and treatment plants are not distinguishable. Overall, it does appear that recently backchecked auditors improve accuracy in control plants despite not being subject to the other components of the intervention.

D Response of Plants to Audit Treatment

The experiment induced a large improvement in auditor reporting. The purpose of environmental regulation is to reduce actual pollution, which may not have improved even with truthful audit reporting. The model suggests that the change in audit reporting may have changed the firm's decision margin to whether to abate or incur a fine, rather than whether to abate or purchase a compliant report. Whether or not firms abate in response to the treatment then depends on the relation of abatement costs to regulatory fines. If the stakes are low and abatement costs are high, so that few firms have abatement costs less than regulatory fines, then we would expect no pollution response. Conversely, firms with abatement costs

[21] We cannot estimate an interaction term for being in the audit treatment times having had any control backchecks before an audit, as no backchecked audit in the treatment during the midline was preceded by a midline control backcheck.

less than the fine expected for a non-compliant report will reduce their pollution emissions in response to the audit treatment.

We find that firms reduce pollution significantly in response to the audit treatment. Table 8 shows regressions of pollution concentrations measured during the endline survey on audit treatment status.[22] The units are standardized in the same manner as in the auditor accuracy tables above. Firms subject to the audit treatment reduced pollution by 0.211 standard deviations (standard error 0.099 standard deviations, significant at the 5% level) on average. This reduction comes almost entirely from a large decrease in water pollutant concentrations, shown in column (3), as opposed to air pollutants, shown in column (5). The volume of effluent emitted (not shown) did not change in response to the experimental treatments, so these reductions in concentrations represent reductions in the total emitted effluent load—less water pollution—amongst treatment firms. The even-numbered columns include auditor-by-year fixed effects to control for the influence of having each auditor in each year separately. In these specifications the overall treatment effect grows to a decrease of 0.456 standard deviations for all pollutants.

The reform of auditor reporting thus caused firms working with independent auditors to reduce pollution emissions. This decrease comes from the water pollutants that are regulatory priorities and which were the original spur for the development of the audit scheme.

VI Conclusion

This paper provides experimental evidence that third-party environmental auditors report more accurately when working under a system that encourages independence. Client plants, in turn, respond to greater audit scrutiny by reducing pollution output.

In the audit treatment, plants were randomly assigned to auditors, rather than having plants select auditors, plant auditors were paid from a central pool, rather than by plants directly, and auditors were backchecked by independently measuring the same pollutant readings that they reported. In addition, in one year of the study auditors were directly paid for the accuracy of their reports.

Over the course of a two-year evaluation, auditors report much more accurately on the

[22] The audit intervention was conducted concurrently with another treatment at the plant level. We restrict the pollution sample in this table to the subset of audit sample firms not subject to the other experimental intervention.

treatment plants than on the control plants. In control plants, misreporting is ubiquitous, especially just beneath the pollution standard, and stronger for the pollutants that the regulator cares most about, as indicated by citations of pollutants in regulatory violations. On average, readings in the treatment group for a standardized pollutant index of water and air pollutants are significantly higher and more accurate. By the end of the second year of the experiment, reports in the treatment group were unbiased. This treatment effect comes largely from the tails of the distribution, where auditors in the treatment are much more likely to report both small violations of the pollutant standard and extremely high, potentially incriminating readings. Additional, non-experimental analysis suggest that the both financial incentives and backcheck contributed to the change in auditor reporting.

Plants randomly assigned to the audit treatment responded to tighter monitoring by reducing their pollution output as measured about 6 months after the end of the audit intervention. This reduction in pollution was on the same scale as the observed change in audit reporting, though many firms in the treatment remained out of compliance with tight regulatory standards. Here again, the effects are stronger for water than air pollutants.

The results observed in this experiment can be interpreted in the context of a simple model of firm and auditor behavior. The two main results of the experiment are that auditor reporting became almost completely accurate and that, while firms abated on average, there were still many firms out of compliance. This is consistent with a model regime where there is heterogeneity in abatement costs amongst firms, many firms have abatement costs greater than the anticipated cost of fines for high reports, and auditors had a significant continuation value from participating in the audit market. The experimental treatments were strong, in the sense that they moved the value of buying falsely low reports sufficiently that the relevant decision margin for firms was between abatement and open non-compliance. The sea change in auditor reporting towards accurate, high reports argues that for many firms the expected regulatory fine was less than abatement cost. The observed reduction in pollution levels shows that for a significant fraction of firms abatement was expected to be less costly than a high audit report.

We attribute these results to greater independence of auditors from client firms and incentives for accurate reporting. The audit reforms raised the cost for auditors of misreporting and, with assignment to plants, removed the potential benefit of ensuring continued business.

Other explanations for why treatment auditors reported more accurately in this study can be ruled out. The treatment effects are not experimental Hawthorne effects of studying auditors, or of auditors anticipating that reputation might matter more under revised audit regulations in the future, as the treatment effects are observed even comparing the behavior of the same audit firms under different incentive systems, and response to the experiment *per se* would occur at the auditor level. The differences in behavior are also not attributable to auditor income effects, whereby receiving a fixed audit rate above cost allowed auditors to perform their jobs better in the treatment. Receiving a larger payment, in the absence of any changes in the cost of auditing effort or the benefits of accurate reporting, rational auditors would pocket the difference rather than spend it to improve audit quality.

These results are encouraging in the specific context of environmental regulation in India. One of the challenges of regulation is gathering accurate information on the performance of regulated firms in the face of agency problems either in third-party reporting or within the regulator itself. Not only the unique environmental audit system in Gujarat but also 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). In this study a package of reforms designed to foster greater auditor independence led to much more accurate reporting and lower pollution within a relatively short amount of time. This reform shows it is feasible for the State Pollution Control Boards (SPCBs) to collect accurate information on pollution and use it to enforce regulatory standards.

The audit scheme in Gujarat was introduced to emulate "external audit[s] in the fields of company law and income tax" and succeeded all too well (Gujarat High Court, 1996). While the core issue of a conflict of interest is common to this study's setting and those of corporate auditing or credit ratings, there are some limitations to how well the results may transpose to those settings. One limitation is that in other regulatory and competitive environments auditors might want to develop a reputation for being accurate. As pointed out by Ronen (2010) and Bolton et al. (2011), though, reputational incentives can cut both ways. Auditors may want to develop a reputation for leniency to attract more business, as long as there are enough naive investors or other clients to accept their recommendations at face value. Moreover, while audit companies are large and may want to internalize the reputational

externalities of misreporting, it is less clear that individual managers or analysts have the same incentives (Bar-Isaac and Shapiro, 2011). The evidence on the quality of corporate financial auditing is not very convincing, as quality is measured *ex post* by instances where the SEC decided to investigate, but the financial crisis gave a window into the quality of the credit ratings. The massive downgrades that had to take place suggest that the ratings were greatly inflated, in particular in the boom years (Ashcraft et al., 2010).

A second limitation on the range of the study results is that the annual rotation that was imposed in our scheme may not be feasible for the financial audits or credit ratings of large firms, where there may be large fixed costs in understanding a firm's finances. Modifying the form of rotation to guarantee auditors would work with a client for, say, five years before being reassigned, unless found to be grossly inadequate under regulatory review, would serve the same purpose, of severing the link between auditors' work and firm satisfaction, while reducing the importance of fixed costs. The details of a modified third-party audit system would clearly need to be adapted to each setting, but this study shows that, in one context, conflicts of interest matter enormously, and reforms that make auditors genuinely independent can bring about truthful reporting and greater regulatory compliance.

VII References

References

Ashcraft, A., Goldsmith-Pinkham, P., and Vickery, J. (2010). Mbs ratings and the mortgage credit boom. *Federal Reserve Bank of New York Staff Reports*, (449).

Bar-Isaac, H. and Shapiro, J. (2011). Credit ratings accuracy and analyst incentives. *American Economic Review, Papers & Proceedings*, 101(3):120–24.

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

Bolton, P., Freixas, X., and Shapiro, J. (2011). The credit ratings game. *The Journal of Finance*, Forthcoming.

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.

Chen, Y., Ebenstein, A., Greenstone, M., and Li, H. (2010). The 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).

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

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 Greenstone, M. (2011). Environmental regulations, air and water pollution and infant mortality in india. *CEEPR Working Paper*, 2011(14).

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

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

International Monetary Fund (2009). Global financial stability report.

Kreps, D. and Wilson, R. (1982). Reputation and imperfect information. *Journal of economic theory*, 27(2):253–279.

Mathis, J., McAndrews, J., and Rochet, J.-C. (2009). Rating the raters: Are reputational concerns powerful enough to discipline rating agencies? *Journal of Monetary Economics.*

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.

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.*

VIII Figures

Figure 1: Audit Readings for Bio-chemical Oxygen Demand (BOD) by Treatment Status

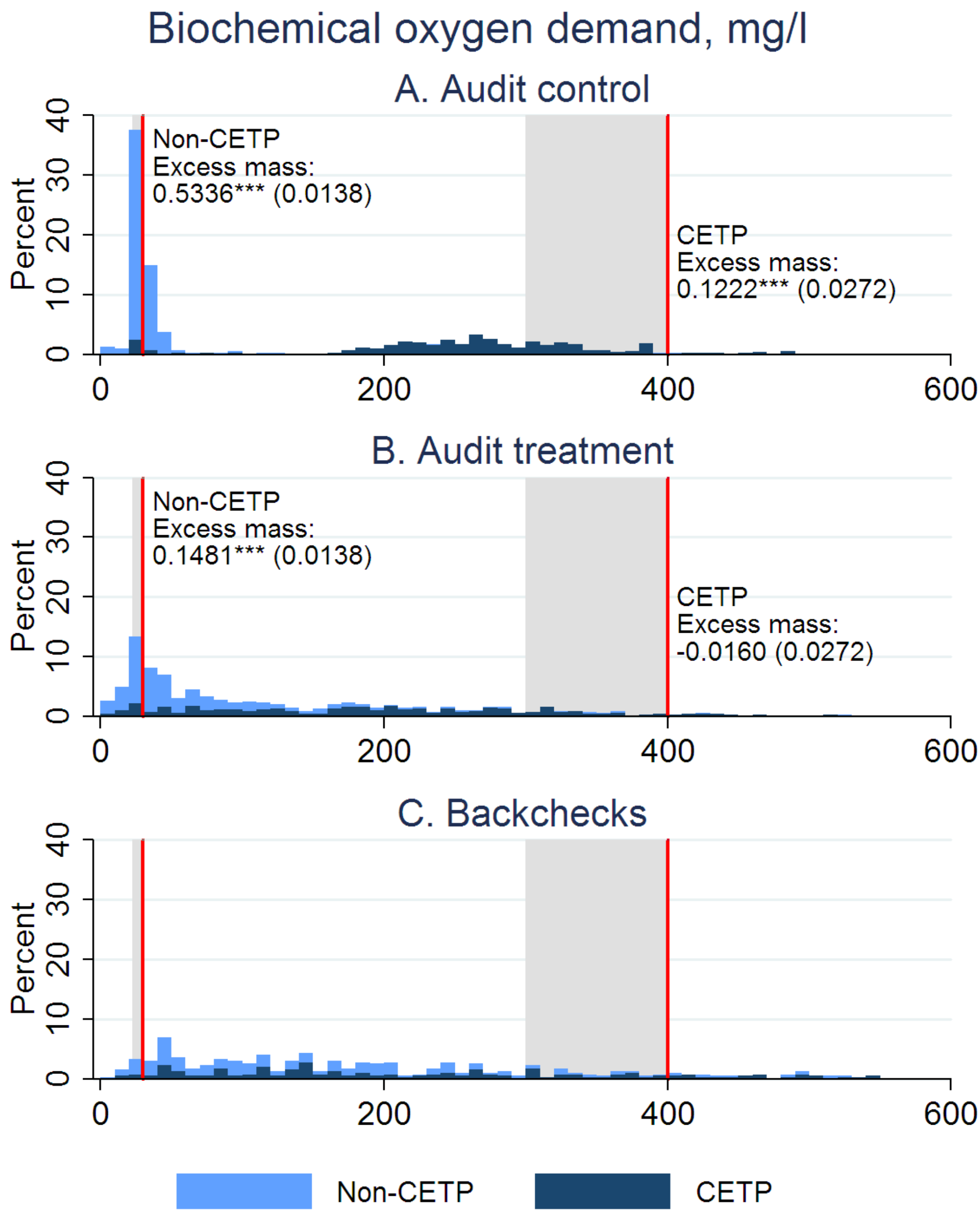


Figure 2: Audit Readings for Suspended Particulate Matter (SPM) by Treatment Status

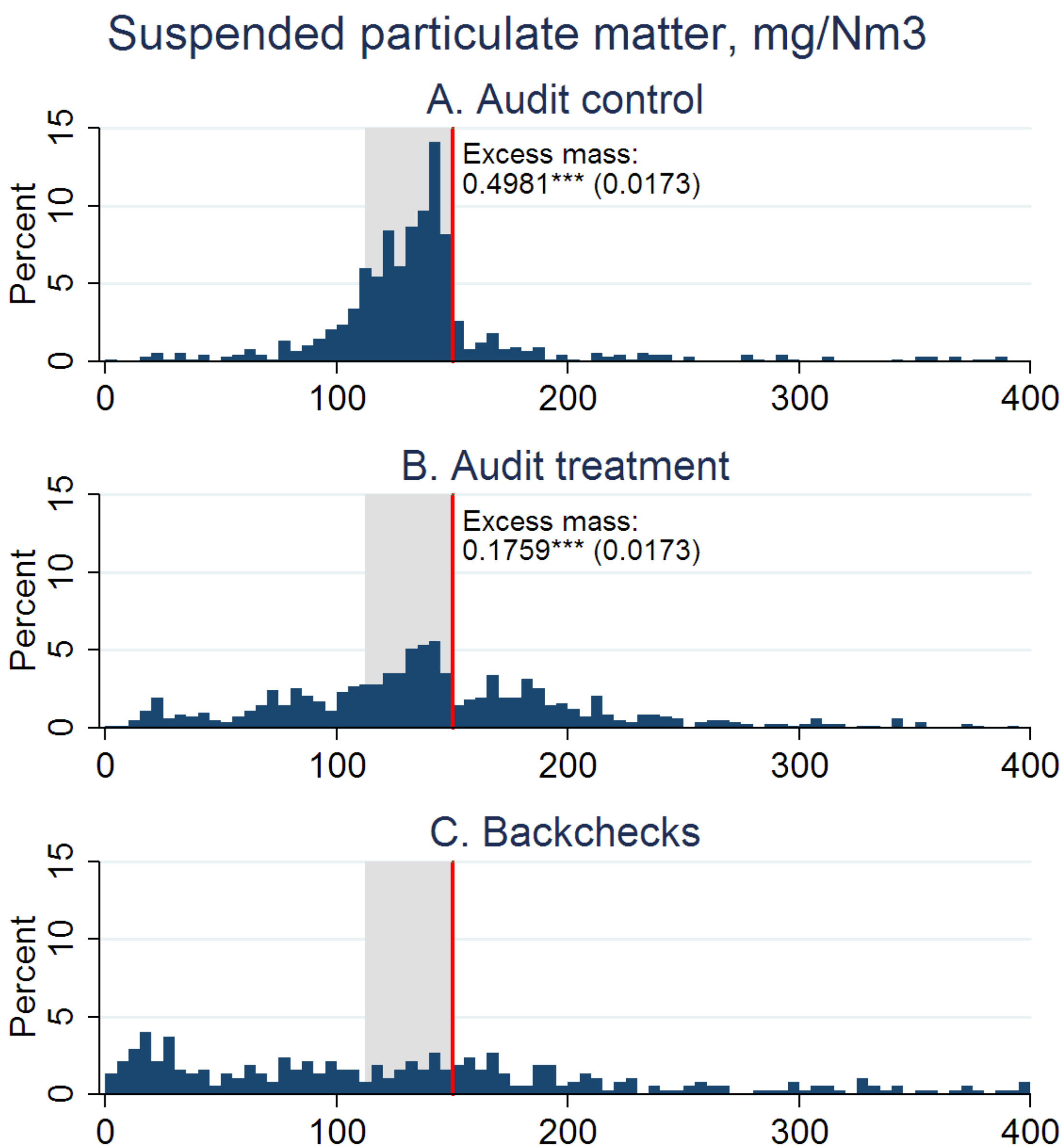


Figure 3: Audit Treatment Effect in Density Bins, All Pollutants

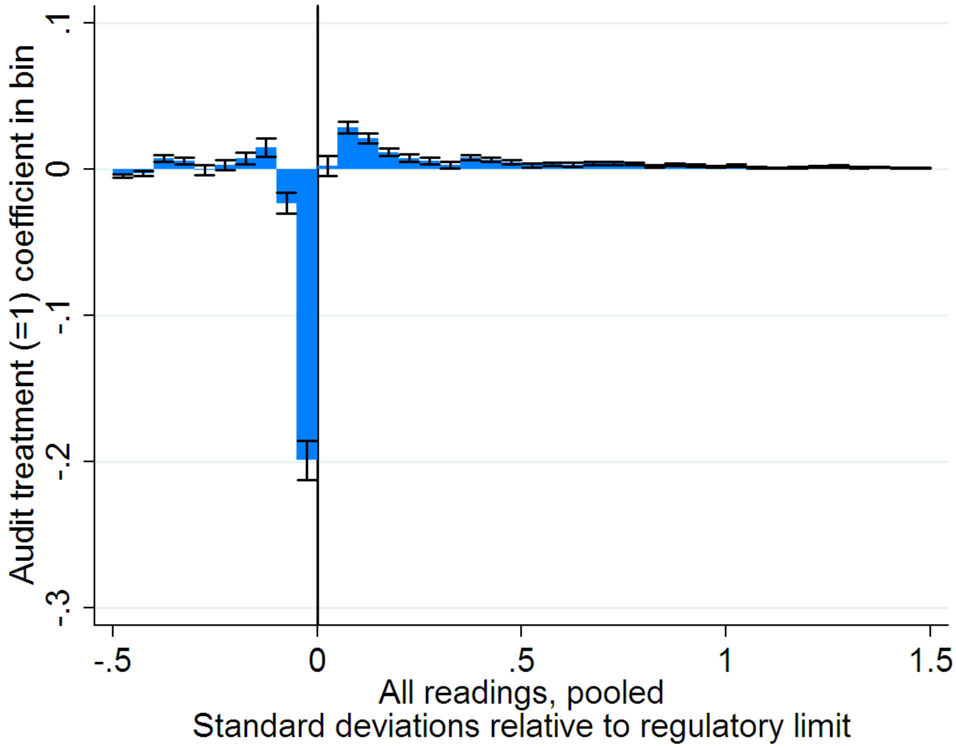


Figure 4: Quantile Regression of Pollutant Report on Treatment Status

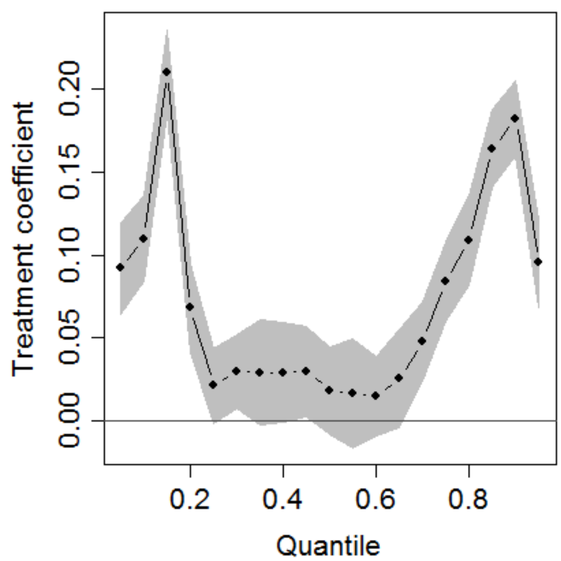
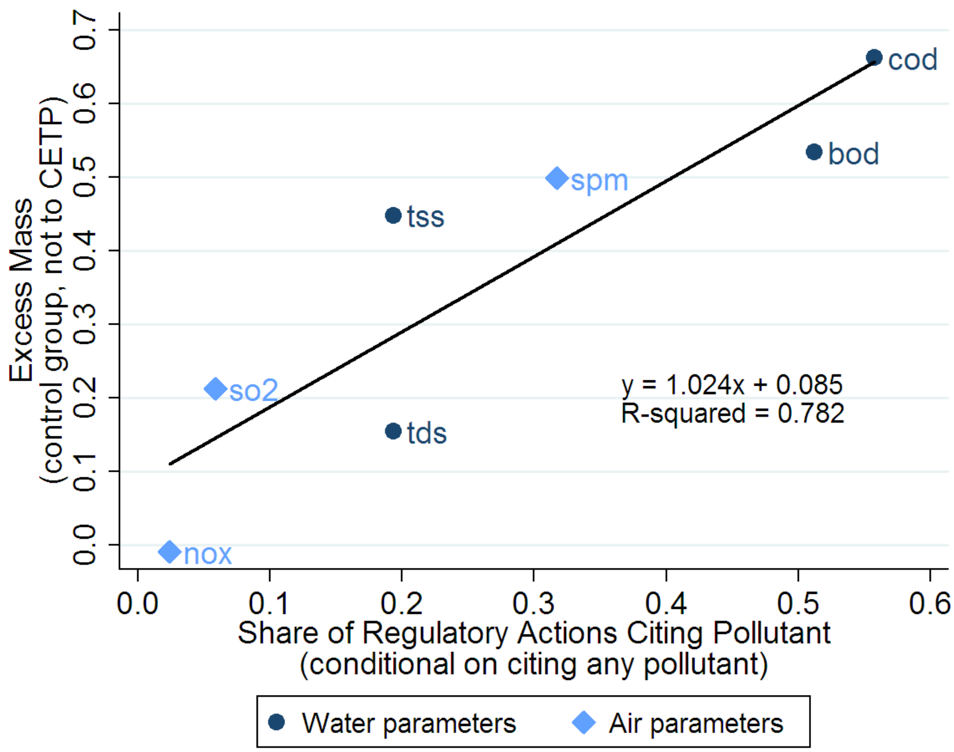


Figure 5: Excess Mass Against Parameter Citations in Violations



IX Tables

Table 1: Audit Treatment Covariate Balance

	(1)	(2)	(3)
	Sample mean [sd]		
	Treatment	Control	Difference
Capital investment Rs. 50m to Rs. 100m (=1)	0.099	0.13	-0.035
	[0.30]	[0.34]	(0.030)
Located in industrial estate (=1)	0.55	0.53	0.020
	[0.50]	[0.50]	(0.046)
Whether audit submitted in prior year (=1)	0.72	0.69	0.025
	[0.45]	[0.46]	(0.042)
Textiles (=1)	0.81	0.83	-0.026
	[0.40]	[0.37]	(0.035)
Dyes and Intermediates (=1)	0.094	0.083	0.011
	[0.29]	[0.28]	(0.026)
Effluent to common treatment (=1)	0.39	0.34	0.053
	[0.49]	[0.48]	(0.044)
Water consumed (kl / day)	490.1	449.6	40.5
	[488.4]	[599.8]	(50.4)
Waste water generated (kl / day)	391.4	355.9	35.5
	[335.4]	[345.5]	(31.3)
Lignite used as fuel (=1)	0.64	0.66	-0.019
	[0.48]	[0.47]	(0.044)
Diesel used as fuel (=1)	0.29	0.26	0.029
	[0.45]	[0.44]	(0.041)
Air emissions from flue gas (=1)	0.80	0.80	0.0024
	[0.40]	[0.40]	(0.037)
Air emissions from boiler (=1)	0.88	0.89	-0.016
	[0.33]	[0.31]	(0.030)
Bag filter installed (=1)	0.21	0.30	-0.090**
	[0.41]	[0.46]	(0.040)
Cyclone installed (=1)	0.094	0.10	-0.0056
	[0.29]	[0.30]	(0.027)
Scrubber installed (=1)	0.41	0.41	-0.0048
	[0.49]	[0.49]	(0.045)
Observations	233	240	

All variables from administrative data prior to the experiment for audit sample. Column (3) shows differences between treatment and control with standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Submission of Audit Reports

	Treatment	Control
Panel A: 2009		
Audit submitted	163	177
Total plants	233	240
Share submitted	0.70	0.74
Panel B: 2010		
Audit submitted	164	153
Total plants	233	240
Share submitted	0.70	0.64

Table 3: Excess Mass Below Pollutant Standards

		(1)	(2)	(3)	(4)	(5)	(6)
		Year 1 (2009)		Year 2 (2010)		Both Years	
		Treatment	Control	Treatment	Control	Treatment	Control
Panel A: Pooled by Pollutant Media							
All		0.0611	0.3223	0.0878	0.2725	0.0743	0.2969
		(0.0113)	(0.0113)	(0.0062)	(0.0062)	(0.0053)	(0.0053)
Water		0.0364	0.3473	0.0746	0.3176	0.0589	0.3362
		(0.0162)	(0.0162)	(0.0079)	(0.0079)	(0.0073)	(0.0073)
Air		0.1058	0.2765	0.1111	0.2074	0.1018	0.2335
		(0.0117)	(0.0117)	(0.0085)	(0.0085)	(0.0072)	(0.0072)
Panel B: Select Individual Pollutants							
COD	Not to CETP	0.1848	0.7176	0.1976	0.5999	0.1870	0.6617
		(0.0181)	(0.0181)	(0.0134)	(0.0134)	(0.0114)	(0.0114)
	CETP	-0.0596	0.1061	0.0092	0.2402	0.0023	0.2019
		(0.0643)	(0.0643)	(0.0315)	(0.0315)	(0.0286)	(0.0286)
TSS	Not to CETP	0.1522	0.4939	0.1637	0.4206	0.1445	0.4477
		(0.0297)	(0.0297)	(0.0242)	(0.0242)	(0.0191)	(0.0191)
	CETP	0.0164	0.4388	0.1276	0.4728	0.0821	0.4639
		(0.0352)	(0.0352)	(0.0148)	(0.0148)	(0.0144)	(0.0144)
TDS		0.0372	0.1103	0.0369	0.1926	0.0356	0.1540
		(0.0347)	(0.0347)	(0.0210)	(0.0210)	(0.0177)	(0.0177)
SPM		0.1864	0.5573	0.1920	0.4692	0.1759	0.4981
		(0.0292)	(0.0292)	(0.0211)	(0.0211)	(0.0169)	(0.0169)
SO_2		0.1394	0.2809	0.1387	0.1622	0.1332	0.2121
		(0.0187)	(0.0187)	(0.0134)	(0.0134)	(0.0113)	(0.0113)

Final-stage effluent and boiler-stack air samples, all audit reports. Standard errors in parentheses calculated by bootstrapping pollutant samples with replacement at the plant-by-year-by-season level from the distribution of backcheck readings. Audit treatment and control distributions are the true distributions of reports and so do not contribute to sampling error.

Table 4: Audit Treatment Effects on Auditor Reporting

	(1) All	(2) All	(3) Water	(4) Water	(5) Air	(6) Air
Panel A: Audit Report Levels on Treatment Status						
Audit treatment	0.103***	0.131***	0.117**	0.131**	0.0852***	0.132***
	(0.0354)	(0.0384)	(0.0527)	(0.0593)	(0.0214)	(0.0216)
Auditor fixed effects	No	Yes	No	Yes	No	Yes
Control mean	-0.291	-0.291	-0.350	-0.350	-0.194	-0.194
Observations	13172	13172	8373	8373	4799	4799
Panel B: Backcheck - Audit Differences on Treatment Status						
Audit treatment	-0.210***	-0.153	-0.152	-0.156	-0.312***	-0.166*
	(0.0726)	(0.0986)	(0.102)	(0.138)	(0.0825)	(0.0951)
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 and year fixed effects. Standard errors clustered at the plant level in parentheses. 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 $Air = \{SO_2, NO_x, \text{SPM}\}$ with $All = Water \cup Air$. Panel A sample of audit reports that reached the regulator over the two years of the experiment. Panel B midline survey sample. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Accuracy of Audit Reports in Treatment

	(1) All	(2) Water	(3) Air
		Panel A: 2009	
Mean backcheck-audit difference	0.379***	0.513***	0.124
	(0.0601)	(0.0682)	(0.1000)
		Panel B: 2010, ex-midline survey	
Mean backcheck-audit difference	0.278***	0.260***	0.310***
	(0.0632)	(0.0725)	(0.107)
		Panel C: 2010, midline survey	
Mean backcheck-audit difference	0.0779	0.176**	-0.0857
	(0.0567)	(0.0774)	(0.0663)

Final-stage effluent and boiler-stack air samples. Backcheck and audit pollutants matched by plant, time and sampling location for all backchecks conducted in the treatment group. Backchecks conducted in a 20% random sample of audit visits in each season except 2010, season 3, during which backchecks were conducted in the midline sample of plants. Standard errors clustered at the plant level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Audit Reports on Treatment Status Over Time

	(1) All	(2) All	(3) All	(4) All	(5) All
Audit treatment	0.103***	0.0789**	0.0849**	0.218***	0.180***
	(0.0354)	(0.0364)	(0.0370)	(0.0627)	(0.0568)
Incentive pay (year=2010)	0.0561**	0.0307	0.100**	0.00187	-0.00610
	(0.0273)	(0.0195)	(0.0440)	(0.0325)	(0.0325)
Incentive pay X treatment		0.0474	0.0446	0.257***	0.261***
		(0.0441)	(0.0436)	(0.0920)	(0.0976)
Years (fractional) from Jan 1, 2009			-0.0703**	0.0290	-0.00168
			(0.0349)	(0.0226)	(0.0192)
Years (fractional) X treatment				-0.220***	-0.169***
				(0.0677)	(0.0631)
Auditor fixed effects	No	No	No	No	Yes
Observations	13172	13172	13172	13172	13172

Regressions include region fixed effects. Standard errors clustered at the plant level in parentheses. 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 $All = Water \cup Air$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Backcheck - Audit Differences on Backcheck Experience

	(1)	(2)	(3)	(4)
Audit treatment	-0.250***	-0.222	-0.255***	-0.327**
	(0.0909)	(0.144)	(0.0901)	(0.149)
Any backchecks on auditor before audit	-0.170*	-0.193		
	(0.0997)	(0.219)		
Any backchecks on auditor before audit X Audit treatment	0.0944	0.0628		
	(0.152)	(0.228)		
Any treatment backchecks on auditor before audit			-0.176*	-0.480**
			(0.101)	(0.216)
Any treatment backchecks on auditor before audit X Audit treatment			0.103	0.221
			(0.151)	(0.221)
Any control backchecks on auditor before audit			-0.103	-0.248
			(0.0917)	(0.319)
Auditor fixed effects	No	Yes	No	Yes
Observations	1248	1248	1248	1248

Regressions include region fixed effects and week fixed effects. Standard errors clustered at the plant level in parentheses. 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 $All = Water \cup Air$. Midline survey sample. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Pooled Endline Pollutant Concentrations on Treatment Status

	(1) All	(2) All	(3) Water	(4) Water	(5) Air	(6) Air
Audit treatment	-0.211**	-0.456*	-0.300*	-0.982*	-0.0528	0.0237
	(0.0993)	(0.266)	(0.159)	(0.528)	(0.0566)	(0.101)
2009 Auditor fixed effects	No	Yes	No	Yes	No	Yes
2010 Auditor fixed effects	No	Yes	No	Yes	No	Yes
Observations	1439	1439	860	860	579	579

Regressions include region fixed effects. Standard errors clustered at the plant level in parentheses. 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 $All = Water \cup Air$. Endline survey data from audit sample in control group of treatment that cross-cut audit treatment. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

X Data Appendix

Table 9: Description of Pollutants

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.
Ammonia 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 particules and liquid droplets with a number of components, including acids, organic chemicals, metals and soil or dust (U.S. Environmental Protection Agency, 2010). Particulat 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., 2010).

XI Model Appendix

The following is a subgame-perfect Nash equilibrium with regulation:

1. Firms abate if and only if $c \leq min\{f, qV_{hl} - \Delta V\}$.

2. Firms offer payments

$$w(\hat{p}|p) = \begin{cases} a - \Delta V + qV_{lh} + \varepsilon & qV_{hl} - \Delta V \leq f, \hat{p} = p_l, p = p_h \\ a & otherwise \end{cases}.$$

3. Auditor submits report

$$\hat{p} = \begin{cases} p_h & p = p_l, \quad w_{ll} < w_{hl} - qV_0 & \cup \quad w_{hl} \geq a \\ p_l & p = p_l, \quad w_{ll} \geq w_{hl} - qV_0 & \cup \quad w_{ll} \geq a \\ p_h & p = p_h, \quad w_{lh} < w_{hh} - \Delta V + qV_{lh} & \cup \quad w_{hh} \geq a \\ p_l & p = p_h, \quad w_{lh} \geq w_{hh} - \Delta V + qV_{lh} & \cup \quad w_{lh} \geq a - \Delta V + qV_{lh} \end{cases}.$$

   If none of these conditions are met, the auditor declines to submit a report.

4. Regulator imposes a fine f if $\hat{p} = p_h$ is observed and disbars auditors observed to have reported falsely.

The definition of an equilibrium with regulation is one in which the regulator imposes a fine f if $\hat{p} = p_h$ is observed and disbars auditors observed to have reported falsely. This equilibrium is unique in this set up to the off-equilibrium path payments of the firm to the auditor and therefore all other equilibria with regulation are payoff equilvalent.

Consider the auditor's actions. In the subgame with $p = p_l$, the auditor receives a higher payoff from (truly) reporting $\hat{p} = p_l$ if $w_{ll} - a + V_0 \geq w_{hl} - a + (1-q)V_0$ and is willing to do so if $w_{ll} - a + V_0 \geq V_0$. These conditions are equivalent to $w_{ll} \geq w_{hl} - qV_0, w_{hl} \geq a$ given that $\Delta V \equiv V_{lh} - V_0$. In the subgame with $p = p_h$. The auditor receives a higher payoff from (falsely) reporting $\hat{p} = p_l$ if $w_{lh} - a + (1-q)V_{lh} \geq w_{hh} - a + V_0$ and is willing to do so if $w_{lh} - a + (1-q)V_{lh} > V_0$. These conditions are equivalent to $w_{lh} \geq w_{hh} - \Delta V + qV_{lh}$ and $w_{lh} \geq a - \Delta V + qV_{lh}$.

Consider the firm's actions in the subgame where it has chosen $p = p_l$. Inducing low reporting requires $w_{ll} - a + V_0 \leq w_{hl} - a + (1-q)V_0$ or $w_{ll} \leq w_{hl} - qV_0$. Inducing a high

report is optimal if $w_{ll} \geq w_{hl} + f$ but doing so requires a $w_{ll} \leq w_{hl} - qV_0$, a contradiction. The firm thus never induces a high report. For a low report it offers payment at cost. The payment for a high report can be any such that $w_{ll} = a > w_{hl} - qV_0 \to w_{hl} < a + qV_0$; we let $w_{hl} = a$ above.

Consider the firm's actions in the subgame where it has chosen $p = p_h$. Given the strategies of the auditor and regulators, inducing low reporting requires $w_{lh} - a + (1-q)V_{lh} \geq w_{hh} - a + V_0, w_{lh} - a + (1-q)V_{lh} \geq w_{hh} - a + V_0$. Inducing false reporting is profitable if $w_{lh} \leq w_{hh} + f$ and is possible if $w_{lh} \geq w_{hh} - \Delta V + qV_{lh}$. The firm will not pay any more than necessary, so inducing false reporting is worthwhile if $w_{hh} - \Delta V + qV_{lh} \leq f + w_{hh} \Leftrightarrow qV_{lh} - \Delta V \leq f$ which is the condition given. The firm will set $w_{lh} \geq a - \Delta V + qV_{lh}$, which is the lowest payment at which the auditor will submit. If the firm is not inducing false reporting it sets $w_{hh} = a$ to induce submission and $w_{lh} \leq a$.

Now consider the firm's choice of whether to abate. The outcome of the subgame without abatement is that total firm costs are $min\{f, qV_{lh} - \Delta V\}$. In the subgame with abatement, abatement c is the only cost. Abateing is therefore rational if $c \geq min\{f, qV_{lh} - \Delta V\}$ as given.