Econometrica Supplementary Material

SUPPLEMENT TO "THE VALUE OF REGULATORY DISCRETION: ESTIMATES FROM ENVIRONMENTAL INSPECTIONS IN INDIA"

(*Econometrica*, Vol. 86, No. 6, November 2018, 2123–2160)

ESTHER DUFLO
Departmen of Economics, Massachusetts Institute of Technology

MICHAEL GREENSTONE
Departmen of Economics, University of Chicago

ROHINI PANDE
Kennedy School, Harvard University

NICHOLAS RYAN
Departmen of Economics, Yale

APPENDIX A: DATA AND SUPPLEMENTARY ANALYSIS

A.1. *Mapping Documents to Regulatory and Plant Actions*

SECTION 2.3 describes the regulatory documents we use to assign actions to the regulator and plants in our model of regulation. This Appendix gives more detail on the documents involved, the mapping of these documents to actions, and the linkage of these actions into chains of related actions.

Our approach is based on rules that the Gujarat Pollution Control Board follows as a regulatory agency. All regulatory actions that GPCB takes regarding regulated plants must be documented. These documents are of distinct types, are all dated, and are often cross-referenced; that is, a document taking a regulatory action will often explicitly cite a prior violation of pollution standards as grounds for that action. Plants are not, like the regulator, obligated to follow rules in their correspondence. However, plants often do respond in writing to regulatory orders, in particular when they want to document that they have made a costly abatement investment.

Table S.I and the discussion in Section 2.3 covered the basic types of documents and their mapping to actions. Many of the documents have a one-to-one correspondence to actions. For the regulator, the action inspect is always indicated by an inspection report, which is a particular type of regulatory document. The action punish is always indicated by a closure direction sent to plants (or the plants' utility). We group multiple documents of varying severity under the action warn, including simple citations noting violations of pollution standards and more severe warnings that threaten closure. These documents all show that the regulator has found a violation, and often threaten action if the plant does not remediate, but they have in common that they do not impose a direct cost on plants. For plants, the main action of interest is whether a plant chooses to comply by installing abatement equipment. This action is typically documented by an invoice or other record from an environmental consultant or vendor that installed the equipment.

Esther Duflo: eduflo@mit.edu
Michael Greenstone: mgreenst@uchicago.edu
Rohini Pande: rohini_pande@harvard.edu
Nicholas Ryan: nicholas.ryan@yale.edu

 https://doi.org/10.3982/ECTA12876

TABLE S.I

MAPPING OF PLAYER ACTIONS TO RAW DOCUMENTS

Action	Document	Description
		Panel A. Regulatory Actions
Inspect	Inspection report	Analysis of air and water samples; report on plant characteristics.
Warn	Letter to plant	Non-threatening letter ordering improvement in pollutant concentrations.
	Citation	Threatening letter demanding explanation for high pollution levels, missing permit to operate, or missing pollution abatement equipment.
	Closure notice	Notice that the plant will be ordered to close in 15 days if the plant does not take action to improve pollution.
Punish	Closure direction	Order the plant to close immediately.
	Utility notice	Notice that water or electricity has been disconnected.
Accept	Revocation of closure direction	Permission to start operation.
	None	No further regulatory action within 60 days.
		Panel B. Firm Actions
Comply	Equipment installed	Notice that pollution abatement equipment has been installed.
	Process installed	Notice that a process has been installed.
	Bond posted	Letter from bank to GPCB explaining that the firm has posted a guarantee against future misconduct.
Ignore	Letter to regulator	Letter of protest to GPCB. Challenges parameter readings and other directives.
	None	Implied by consecutive GPCB actions without plant response.

Regulatory interactions with plants occur in groups of related actions we call chains. Many actions are responses to earlier actions. For example, if the regulator inspects a plant and finds a violation, this may result in a later action of warn or punish, or in the plant possibly choosing to comply to avoid punishment. For the same plant, we link actions in a chain by using explicit references in the documents underlying each action and the dating of those documents, as follows.

Stage 1. *Link documents that reference one another.*

(i) Link documents if one document explicitly cites (by document number or date) an inspection or other earlier document.

(ii) If there is no exact match, link documents if a near match exists that differs by at most one digit.

Stage 2. *Link documents dated close to one another.*

(i) Keep all documents already linked in Stage 1 together (we may call these sub-chains).

(ii) Link remaining documents (or groups of documents) to those that happened soon after and are plausibly logically related, using the following rules for each pair of candidate documents.

(a) Documents to link if, within 30 days, the following actions hold:

- Any regulator action is followed by plant ignore.
- Any regulator action is followed by regulator inspect.
- Regulator punish (closure notice, closure direction, utility confirmation of action) is followed by plant ignore or comply (equipment installation, process installation, bank guarantee posted).

(b) Documents to link if, within 60 days, the following conditions hold:

- Regulator inspect is followed by warn (letter, show cause notice, closure notice, closure direction).
- Regulator inspect is followed by regulator accept.
- Plant comply is followed by regulator accept (revocation of closure direction).

Stage 3. *Impute missing actions to enforce structure of the plant's problem*.

(i) *Chains start with the action inspect.* If a chain does not start with an inspection, append earlier inspections that occurred within 1 month of chain start. If no such recent inspection occurred, truncate the chain before first inspection.

(ii) *Chains end with the action accept.* Impute regulatory action accept at the end of the chain if there are no further followup actions within 60 days.

(iii) *Chains alternate moves between the plant and regulatory machine.* Impute the plant move ignore between any consecutive GPCB actions.

In Stage 3, we impute actions to enforce an alternate-move structure to the plant's problem. The main assumption in imputing moves is that the plant had an opportunity to respond to all regulatory actions, even if it in fact did not respond. That is, we take the absence of a response as the action ignore. Similarly, we assume the regulator could have continued pursuing plants at the end of each chain, and that if it does not, this represents the action accept. This assumption is weak because our data are comprehensive and all regulatory actions against plants are documented, so that if the regulator did later punish or warn a plant, this action would have been observed.

Table S.II gives an example of the result of this linking process. GPCB initially inspects a plant and the plant does nothing. GPCB, presumably on reviewing the inspection report and pollution samples, then orders that the plant be closed. The plant responds, several weeks later, by installing abatement equipment. GPCB repeatedly inspects the plant over

TABLE S.II

SAMPLE CHAIN[a]

Round (1)	Player (2)	Action (3)	Document (4)	Date (5)
1	GPCB	Inspect	Inspection report	2008-09-05
2	Plant	Ignore		2008-09-05
3	GPCB	Punish	Closure direction	2009-01-12
4	Plant	Comply	Equipment installed	2009-01-28
5	GPCB	Inspect	Inspection report	2009-01-31
6	Plant	Ignore		2009-01-31
7	GPCB	Inspect	Inspection report	2009-02-04
8	Plant	Ignore		2009-02-04
9	GPCB	Punish	Closure direction	2009-05-22
10	Plant	Comply	Process installed	2009-05-30
11	GPCB	Inspect	Inspection report	2009-06-16
12	Plant	Comply	Process installed	2009-06-16
13	GPCB	Accept	Revocation of closure direction	2009-06-24

[a]The table displays a 13-round chain of interactions between GPCB and one plant during the experiment. Column 3 indicates the category of action, while column 4 reports the underlying document to which the action corresponds. Ignore actions by the plant in rounds 2, 6, and 8 have been imputed based on adjacent actions in the chain and, hence, column 4 is left blank in these rounds. All chains begin with a regulatory inspection, inspect. The players then alternate moves until the regulator decides to accept the plant's compliance for the time being, which terminates the chain. Table 1 in the main paper describes the way in which the actions are mapped to the underlying documents, and the Data Appendix provides a full explanation of the rules used to construct the chains.

the next several months, apparently finds the remedy inadequate, and again orders the plant closed. The plant responds by modifying its production process. GPCB is then satisfied and revokes the closure order. This example gives a sense of the rich back-and-forth that is possible to observe in the chained action data.

A.2. *Abatement Costs*

Measures of plant abatement cost come from the end-line survey. The survey asked plants to describe each piece of pollution control "equipment installed or upgrades made (including routine maintenance such as filter changes, etc.)" since the start of the experiment, to record the corresponding expenditure and to verify the equipment or upgrade took place.

We flag these abatement expenditures as either capital or maintenance expenditures and sum them to the plant level. The determination of whether an expenditure is a capital or maintenance expenditure is based on string matching with the text description of each investment. Enumerators indicated maintenance using the words "maintenance" or "change," as in the action of changing a filter or other replaceable part. Expenditures with descriptions containing any of the following strings are therefore coded as maintenance: chang, mainten, maintain, maintan, (bag.*bag). The (bag.*bag) string captures variations on "change of bag filter bag," a common maintenance activity for air pollution control devices. Expenditures containing none of these strings are coded as capital expenditures.

For the purposes of Table II, panel A, which compares capital and maintenance costs, we amortize capital costs, which are reported as lump sums, into an equivalent annual expenditure. We calculate the constant annual expenditure such that the sum of the present value of the expenditure over the equipment lifespan equals the observed up-front capital expenditure (with an interest rate of $r = 0.20$ and a 10-year equipment lifespan).

A.3. *Experimental Integrity: Covariate Balance and Attrition*

This subsection verifies the integrity of the experiment both ex ante and ex post. First, we check the balance of covariates before the experiment started. Then we check for differential attrition during the experiment. Finally, we check our model assumption that the regulator acted similarly against treatment and control plants conditional on the results of an inspection.

Table S.III describes the cross-cutting experimental design. Plants were assigned to inspection treatment status conditional on audit treatment status (Duflo et al. (2013)). Since only certain plants are eligible for audits, there are three possible audit treatment statuses:

TABLE S.III

EXPERIMENTAL DESIGN: TREATMENT ASSIGNMENTS[a]

	Inspection Control	Inspection Treatment	Total
Audit control	120	120	240
Audit treatment	116	117	233
Not audit eligible	244	243	487
Total	480	480	960

[a]The table reports the number of plants assigned to each combination of the inspection treatment and the audit treatment of Duflo et al. (2013). Inspection treatment status is either control or treatment. With respect to audit, only some plants are audit eligible (see text). Conditional on being eligible for audit, plants are assigned to inspection treatment or control.

not audit eligible, and, conditional on being eligible, audit control or treatment. The inspection treatment status is therefore orthogonal to audit treatment status and eligibility.

Table S.IV compares the balance of the inspection treatment assignment on fixed plant characteristics in panel A and on regulatory interactions between plants and the regulator, such as inspections and violations of pollution standards, in panel B. The table uses

TABLE S.IV

INSPECTION TREATMENT COVARIATE BALANCE[a]

	Control (1)	Treatment (2)	Difference (3)
Panel A. Plant Characteristics			
Capital investment (rupees) 50–100m (=1)	0.087	0.071	−0.017
	[0.28]	[0.26]	(0.017)
Located in industrial estate (=1)	0.33	0.37	0.032
	[0.47]	[0.48]	(0.027)
Textiles (=1)	0.45	0.45	−0.0092
	[0.50]	[0.50]	(0.020)
Dyes and intermediates (=1)	0.13	0.16	0.027
	[0.34]	[0.36]	(0.022)
Effluent to common treatment (=1)	0.37	0.35	−0.021
	[0.48]	[0.48]	(0.031)
Waste water generated (kl/day)	192.1	196.8	4.30
	[310.9]	[316.4]	(16.2)
Air emissions from boiler (=1)	0.50	0.52	0.019
	[0.50]	[0.50]	(0.020)
Panel B. Regulatory Interactions in Year Prior to Study			
Number of inspections	1.22	1.25	0.026
	[1.32]	[1.32]	(0.079)
Inspections below prescribed (=1)	0.42	0.39	−0.031
	[0.49]	[0.49]	(0.029)
Number of pollution readings	3.64	3.92	0.28
	[5.65]	[5.58]	(0.31)
Pollution reading ever collected (=1)	0.40	0.44	0.048*
	[0.49]	[0.50]	(0.027)
Any pollution reading above limit (=1)	0.34	0.38	0.031
	[0.48]	[0.48]	(0.026)
Citations	0.22	0.20	−0.023
	[0.51]	[0.55]	(0.034)
Closure warnings	0.056	0.052	−0.0044
	[0.31]	[0.32]	(0.020)
Closure directions	0.075	0.077	0.0019
	[0.31]	[0.34]	(0.021)
Bank guarantees posted	0.019	0.029	0.010
	[0.15]	[0.21]	(0.012)
Equipment mandates	0.24	0.25	0.0082
	[0.54]	[0.53]	(0.029)
Any utility disconnection (=1)	0.010	0.0021	−0.0083
	[0.10]	[0.046]	(0.0051)
Observations	480	480	

[a]The table tests for the balance of covariates by inspection treatment status using administrative data from the regulator covering the year prior to the experiment. Columns 1 and 2 show means; standard deviations are given in brackets. Column 3 shows the coefficient on treatment from regressions of each characteristic on treatment and region fixed effects. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

administrative data and the regulatory interactions are measured over the last full calendar year (2008) prior to the experiment starting in August 2009. Each row considers a separate plant characteristic: columns 1 and 2 report the means for control and treatment plants, respectively, using administrative data for the year prior to the experiment. Column 3 reports the coefficient α_2 on the inspection treatment dummy T_j from the following regression, where for each outcome Y_j for plant j in region r,

$$Y_{jr} = \alpha_r + \alpha_1 \text{AuditSample}_j + \alpha_2 T_j + \epsilon_j, \tag{7}$$

where α_r are region effects and AuditSample_j is a dummy for a plant belonging to the audit sample (i.e., being audit eligible, rather than being assigned to the audit treatment).[25]

We find that inspections, pollution readings, and citations are balanced by treatment assignment. Of 18 baseline measures reported, there is a significant difference between the treatment and control groups at the 10% level on only one measure.

Table S.V reports overall levels of attrition in the experiment. About 18% of plants did not complete complete the end-line survey. Most of this attrition, 13 percentage points, was due to plants that closed during the experiment. Table S.VI compares attrition across the treatment and the control groups. One may be concerned that the extra scrutiny of treatment plants would drive them out of business. We find, to the contrary, that attrition is not differential across treatment arms (the point estimate for the effect of treatment on attrition is negative).

In the model, we assume that the regulatory machine acts similarly against both treatment and control plants (Section 5). This assumption is based on the design of the experiment, since the officials who decide to act against plants were not informed of whether an inspection report came from a treatment or a control plant. Here we present additional statistical evidence supporting that this assumption held in practice.

Table S.VII regresses a dummy for regulatory machine acceptance (i.e., leaving the plant alone) in a given round on the observable characteristics noted during an inspection, treatment status, and interactions of treatment status and observables. Columns 1–6 add progressively richer specifications of observables and interactions. In column 1, there

TABLE S.V

ATTRITION IN THE END-LINE SURVEY[a]

	N (1)	% (2)
Survey completed	791	82.4
Plant closed	124	12.9
Plant refused survey	5	0.5
Other	40	4.2
Total	960	100.0

[a]The table shows how many plants completed the end-line survey, and the reasons for attrition for those that did not. "Plant closed" includes plants that were permanently closed (111), plants that were temporarily closed, and plants where production was temporarily suspended. "Refused survey" includes plants that were operating at the time of the visit, but that refused to respond to the questions in the survey. "Other" includes plants that moved to an unknown address and plants for which an incorrect address had been recorded.

[25]Since only one region, Ahmedabad, contains both audit-eligible and -ineligible plants, this specification is equivalent to a full set of region × eligible effects.

TABLE S.VI

END-LINE ATTRITION BY INSPECTION TREATMENT STATUS[a]

	Treatment (1)	Control (2)	Difference (3)
Survey completed	0.840	0.808	0.031
	[0.367]	[0.394]	(0.024)
Plant closed	0.123	0.135	−0.013
	[0.329]	[0.343]	(0.022)
Plant refused survey	0.008	0.002	0.006
	[0.091]	[0.046]	(0.005)
Other	0.029	0.054	−0.025*
	[0.168]	[0.227]	(0.013)
Observations	480	480	

[a]The table shows differences in end-line responses and reasons for attrition between the treatment and the control groups. Columns 1 and 2 show means; standard deviations are given in brackets. Column 3 shows the coefficient on treatment from regressions of each characteristic on inspection treatment assignment, region fixed effects, and audit sample control. Reported are treatment effects, with region controls. "Plant closed" includes plants where production was temporarily suspended and plants that were temporarily or permanently closed. "Refused survey" includes plants that were in production at the time of the visit, but that refused to respond to the questions in the survey. "Other" includes plants that moved to an unknown address and plants for which an incorrect address had been recorded. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

are no controls. Column 2 controls for round of the game and its interactions with treatment, column 3 controls for pollution, column 4 controls for pollution and its interactions with treatment, and columns 5 and 6 additionally control for past regulatory and plant actions. In each column, we report at the bottom the p-value from an F-test that the coefficients on treatment and all treatment interactions with observables are jointly zero. If the regulator followed up with treatment plants differently, we would expect the treatment main effect or the interaction of the treatment and some observable to be significant (for example, if the regulator did not pursue treatment plants after finding high pollution readings). We fail to reject the joint null with p-values between approximately 0.20 and 0.60. We conclude that the regulatory machine treats plants similarly conditional on the facts observed in an inspection.

A.4. *Letter Treatment*

Table S.VIII reports the results of a treatment shortly before the end-line survey that sent plants a letter reminding them of their obligations to meet emissions limits. The letter had no significant effect on pollution or compliance.

A.5. *Compliance Placebo Checks*

Table S.IX reports the results of placebo regressions showing treatment effects on compliance at various compliance thresholds. Column 1 shows the regression of compliance on treatment, where compliance is defined as a pollution reading below the true standard $\overline{p}$, and columns 2–4 show the same regression with placebo standards set at several multiples of the true level. There is a significant effect, at the 10% level, only at the true threshold (column 1).

TABLE S.VII

PROBABILITY OF REGULATOR ACCEPTANCE ON TREATMENT AND ROUND[a]

	(1)	(2)	(3)	(4)	(5)	(6)
Inspection treatment assigned (=1)	0.0129	0.00779	0.0127	0.0164	0.0106	−0.0280
	(0.0112)	(0.00926)	(0.0110)	(0.0137)	(0.00835)	(0.0207)
Constant	0.827***	0.869***	0.863***	0.861***	0.891***	0.910***
	(0.00831)	(0.00701)	(0.00894)	(0.00974)	(0.00700)	(0.0134)
Period						
$t > 3$		−0.219***			−0.325***	−0.323***
		(0.0252)			(0.0350)	(0.0557)
$t > 5$		−0.0367			0.0293	0.0548
		(0.0479)			(0.0366)	(0.0610)
$t > 7$		−0.0571			−0.0224	−0.125*
		(0.0634)			(0.0396)	(0.0682)
Period × treatment						
$t > 3\times$ inspection treatment		0.0219				0.0322
		(0.0344)				(0.0464)
$t > 5\times$ inspection treatment		−0.0610				−0.0248
		(0.0638)				(0.0790)
$t > 7\times$ inspection treatment		0.0757				0.0920
		(0.0821)				(0.0961)
Lagged regulatory actions						
Warn, lag 1					0.169***	0.150***
					(0.0364)	(0.0477)
Punish, lag 1					−0.129***	−0.101*
					(0.0413)	(0.0535)
Lagged plant actions						
Firm: protest, lag 1					0.0482*	0.0535
					(0.0272)	(0.0363)
Firm: comply, lag 1					0.248***	0.259***
					(0.0329)	(0.0416)
Last pollution reading						
0–1x			−0.0102	0.0140	0.000305	
			(0.0158)	(0.0211)	(0.0119)	
1–2x			−0.0661***	−0.0534***	−0.0524***	−0.0661***
			(0.0134)	(0.0207)	(0.0105)	(0.0205)
2–5x			−0.109***	−0.120***	−0.0896***	−0.119***
			(0.0151)	(0.0251)	(0.0120)	(0.0216)
>5x			−0.160***	−0.178***	−0.116***	−0.144***
			(0.0260)	(0.0375)	(0.0189)	(0.0308)
Poll. reading × treatment						
0–1$x\times$ inspection treatment				−0.0424		
				(0.0309)		
1–2$x\times$ inspection treatment				−0.0221		0.0239
				(0.0271)		(0.0280)
2–5$x\times$ inspection treatment				0.0186		0.0497*
				(0.0313)		(0.0291)
>5$x\times$ inspection treatment				0.0330		0.0452
				(0.0516)		(0.0428)
p-value for F-test of null: All × inspection treatment terms = 0	0.249	0.608	0.245	0.388	0.204	0.315
Inspection control mean	0.827	0.827	0.827	0.827	0.827	0.827
Observations	8897	8897	8897	8897	8897	4089

[a]Does not include region fixed effects. Omitted actions are ignore (for regulator) and inspect (for plant). The omitted pollution reading for column (6) is "No pollution reading taken." Standard errors clustered at plant level are given in parentheses. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

TABLE S.VIII
END-LINE POLLUTION AND COMPLIANCE ON LETTER TREATMENT[a]

	Pollution (1)	Compliance (2)
Inspection treatment assigned (=1)	−0.0160	0.0248
	(0.0866)	(0.0238)
Letter treatment assigned (=1)	−0.0482	0.0311
	(0.0928)	(0.0241)
Inspection treatment × letter treatment (=1)	0.0326	−0.00340
	(0.130)	(0.0345)
Inspection and letter control mean	0.652	0.595
Observations	4168	4168

[a]The table shows regressions of pollution (column 1) and compliance (column 2) on inspection and letter treatment assignments. Observations are at the plant-by-pollutant level, where pollution consists of air and water pollution readings for each plant, taken during the end-line survey, and each pollutant is standardized by dividing by its standard deviation. Compliance is a dummy for each pollutant being below its regulatory standard. The table regresses these outcomes on inspection treatment assignment, letter treatment assignment, and inspection × letter treatment. The letter treatment was a letter sent by the regulator to plants shortly before the end-line survey that reiterated the terms of plants' environmental consents and reminded them of their obligations to meet emissions limits. Specifications also include region fixed effects. Standard errors clustered at the plant level are given in parentheses. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

A.6. *Regulatory Targeting Using End-Line Survey Data*

Table S.X presents additional evidence on regulatory targeting, plant-level regression estimates for how inspections depend on a plant's underlying pollution. To get at targeting under the status quo, we restrict the regression to the control group. To measure the regulator's response to latent pollution levels, we run the regression in the year after the experiment ended, using our own end-line survey readings, which were not reported to the regulator, as the measure of pollution. Columns 1–4 use a categorical dummy as the independent variable, where missing readings are coded zero (and indicated by a separate dummy variable), readings beneath the standard $\overline{p}$ are coded 1, between $(\overline{p}, 2\overline{p}]$

TABLE S.IX
PLACEBO CHECK OF ALTERNATE COMPLIANCE THRESHOLDS[a]

	$[0, \overline{p})$ (1)	$[0, 2\overline{p})$ (2)	$[0, 5\overline{p})$ (3)	$[0, 10\overline{p})$ (4)
Inspection treatment assigned (=1)	0.0366*	0.0144	0.00323	−0.000368
	(0.0213)	(0.0193)	(0.0131)	(0.00824)
Audit treatment assigned (=1)	0.0288	0.0154	0.0123	0.0166*
	(0.0258)	(0.0238)	(0.0162)	(0.00917)
Audit × inspection treatment (=1)	−0.0365	−0.0245	−0.0109	−0.0106
	(0.0353)	(0.0316)	(0.0214)	(0.0116)
Inspection and audit control mean	0.614	0.813	0.928	0.975
Observations	4168	4168	4168	4168

[a]The table presents regression estimates for compliance on inspection treatment assignment and audit treatment assignment, using pollution levels taken from the end-line survey. Compliance is defined as pollution below N times the limit $\overline{p}$, with $N = 1, 2, 3, 5, 10$. Pollution is standardized by back check standard deviation. Standard errors are given in parentheses. Audit treatment and treatment interaction controls, and year and region fixed effects are included. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

TABLE S.X

REGULATORY TARGETING ON UNOBSERVED POLLUTION IN THE CONTROL GROUP[a]

	Dependent Variable: Number of Inspections in 1 Year After End-Line Survey				
	(1)	(2)	(3)	(4)	(5)
End-line pollution bin (0–4)	0.170*	0.173*	0.182*	0.172*	
	(0.0978)	(0.103)	(0.101)	(0.103)	
End-line pollution $\in [1\overline{p}, 2\overline{p})$ (=1)					0.459
					(0.278)
End-line pollution $\in [2\overline{p}, 5\overline{p})$ (=1)					0.539
					(0.349)
End-line pollution $\geq 5\overline{p}$ (=1)					0.665**
					(0.331)
Constant	2.058***	1.995***	1.898***	1.893***	1.943***
	(0.302)	(0.463)	(0.461)	(0.532)	(0.486)
Plant characteristics	Yes	Yes	Yes	Yes	Yes
Audit treatment assignment		Yes	Yes	Yes	Yes
Recent regulatory actions			Yes	Yes	Yes
Recent pollution readings				Yes	Yes
Mean dependent variable	1.392	1.392	1.392	1.392	1.392
F-stat. p-value					0.00859
R^2	0.213	0.213	0.259	0.284	0.285
Observations	388	388	388	388	388

[a]The table regresses the number of regulatory inspections in the year after the end-line survey on pollution readings as measured during the survey, in the inspection control group of plants. As end-line pollution readings were not reported to the regulator, these readings represent an unobserved component of pollution, from the regulator's perspective. The end-line pollution bin is a categorical variable that takes the value of 0 for plants with no pollution readings, 1 if pollution is in $[0, \overline{p})$, 2 if in $[\overline{p}, 2\overline{p})$, 3 if in $[2\overline{p}, 5\overline{p})$, and 4 if above $5\overline{p}$. All specifications also include a dummy for a plant having no pollution reading. Column 5 separates the pollution bin into a set of dummy variables and reports the coefficient on each; the omitted dummy is having end-line pollution bin equal to 1 (i.e., pollution beneath the standard). The F-test reported is for the joint significance of the end-line pollution bin dummies. Plant characteristics include dummies for size, use of coal or lignite as fuel, high waste water generated, dye sector, textile sector, and region. Recent regulatory actions include the number of regulatory actions of several types against the plant in the year before the end-line. Recent pollution readings include dummies for pollution bins at the most recent regulatory inspection before the end-line. Robust standard errors are given in parentheses. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

are coded 2, between $(2\overline{p}, 5\overline{p}]$ are coded 3, and $5\overline{p}$ and above are coded 4; column 5 shows dummies for the underlying pollution categories.

Table S.X, column 1 shows that, conditional on plant characteristics, higher end-line survey pollution readings predict more inspections, even though the regulator does not observe these readings. Each higher category of pollution results in 0.170 more inspections per year (standard error 0.0978) relative to a compliant plant.[26] Regulatory targeting on unobserved pollution remains as strong when adding controls for audit treatment status during the experiment (column 2) and recent regulatory penalties (column 3). Perhaps most strikingly, when adding recent pollution readings that the regulator took as controls, we find the coefficient on end-line survey pollution readings remains unchanged (column 4). In column 5, we separate the components of the categorical end-line pollution variable and find that belonging to the highest pollution bin is associated with 0.665 additional inspections per year (standard error 0.331). The dummies for pollution bins above the standard are jointly significant in predicting later inspections ($F_{4,367} = 3.46$,

[26]Plant observable characteristics like belonging to a dirtier sector, being of greater scale, and generating more waste water also predict more inspections (coefficients not reported).

TABLE S.XI
PARAMETERS FOR TARGETING MODEL SIMULATION[a]

Description	Parameter (1)	Value (2)
Pollution equation	ϕ	$[0.5\ 0.5\ -0.1\ -1.5]'$
Inspection equation	β	$[0.3\ -0.1\ 0.25]'$
Maintenance cost	μ_c	2
	σ_c	1
Pollution shocks	σ_1	$\in \{0.31, 0.43, 0.61\}$
	σ_2	$\in \{0.73, 0.66, 0.50\}$
Targeting parameter	ρ	1

[a]The table gives the parameters for the Monte Carlo simulation of the targeting model. Pollution parameters are the coefficients on $[X_j$ run] in the pollution equation and inspection parameters are the coefficients on X_j. Other parameters are described in Section 4.

p-value < 0.01). The regulator appears to target inspections based on signals of plant pollution beyond its own past pollution readings.

APPENDIX B: SIMULATION OF TARGETING STAGE

This appendix describes the Monte Carlo simulation of optimal regulatory targeting.

B.1. *Setup*

We set $N = 900$ plants and endow plants with exogenous observables $X_j = [X_1\ \ X_2\ \ X_3]$, where $X_1, X_2, X_3 \in \{0, 1\}$, $\Pr(X_1 = 1) = 0.25$, and $\Pr(X_2 = 1) = \Pr(X_3 = 1) = 0.50$. We assume the parameter values in Table S.XI.

In each simulation, we take the following steps.

(i) Draw shocks for each plant (holding the draws fixed across simulations).

(ii) Construct latent levels of pollution if the plant does not abate.

(iii) Solve the regulator's targeting problem subject to a budget of $\overline{N} = 1.47$ inspections per plant.

The nine different simulations vary on the two dimensions of penalty function shape and regulatory information.

We use three different penalty functions that yield approximately the same average level of penalties but have different curvature. Figure S.1 shows these alternate functions. The horizontal axis in the figure is the pollution found at a plant on initial inspection, in units of multiples of the regulatory standard. The vertical axis is the expected discounted value of the penalty function, in thousands of U.S. dollars. All penalty functions intersect the same penalty values at pollution readings of $1\times$ and $7.5\times$ the regulatory standard. The three functions differ in their curvature. The dotted line is the penalty function as estimated in the penalty stage, interpolated between discrete values using a piecewise-cubic Hermite interpolating polynomial (pchip), a type of spline that preserves monotonicity between knots. The estimated penalty function shows sharply increasing marginal penalties at the regulatory standard and decreasing marginal penalties beyond that. The solid line is a linear penalty function (constant marginal penalty) and the dashed line is a quartic penalty function (increasing marginal penalty).

The second dimension on which the simulations vary is the information available to the regulator. We keep $\sigma_1^2 + \sigma_2^2 = 0.625$ across simulations and vary the fraction of the

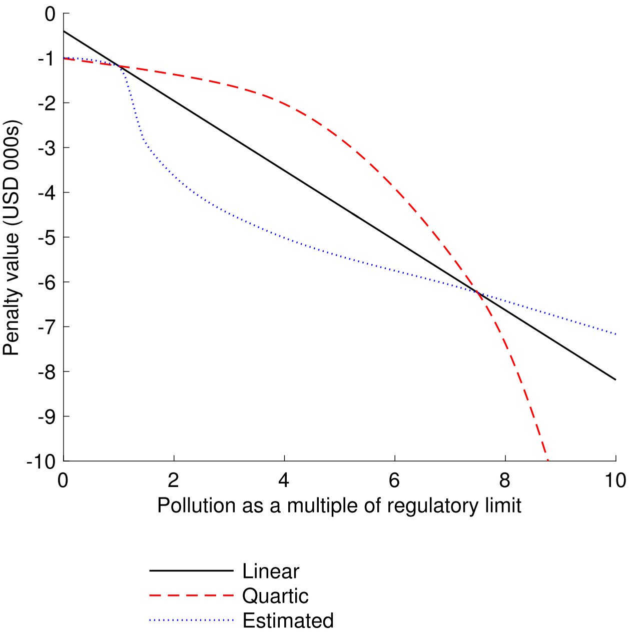

FIGURE S.1.—Alternate penalty function forms. The figure shows alternate functional forms for the penalty function, which gives the expected discounted value of the penalty stage, at the time of an initial inspection, as a function of plant pollution. The value is measured in thousands of U.S. dollars and pollution is measured as a multiple of the regulatory limit; that is, a plant with pollution equal to 2 has a reading double the limit. The solid line is a linear penalty function (constant marginal penalty), the dashed line is a quartic penalty function (increasing marginal penalty), and the dotted line is the penalty function as estimated in the penalty stage. The estimated penalty function is interpolated between discrete values using a piecewise-cubic hermitian interpolating polynomial (pchip), a type of spline that preserves monotonicity between knots. All penalty functions are set to intersect the estimated penalty function at pollution values of 1.0 (the standard) and 7.5 (roughly the limit of the pollution values observed).

variance in total pollution that is observable to the regulator, setting σ_1 such that $\sigma_1^2/(\sigma_1^2 + \sigma_2^2) \in \{0.15, 0.30, 0.60\}$. In total there are therefore nine different simulation runs with each combination of regulatory information and penalties.

B.2. *Simulation Results*

This section presents results from simulations of the targeting model to show how regulatory penalties and information shape optimal inspection targeting.

The optimal targeting rule will depend on several aspects of the model. First, since the plant's value of regulation from the penalty stage varies with pollution, the regulator can induce more abatement by allocating inspections to plants with high pollution and, therefore, high expected penalties. Second, plant reductions in pollution are proportional to the level of pollution, so allocating inspections to higher-polluting plants will have higher yield in abatement if those plants do abate. These forces both suggest allocating more inspections to plants with higher observable pollution shocks. Third, however, plants have idiosyncratic maintenance costs, and if inspections are very concentrated on the plants with the highest observable pollution shocks, the regulator may miss chances to induce lower-cost plants to run their equipment. The relative strength of these forces—how much the regulator should go after the plants it thinks are the dirtiest—will depend

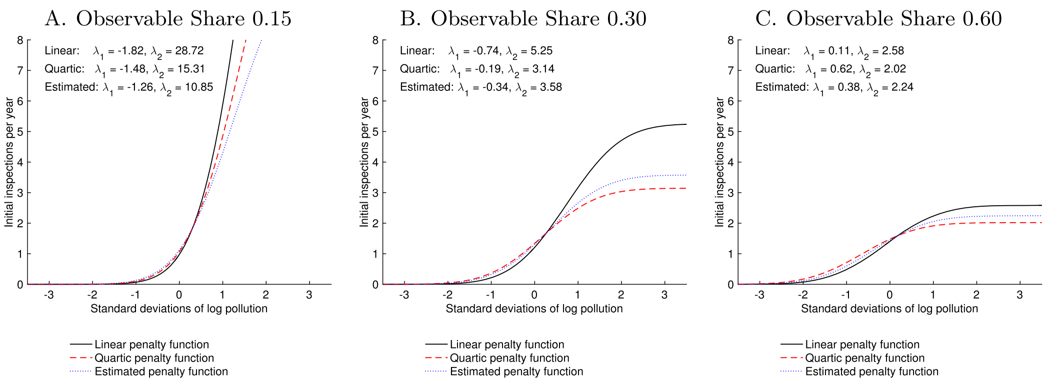

FIGURE S.2.—Optimal targeting rules as observable share of pollution varies. The figure compares the optimal inspection targeting function in Monte Carlo simulations of the targeting stage of the model that vary (i) the share of variation in pollution that is observable by the regulator (across panels) and (ii) the shape of the penalty function (across curves within a panel). The share of variation observable to the regulator is equal to $\sigma_1^2/(\sigma_1^2+\sigma_2^2) \in \{0.15, 0.30, 0.60\}$ across panels. Each curve within a panel shows the optimal targeting rule, giving inspections as a function of the standard deviations of observed pollution, that solves the regulator's problem of minimizing pollution subject to an inspection budget, in a Monte Carlo simulation of the model. The curves within a panel differ in the shape of the regulatory penalty function. The alternate shapes of the penalty function are shown in Figure S.1 and explained in the notes for that table.

on the accuracy of the regulator's signal of pollution and the shape of the penalty function with respect to measured pollution.

Figure S.2 shows the optimal targeting functions—the key endogenous object of interest—from each of these nine simulations. The three panels vary in the share of pollution observable to the regulator, and the three lines within each panel vary in the shape of the penalty function. We plot all the optimal targeting functions for a hypothetical plant for which the argument of the targeting function $X_j'\beta = 0$, aside from the contribution of the observed pollution shocks.

Consider first the variation in targeting functions induced by the regulator's information. The solid line in all figures represents the optimal targeting function, of the form (5), given a linear value function of the penalty stage (constant marginal penalties for higher initial pollution). In Figure S.2(A), when the share of variation in pollution observed by the regulator is small, the optimal penalty function is steep. The maximum inspection parameter $\lambda_2 = 28.72$ indicates that a plant with an arbitrarily high u_{1j} shock would receive at most 28.72 inspections per year, and the shift parameter $\lambda_1 = -1.82$ implies that plants would have to draw extremely high shocks to receive these high inspections: a plant with a $3\sigma_1$ shock and $X_j'\beta = 0$ would receive only about four initial inspections per year. The regulator knows little about pollution and puts its eggs in one basket by aggressively going after the plants observed to be dirtiest. As the share of pollution observable to the regulator decreases (Figure S.2(B) and (C)), the optimal targeting function (solid curve) gets much flatter and shifts leftward. The regulator, more confident in its signal and, therefore, that a plant with moderately high observed pollution will run, spreads inspections around to pick up a broader set of plants that may have low abatement costs.

The variation in the shape of the penalty function is not as important as information for the optimal targeting rule, though it does change the rule somewhat. For example, if the regulator has very little information (Figure S.2(A)), the optimal targeting rule is somewhat more concentrated in high polluting firms under a linear penalty function then under the estimated penalty function, which has decreasing marginal penalties at high pollution levels. This difference makes sense: under a linear penalty function, inspecting plants that are highly polluting has an extra kick, since marginal penalties do not taper off at high levels of pollution. We see a similar ordering for higher levels of information in Figure S.2(B) and (C). Perhaps most subtly, the estimated penalty function (dashed line) with increasing, then decreasing marginal penalties, yields less concentrated inspections than the quartic penalty function under low information but more concentrated inspections under high information. This reversal happens because, under low information, the regulator is inspecting mainly the plants with the highest observable shocks, so the region of the penalty function at very high pollution is relatively more important. At this high range, the estimated penalty function has decreasing marginal penalties, unlike the quartic penalty function. Under high information, the regulator is inspecting a broader range of plants, so the region of the estimated penalty function near the standard—which has sharply increasing marginal penalties—comes into play, and it is better for the regulator to concentrate inspections slightly more.

The regulator's targeting problem captures the trade-offs involved in setting a targeting rule in an economically rich way. The parsimonious probit link form (5), governed by parameters λ, allows a range of interesting targeting rules, from those that are very steeply increasing to rules close to linear in pollution to rules that are nearly flat.

APPENDIX C: ESTIMATION

C.1. *Penalty Stage Estimation*

Here we provide further details of the preliminary estimation steps involved in forming the penalty stage likelihood.

C.1.1. *Actions and Plant Payoffs in a Penalty Stage Round*

In all even rounds, the plant may comply or ignore the regulatory machine. To comply, a plant must pay a constant $\pi_j(a_{jt} = \text{comply}|s_t) = -k$ to install abatement equipment. We assume all plants have costs for installing abatement capital equal to the average value of abatement capital costs observed in our sample, conditional on installation, which is $k =$ \$17,000. To ignore the regulator costs the plant nothing today, but may increase the risk of future regulatory action.

If the machine punishes, the plant payoff takes one of two forms depending on the specification. Most simply, we estimate specifications where plant penalties are a constant $h(p_{jt}) = -h$. We also estimate specifications where plant penalties depend on pollution,

$$\begin{aligned}\pi_j(a_{Rt} = \text{punish}|s_t) &= -h(p_{jt}) \\ &= -\big(\tau_0 \mathbf{1}\{\overline{p} < p_{jt} < 2\overline{p}\} + \tau_1 \mathbf{1}\{2\overline{p} \le p_{jt} < 5\overline{p}\} + \tau_2 \mathbf{1}\{5\overline{p} \le p_{jt}\}\big),\end{aligned}$$

where p_{jt} is the pollution reading and the legally mandated pollution threshold is $\overline{p}$. The payoff for punishment is the cost to the plant of temporary closure and any remediation.

If the machine inspects, the plant has payoff

$$\begin{aligned}\pi_j(a_{Rt} = \text{inspect}|s_t) &= -b(p_{jt}, a_{j-}) \\ &= -\big(1 - \mathbf{1}\{a_{j-} = \text{comply}\}\big) \\ &\quad \times \big(\nu_0 \mathbf{1}\{\overline{p} < p_{jt} < 2\overline{p}\} + \nu_1 \mathbf{1}\{2\overline{p} \le p_{jt} < 5\overline{p}\} + \nu_2 \mathbf{1}\{5\overline{p} \le p_{jt}\}\big).\end{aligned}$$

The subscripts in a_{j-} reference the prior action of the plant. If the regulator is to move at turn t, then a_{R-} will be the regulatory machine's prior action at $t-2$; if the plant is to move at t, then a_{R-} will have been taken at $t-1$. The payoff for inspection captures the possible disruption posed by inspections and any bribes plants give on being inspected. This function specifies that inspections are costless for plants that have recently complied, but for plants that have not recently complied, inspections incur costs that depend on pollution emissions. We also estimate specifications where inspections are costless for all plants.

Warnings are costless to the plant, but continue the stage and obligate the plant to respond. Finally, the machine may accept that the plant is compliant, which costs the plant nothing and ends the penalty stage.

C.1.2. *States and State Transitions*

We specify the common state of the game to comprise the pollution reading, the last two actions of the regulator and plant, and the penalty stage round:

$$s_t = \{p_{jt}, a_{j-}, a_{R-}, \mathbf{1}\{t > 2\}, \mathbf{1}\{t > 4\}, \mathbf{1}\{t > 6\}\}.$$

The pollution reading p_{jt} is the maximum reading for any pollutant observed in the most recent inspection; if no inspection occurs in round t, then pollution is recalled from the last inspection within the stage. We specify the round as entering with several dummies to allow regulatory actions to respond flexibly to the selection of plants that may occur across rounds.[27]

The state transition after the plant moves is wholly deterministic, because the plant affects only how its own action is recorded in the state: if it chooses to comply today, then $a_{j-} =$ comply tomorrow. The transition after the regulator moves has a deterministic part, for the machine's action, and a stochastic part, the current plant pollution reading. We use a simple count estimator for the pollution state transition when the machine moves:

$$\Pr(p' | p_{jt}, a_{Rt}) = \frac{\sum_{j,c,t} \mathbf{1}\{p_{j,t+1} = p' | p_{jt}, a_{Rt}\}}{\sum_{j,c,t} \mathbf{1}\{p_{jt}, a_{Rt}\}}.$$

The pollution state may transition only if $a_{Rt} =$ inspect. We restrict the pollution transition to depend on past pollution and the machine's move, but not the plant's past moves.[28] When the machine accepts, the penalty stage ends and the firm draws a new u_{2jm}.

We treat each chain of interactions between the plant and the regulatory machine as independent, that is, $u_{2j,m+1}$ is independent of u_{1j} and u_{2jm}. The regulator continues to observe part of pollution, u_{1j}, but conditional on this observation does not, for example, use past readings from the penalty stage to determine targeting.[29]

[27]In theory, the whole history of player actions could enter the state. We found that enriching the state in plausible directions, such as including further lags, did not help predict regulatory actions.

[28]The count estimator may be biased for low-probability events in finite samples, so that conditioning on more past actions will leave many cells empty (e.g., the probability of pollution transitioning from above 5 times the standard to between 1 and 2 times given that the plant complied and the regulator inspected). Nonetheless, we find the count estimator preferable to smooth alternatives, such as an ordered logit model, because it does not restrict state transition patterns.

[29]This assumption is for tractability but is also empirically reasonable. The average time between chains (about 5 months) is much larger than the average time between actions within a chain (2 weeks). Recent

C.1.3. *Regulatory Machine Conditional Choice Probabilities*

The plant knows the machine action probabilities in each possible future state. We use a multinomial logit model to estimate these conditional probabilities, where

$$\Pr(a_{Rt} = a|s_t) = \frac{\exp\bigl(q(s_t)'\omega_a\bigr)}{\sum_{a'} \exp\bigl(q(s_t)'\omega_a\bigr)},$$

ω_a is a vector of coefficients for each action, and $q(s_{it})$ is a vector of state values: dummies for the possible most recent actions, categorical bins for the observed pollution level p_{it}, and dummies for the stage of the game.

C.1.4. *Action-Specific Values*

We calculate action-specific values for the plant using backward induction. We assume the game is finite and that the regulatory machine will always accept in period $T = 35$, which is well beyond the ultimate round of $t = 19$ actually observed in the data.[30] We further assume the plant does not anticipate any change in future value from actions beyond the current penalty stage. In the plant's problem, which is then finite, we infer action-specific values using the state transitions and choice probabilities, starting at the final round.[31] We use a discount factor of $\delta = 0.991$ between rounds, which has been calibrated, given the average round duration, to match the annual returns on capital for Indian firms found by Banerjee and Duflo (2014).

C.2. *Targeting Stage Estimation*

Section 5.2 describes the moments used to estimate the targeting stage of the model. This Appendix derives the expressions for these moments in the model. Given the form of the inspection targeting function and the assumed distributions of cost and pollution shocks, most of the model moments have concise analytic forms.

- **Pollution equation**. We use T_j as an instrument for N_j in the pollution equation (1):

$$g_{1j}(\phi) = Z_j'\bigl(\log P_j - \phi_0 - X_j'\phi_1 - \text{run}_j\phi_2\bigr), \tag{8}$$

where $Z_j = [1\ X_j'\ T_j]'$.

---

pollution readings do not change regulatory targeting of inspections (Appendix Table S.X, column 4). Last, the regulator has a short memory in practice; of those actions that explicitly cite a prior inspection, 93% of the time the inspection cited is the most recent prior inspection.

[30]Given that the probability of regulatory machine acceptance in any given round acts like a discount factor, this assumption on the game length is conservative in that late rounds matter very little for plants' expected values.

[31]When $t = T$, then $v_j(a_{Rt}|s_t) = 0$. At $t = T - 1$, the plant's value equals its one-period profit plus an action-specific shock, $v_j(a_{jt}|s_t) = \pi_j(a_{jt}|s_t) + e_j(a_{jt}, s_t)$. The regulator always moves as estimated in the data. At moves $t = T - 3$ and all earlier moves of the plant, the plants action-specific value is found with the empirical analogue to equation (6), where the plant's profit in a given round depends on the parameters θ_{uP}. For the estimation, we restrict the sample to all plant actions taken in round $t = 4$ and after, omitting $t = 2$ on the grounds that we believe the plant often does not have a chance to respond to the regulator in $t = 2$ before another regulatory action is observed (see discussion in Section 2).

• **Inspection equation**. The targeting rule allows for an analytic expression of the expected number of inspections, given plant characteristics X_j. We calculate the expected number of initial inspections for a plant with certain observable characteristics as

$$\begin{aligned}\mathbb{E}\big[\mathcal{I}(u_{1j}|X_j, T_j, \lambda, \beta, \rho)\big] &= \int \phi(u_1/\sigma_1)\mathcal{I}(u_{1j}|X_j, T_j, \lambda, \beta, \rho)\, dF(U_1) \\ &= \int \phi(u_1/\sigma_1)\lambda_2\Phi\left(\frac{\lambda_1 + X_j'\beta_1 + T_j\beta_2 + u_1}{\rho}\right) dF(U_1) \\ &= \lambda_2\Phi\left(\frac{\lambda_1 + X_j'\beta_1 + T_j\beta_2}{\sqrt{\rho^2 + \sigma_1^2}}\right),\end{aligned}$$

using the distribution of u_1.[32] Therefore, for any X_j, T_j, and candidate parameters, we can calculate the expected number of initial inspections by integrating over the observable pollution shocks. With this expression, we form moments

$$g_{2j}(\beta) = Z_j'\big(\mathbb{E}\big[\mathcal{I}(u_{1j}|X_j, T_j, \lambda, \beta, \rho)\big] - I_j\big). \tag{9}$$

Similarly, we can calculate the expected value of squared inspections as

$$\begin{aligned}\mathbb{E}\big[\mathcal{I}^2(u_{1j}|X_j, T_j, \lambda, \beta, \rho)\big] &= \int \phi(u_1/\sigma_1)\mathcal{I}^2(u_{1j}|X_j, T_j, \lambda, \beta, \rho)(u_1|X_j, T_j; \beta)\, dF(U_1) \\ &= \lambda_1^2 \int \phi(x)\Phi^2\left(\frac{x-a}{b}\right) dx\end{aligned}$$

for $x = u_1/\sigma_1 \sim \mathcal{N}(0, 1)$, $a_j = [-(\lambda_1 + X_j'\beta_1 + T_j\beta_2)/\sigma_1]$, and $b = \sqrt{\rho^2 + \sigma_1^2}$. This integral can be represented as a bivariate normal cumulative distribution function $\mathcal{N}_2$. If we let

$$Z_1, Z_2 \sim \mathcal{N}\left(\mu = \begin{bmatrix} -a_j \\ -a_j \end{bmatrix}, \Sigma = \begin{bmatrix} \rho^2 + \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \rho^2 + \sigma_1^2 \end{bmatrix}\right)$$

and represent the joint cumulative distribution function (CDF) of these variables as $\Pr(Z_1 \leq z \cup Z_2 \leq z) = \mathcal{F}(z, \mu, \Sigma)$, then the expected value of inspections squared is

$$\mathbb{E}\big[\mathcal{I}^2(u_{1j}|X_j, T_j, \lambda, \beta, \rho)\big] = \lambda_2^2\mathcal{F}\big([0\ 0]', \mu, \Sigma\big).$$

We form moments as

$$g_{3j}(\beta) = Z_{3j}'\big(\mathbb{E}\big[\mathcal{I}^2(u_{1j}|X_j, T_j, \lambda, \beta, \rho)\big] - I_j^2\big), \tag{10}$$

where $Z_{3j} = [1\ T_j]'$.

[32]Let $X \sim \mathcal{N}(0, 1)$. Then

$$I(a, b) = \int \phi(x)\Phi\left(\frac{x-b}{a}\right) dx = 1 - \Phi\left(\frac{b}{\sqrt{a^2+1}}\right) = \Phi\left(\frac{-b}{\sqrt{a^2+1}}\right).$$

Recall that $u_1 \sim \mathcal{N}(0, \sigma_1)$, so $u_1/\sigma_1 \sim \mathcal{N}(0, 1)$.

- **Abatement cost moments**. We use the observed probabilities of run and expected abatement costs, conditional on running, to form

$$g_{4j}(\phi, \mu, \sigma) = \Pr(\text{Run} = 1|\phi, \mu, \sigma) - \mathbf{1}\{c_j > 0\}, \tag{11}$$

$$g_{5j}(\phi, \mu, \sigma) = \mathbb{E}[c_j|\text{Run}, \phi, \mu, \sigma] - \mathbf{1}\{c_j > 0\}c_j. \tag{12}$$

The probability of running and expected cost, conditional on running, are both functions of the distributions of c_j, u_1, and u_2. There is no convenient analytic form for the truncated distribution of c_j, from which to form moments, since this would be the expectation of a truncated sum of log-normal and normal components, passed through a nonlinear penalty function $V_0(\cdot)$. Therefore, we partly simulate the relevant probability and expectation.

Let $\overline{c}_j = I_j(V_0(P_j) - V_0(P_j^*))$ so that run $= \mathbf{1}\{c_j < \overline{c}_j\}$. The threshold $\overline{c}_j$ is a function of plant pollution shocks u_1, u_2 and model parameters (including parameters on targeting, to determine I_j). For candidate parameters and a draw of shocks, we can calculate $\overline{c}_j$ for each plant. Then the probability of run in a simulation s is

$$\begin{aligned}\Pr^s(\text{run} = 1|\phi, \mu, \sigma) &= \Pr\big(c_j < \overline{c}_j^s|\phi, \mu, \sigma\big) \\ &= \Phi\bigg(\frac{\log \overline{c}_j^s - \mu_c}{\sigma_c}\bigg).\end{aligned}$$

Across S simulation draws for each pollution shock, we calculate

$$\Pr(\text{run} = 1|\phi, \mu, \sigma) = \sum_{s=1,\dots,S} \Phi\bigg(\frac{\log \overline{c}_j^s - \mu_c}{\sigma_c}\bigg)/S.$$

We have used simulation over two dimensions of the shock, and used an analytic expression for the truncated moment over the third.

For the expected value of maintenance costs, conditional on maintenance, we use the moments of the truncated log-normal distribution.[33] We form the expected cost of maintenance, conditional on running equipment, as

$$\mathbb{E}[c_j|\text{run}, \phi, \mu, \sigma] = \exp\big(\mu_c + \sigma_c^2/2\big)\frac{\Phi(-\sigma_c + b_{0j})}{\Phi(b_{0j})},$$

where $b_0 = (\log \overline{c}_j^s - \mu_c)/\sigma_c$. Again, because the value $\overline{c}_j^s$ depends on the simulation draws, we take the expected value of c_j as the mean of this expression over simulation draws.

This simulation is somewhat complicated: plants only decide on whether to run abatement equipment based on their expectation of the regulator's targeting, and this targeting

[33] For truncation from above with $x \sim \mathcal{N}(\mu, \sigma)$ and $y = e^x$, these are

$$\mathbb{E}[y|y \leq b] = \exp\big(\mu + \sigma^2/2\big)\frac{\Phi(-\sigma + b_0)}{\Phi(b_0)},$$

$$\mathbb{E}\big[y^2|y \leq b\big] = \exp\big(2\mu + 2\sigma^2\big)\frac{\Phi(-2\sigma + b_0)}{\Phi(b_0)},$$

where $b_0 = (\log b - \mu)/\sigma$.

rule depends on parameters. Thus the simulation involves first solving the targeting rule for given parameters, then simulating $\overline{c}_j^s$ for each draw of pollution shocks, and finally calculating the probability of abatement and expected value of abatement over simulation draws.

- **Variance of pollution shocks**. We wish to estimate the components of σ for the standard deviation of the observed and unobserved pollution distributions.

From the pollution equation, the variance of log pollution is equal to the sum of the variances of the two independent shocks (recall, $\varepsilon_2 = u_1 + u_2$). Therefore,

$$\begin{aligned} g_{6j}(\beta, \phi, \sigma) &= \mathbb{E}\big[\varepsilon_2^2 | \beta, \phi, \sigma\big] - \hat{\varepsilon}_2^2 \\ &= \sigma_1^2 + \sigma_2^2 - \hat{\varepsilon}_2^2, \end{aligned}$$

where we can form our empirical estimate of the pollution residual as $\hat{\varepsilon}_2 = P_j - \hat{\phi}_0 - X_j'\hat{\phi}_1 - \text{run}_j\hat{\phi}_2$. This moment identifies the sum of the variances of the observed and unobserved shocks.

- **Covariance of pollution and inspection shocks**. Finally, we are interested to separate the effect of observed and unobserved pollution shocks. The key idea is that only observable pollution shocks result in more inspections. We form an additional moment

$$g_{7j}(\beta, \phi, \sigma) = \mathbb{E}[\varepsilon_2 \cdot \mathcal{I} | \theta] - \hat{e}_{2j} \times I_j.$$

This moment is related to the covariance of the pollution shock and the level of inspections. Intuitively, if the inspection decision is based on a part of the pollution error not observed by the econometrician, then the pollution residual and inspections will covary.

We derive a prediction for $\mathbb{E}[\varepsilon_2 \cdot \mathcal{I} | \theta]$ in the model,

$$\begin{aligned} \mathbb{E}[\varepsilon_2 \cdot \mathcal{I} | \theta] &= \mathbb{E}\big[(u_1 + u_2) \cdot \mathcal{I}(u_1 | \theta)\big] \\ &= \mathbb{E}\big[u_1 \cdot \mathcal{I}(u_1 | \theta)\big] + \mathbb{E}\big[u_2 \cdot \mathcal{I}(u_1 | \theta)\big] \\ &= \mathbb{E}\big[u_1 \cdot \mathcal{I}(u_1 | \theta)\big], \end{aligned}$$

where the third line follows because the pollution shock u_2 is unobserved by the regulator and cannot affect inspection decisions. This is the key idea for identifying targeting: the inspection targeting function depends only on the observable part of the shock.

We proceed by substituting in the targeting function and integrating by parts to yield the desired moment in the model:

$$\mathbb{E}[\varepsilon_2 \cdot \mathcal{I} | \theta] = \lambda_2 \frac{\sigma_1}{\rho} \phi\left(\frac{-\big(\lambda_1 + X_j'\beta_1 + T_j\beta_2\big)}{\sqrt{\rho^2 + \sigma_1^2}}\right).$$

This moment will depend on the observable characteristics of each plant. Because $\lambda_2 > 0$, $\sigma_1 > 0$, and $\rho > 0$, this correlation is expected to be positive: plants with high pollution shocks are expected to have higher inspections, due to regulatory targeting.

C.3. *First-Order Conditions of Regulator's Targeting Problem*

The regulator's objective function given parameters and firm characteristics is

$$\lambda_0^*, \lambda_1^* \in \underset{\lambda_1,\lambda_2}{\arg\min} \sum_{j=1,\dots,N} \int\int \mathcal{F}\big(\mathcal{I}(u_{1j}|X_j, T_j, \lambda, \beta, \rho) \times \big(V_0(P_j) - V_0(\tilde{P}_j)\big)|\phi, \mu_c, \sigma_c\big)\big) \times \tilde{P}\big(1-e^{\phi_2}\big)\, dF(U_2)\, dF(U_1) \tag{13}$$

$$\text{such that} \sum_{j=1,\dots,N} \int \mathcal{I}(u_{1j}|X_j, T_j, \lambda, \beta, \rho)\, dF(U_1) = N \cdot \bar{I}. \tag{14}$$

The optimal targeting rule is of the form

$$\mathcal{I}^*(u_{1j}|X_j, T_j, \lambda, \beta, \rho) = \lambda_2^* \Phi\bigg(\frac{\lambda_1 + X_j'\beta_1 + T_j\beta_2 + u_1}{\rho}\bigg). \tag{15}$$

In practice, the integrals over pollution shocks are approximated with draws of u_1 and u_2. Therefore, we write this objective function as a sum over simulations $s = 1, \dots, S$. Let $\mathcal{A}(\lambda|\cdot)$ represent the amount of abatement achieved with targeting parameters λ. The cost distribution is log normal, so

$$\mathcal{A}(\lambda|X_j, T_j, \beta, \phi, \sigma, u_s) = \sum_j \sum_s \Phi\bigg(\frac{\log\big[\mathcal{I}(u_{1j}|X_j, T_j, \lambda, \beta, \rho)\Delta_{js}\big] - \mu_c}{\sigma_c}\bigg) \times \tilde{P}\big(1-e^{\phi_2}\big),$$

where $\Delta_{js} = V_0(P_j) - V_0(\tilde{P}_j)$ is the reduction in expected penalties from taking the abatement action run. We omit arguments and substitute the form of the targeting function:

$$\begin{aligned}\mathcal{A}(\lambda) &= \sum_j \sum_s \Phi\bigg(\frac{\log\big[\mathcal{I}(u_1)\Delta_{js}\big] - \mu_c}{\sigma_c}\bigg) \times \tilde{P}\big(1-e^{\phi_2}\big) \\ &= \sum_j \sum_s \Phi\bigg(\frac{\log\big[\lambda_2\Phi\big((\lambda_1 + \beta_1 X_j + \beta_2 T_j + u_1)/\rho\big)\big)\Delta_{js}\big] - \mu_c}{\sigma_c}\bigg) \times \tilde{P}\big(1-e^{\phi_2}\big).\end{aligned}$$

Let $Z_{js} = (X_J'\beta_1 + T_j\beta_2 + u_1)/\rho$. Then we have

$$\begin{aligned}\mathcal{A}(\lambda) &= \sum_j \sum_s \Phi\bigg(\frac{\log\big[\lambda_2\Phi(\lambda_1/\rho + Z_{js})\Delta_{js}\big] - \mu_c}{\sigma_c}\bigg) \times \tilde{P}\big(1-e^{\phi_2}\big) \\ &= \sum_j \sum_s \Phi\bigg(\frac{\log\lambda_2 + \log\big[\Phi(\lambda_1/\rho + Z_{js})\big] + \log[\Delta_{js}] - \mu_c}{\sigma_c}\bigg) \times \tilde{P}\big(1-e^{\phi_2}\big).\end{aligned}$$

Let the argument of the cost distribution be denoted by C_{js}. Taking the derivative with respect to λ_2 yields

$$\partial\mathcal{A}(\lambda)/\partial\lambda_2 = \sum_j \sum_s \phi(C_{js})\frac{1}{\lambda_2\sigma_c} \times \tilde{P}\big(1-e^{\phi_2}\big).$$

With respect to λ_1,

$$\partial\mathcal{A}(\lambda)/\partial\lambda_2 = \sum_j\sum_s \phi(C_{js})\frac{1}{\sigma_c}\frac{1}{\Phi(\lambda_1/\rho + Z_{js})}\phi(\lambda_1/\rho + Z_{js})\frac{1}{\rho}\times\tilde{P}\big(1-e^{\phi_2}\big).$$

We can write the Lagrangian of the optimal targeting problem. Omitting arguments,

$$\begin{aligned}\mathcal{L}(\lambda_1,\lambda_2) &= \sum_{j=1}\int\int \mathcal{F}\big(\mathcal{I}_j(u_1)\big(V_0(P_j)-V_0(\tilde{P}_j)\big)\big)\times\tilde{P}_j\big(1-e^{\phi_2}\big)\,dF(U_1)\,dF(U_2)\\ &\quad-\gamma\Big(\sum_{j\in\mathcal{C}}\int\mathcal{I}_j(u_1)\,dF(U_1)-N\cdot\bar{I}\Big)\\ &= \sum_j\sum_s\Phi\bigg(\frac{\log\big[\mathcal{I}(u_{1j}|X_j,T_j,\lambda,\beta,\rho)\Delta_{js}\big]-\mu_c}{\sigma_c}\bigg)\times\tilde{P}_j\big(1-e^{\phi_2}\big)\\ &\quad-\gamma\bigg(\sum_{j\in\mathcal{C}}\lambda_2\Phi\bigg(\frac{\lambda_1+X_j'\beta_1+T_j\beta_2}{\sqrt{\rho^2+\sigma_1^2}}\bigg)-N\cdot\bar{I}\bigg).\end{aligned}$$

Here we denote by $\mathcal{C}$ the set of plants j in the control group. Thus the optimality constraint is that the targeting rule is optimal in the control group. We then calculate the first-order conditions of the constrained problem:

$$\frac{\partial\mathcal{L}}{\partial\lambda_2} = \frac{\partial\mathcal{A}}{\partial\lambda_2} - \gamma\sum_j\Phi\bigg(\frac{\lambda_1+X_j'\beta_1+T_j\beta_2}{\sqrt{\rho^2+\sigma_1^2}}\bigg) = 0, \tag{16}$$

$$\frac{\partial\mathcal{L}}{\partial\lambda_2} = \frac{\partial\mathcal{A}}{\partial\lambda_1} - \gamma\sum_j\lambda_2\phi\bigg(\frac{\lambda_1+X_j'\beta_1+T_j\beta_2}{\sqrt{\rho^2+\sigma_1^2}}\bigg)\frac{1}{\sqrt{\rho^2+\sigma_1^2}} = 0, \tag{17}$$

$$\frac{\partial\mathcal{L}}{\partial\gamma} = \sum_{j\in\mathcal{C}}\lambda_2\Phi\bigg(\frac{\lambda_1+X_j'\beta_1+T_j\beta_2}{\sqrt{\rho^2+\sigma_1^2}}\bigg) - N\cdot\bar{I} = 0. \tag{18}$$

The optimal targeting parameters λ_1 and λ_2 satisfy these first-order conditions with the Lagrange multiplier γ representing the shadow value of the inspection budget constraint.

APPENDIX D: ROBUSTNESS AND SENSITIVITY ANALYSIS

This Appendix studies the robustness of the structural estimates. Section D.1 considers the robustness of estimated targeting parameters with respect to the two calibrated parameters σ_c and ρ. Section D.2 considers the sensitivity of a broader set of parameters of interest with respect to variation in the underlying moments in the data, using the measure of Andrews, Gentzkow, and Shapiro (2017).

D.1. *Robustness to Calibrated Parameters*

The targeting stage estimation fixed the values of σ_c and ρ. Fixing these values reduces the number of free parameters in the maintenance cost distribution from two to one and in the targeting function from three to two. This subsection studies how the values of these parameters affect the values of estimated parameters.

Table S.XII shows the baseline targeting stage estimates in column 1 and estimates with alternate values of the fixed parameters in columns 2–4. The calibrated parameters

TABLE S.XII

ROBUSTNESS OF TARGETING ESTIMATES TO CALIBRATED PARAMETERS[a]

$\sigma_c =$ $\rho =$	0.50 0.25 (1)	0.25 0.25 (2)	1.00 0.25 (3)	0.50 0.15 (4)	0.50 0.35 (5)
Panel A. Targeting and Pollution Equations					
Pollution Equation					
Run equipment (=1)	−0.742	−0.604	−1.073	−0.703	−0.786
	(0.307)	(0.233)	(0.550)	(0.282)	(0.333)
Audit treatment	−0.102	−0.097	−0.111	−0.106	−0.099
	(0.085)	(0.084)	(0.089)	(0.085)	(0.086)
Audit treatment × inspection treatment	0.066	0.059	0.082	0.069	0.069
	(0.108)	(0.107)	(0.114)	(0.107)	(0.109)
Audit sample	0.613	0.607	0.624	0.618	0.601
	(0.137)	(0.135)	(0.143)	(0.137)	(0.138)
Region: Ahmedabad	−0.201	−0.186	−0.232	−0.214	−0.183
	(0.132)	(0.130)	(0.138)	(0.131)	(0.132)
Region: Surat	−0.371	−0.345	−0.426	−0.382	−0.351
	(0.164)	(0.161)	(0.177)	(0.163)	(0.166)
Constant	−0.004	−0.034	0.067	0.002	−0.010
	(0.103)	(0.096)	(0.129)	(0.100)	(0.105)
Targeting Equation					
Inspection targeting shift parameter (λ_1)	−0.219	−0.220	−0.203	−0.075	−0.457
	(0.066)	(0.066)	(0.062)	(0.034)	(0.137)
Inspection targeting level parameter (λ_2)	10.043	10.162	9.304	6.951	18.445
	(3.124)	(3.164)	(2.598)	(1.233)	(11.751)
Inspection treatment	0.162	0.161	0.168	0.121	0.182
	(0.025)	(0.025)	(0.025)	(0.017)	(0.036)
Audit treatment	−0.005	−0.005	−0.007	−0.006	−0.003
	(0.017)	(0.017)	(0.018)	(0.013)	(0.020)
Audit treatment × inspection treatment	0.016	0.015	0.018	0.015	0.012
	(0.021)	(0.020)	(0.022)	(0.016)	(0.023)
Audit sample	0.095	0.095	0.098	0.070	0.110
	(0.024)	(0.024)	(0.025)	(0.018)	(0.030)
Region: Ahmedabad	−0.221	−0.222	−0.232	−0.178	−0.237
	(0.044)	(0.044)	(0.045)	(0.033)	(0.057)
Region: Surat	−0.178	−0.179	−0.186	−0.144	−0.193
	(0.040)	(0.040)	(0.041)	(0.030)	(0.050)
Panel B. Distributions of Pollution and Maintenance Cost Shocks					
Standard deviation of observed pollution shock (σ_1)	0.111	0.110	0.117	0.093	0.112
	(0.022)	(0.022)	(0.023)	(0.016)	(0.028)
Standard deviation of unobserved pollution shock (σ_2)	0.866	0.855	0.899	0.862	0.871
	(0.042)	(0.037)	(0.073)	(0.041)	(0.045)
Mean of log maintenance cost (μ_c)	1.859	1.637	2.350	1.852	1.870
	(0.316)	(0.317)	(0.337)	(0.315)	(0.315)

[a]The table reports estimates of the targeting stage of the model under alternate values of the calibrated parameters σ_c and ρ. Each column shows one set of estimates of the unconstrained targeting model, that is, without imposing that the regulator's inspection rule is optimal. We use $S = 200$ simulations for each set of estimates. Column 1 shows the baseline estimates. (These estimates differ very slightly from those reported in Table 10 of the paper because Table 10 uses $S = 5000$ simulations.)

are shown in the column headers. Columns 2 and 3 change the value of σ_c from the baseline value of $\sigma_c = 0.50$ to $\sigma_c = 0.25$ and $\sigma_c = 1.00$, respectively, and columns 4 and 5 change the value of ρ from the baseline value of $\rho = 0.25$ to $\rho = 0.15$ and $\rho = 0.35$, respectively.

First consider the effect of altering σ_c on the parameter estimates. Since σ_c is a cost parameter and it affects the willingness to run maintenance equipment to reduce pollution, we expect and indeed find that there are very small effects of this parameter on the targeting equation coefficients. There are two main effects of altering σ_c: first, a higher σ_c in column 3, relative to column 1, increases the estimated μ_c and increases the efficacy of abatement (run equipment coefficient). The model accommodates higher dispersion of cost by moving the cost distribution up and increasing the efficacy of abatement, raising both the gross costs and benefits of abatement, so that the model can still match the moments on the share of plants that are willing to run and their cost conditional on run. The changes in estimates are noticeable but not very large; multiplying σ_c by a factor of 4 (from column 2 to 3) increases the efficacy of abatement by a factor of $-1.07/-0.60 = 1.78$, and both estimates are within 1 standard error of our baseline estimate in column 1. For comparison, in Table VI, the change in the estimated abatement efficacy from imposing the constraint of optimal targeting is much larger.

Next consider the change from altering ρ on the parameter estimates. Since ρ is an inspection targeting parameter, we expect and find that there are very small effects of changing this parameter (in columns 4 and 5) on the coefficients in the pollution equation (relative to column 1 baseline estimates). Changing ρ, which is the denominator of the argument of the targeting function, has predictably larger effects on the coefficients β and λ in the targeting function. In particular, for a smaller ρ (in column 4), we see smaller β estimates for the targeting equation, and for a larger ρ (in column 4), we see larger β estimates. These changes are roughly but not exactly proportional; for example, the inspection treatment coefficient normalized by ρ is $0.162/0.25 = 0.648$ in column 1 and $0.121/0.15 = 0.807$ in column 4. The argument of the targeting functions, for a plant with the same observables and u_1 shock, therefore changes with ρ somewhat, and the estimated parameters λ also change to fit the moments in the data. For example, in the column 4 estimates, the maximum inspections parameter λ_2 moves down for a smaller ρ and the shift parameter λ_1 becomes less negative, which offset the effects of changes in the β/ρ coefficients on inspections.

These counteracting shifts in estimated parameters will matter to the extent that the targeting function fits the data moments differently with different values of ρ. To get a sense of the net effect of these changes, in Table S.XIII we give the values of the main targeting moments, expected inspections, and expected inspections squared, from the model, calculated under each set of fixed and estimated parameters. The column headers are the same as in Table S.XII.

Looking across the columns of Table S.XIII, we see that the expected inspections moment ranges from 2.15 in the baseline case to as high as 2.18 and as low as 2.13. The range of expected inspections hardly varies, depending on whether the targeting function has a moderate ρ or, say, a lower ρ with smaller β estimates and shifts in λ, as in column 4. The expected inspections squared moment is only somewhat more variable, with a range from 6.95 to 7.36 depending on the fixed value of ρ. Because the targeting moments can be fitted about equally well with different values of ρ, offset by changes in the estimated $(\hat{\beta}, \hat{\lambda})$, the values of these parameters are not separately well identified, and estimation runs allowing a free ρ yielded very imprecise targeting parameters.

TABLE S.XIII

ROBUSTNESS OF EXPECTED INSPECTIONS TO CALIBRATION[a]

$\sigma_c =$	0.50	0.25	1.00	0.50	0.50
$\rho =$	0.25	0.25	0.25	0.15	0.35
	(1)	(2)	(3)	(4)	(5)
$\mathbb{E}$[inspections]	2.147	2.149	2.150	2.177	2.126
$\mathbb{E}$[inspections^2]	7.111	7.136	7.131	7.358	6.949

[a]The table reports how the model predictions for moments of expected inspections and expected inspections squared change depending on the value of the calibrated parameters σ_c and ρ. The rows show the values of the two moments and the columns show the predictions of the model at the estimated parameters with each set of calibrated parameters (shown in the column headers). The values of the expected inspections and expected inspections squared moments in the data are 2.20 and 7.53, respectively.

Finally, consider the effects of both calibrated parameters on the estimated standard deviations of pollution shocks in Table S.XII, panel B. The baseline estimate is $\hat{\sigma}_1 = 0.111$. Doubling or halving σ_c has nearly no effect on the estimated σ_1 ($\hat{\sigma}_1 = 0.110$ and $\hat{\sigma}_1 = 0.117$, respectively), and decreasing and increasing ρ also has small effects ($\hat{\sigma}_1 = 0.093$ and $\hat{\sigma}_1 = 0.112$, respectively). All these changes are within 1 standard error of the original estimate and typically far smaller. The effects of the fixed parameters on the standard deviations of unobserved pollution σ_2 are also small.

We conclude that the interpretation of the parameter estimates is robust to variations in the fixed parameters σ_c and ρ. In particular, the significance of direct changes in the ρ parameter for targeting appear to be offset by changes in the estimated λ and β parameters. The fixed parameters do not affect the key finding that the regulator has little information on pollution.

D.2. *Sensitivity of Parameter Estimates to Moments*

D.2.1. *Sensitivity Matrix: Definition*

We estimate the targeting function through the generalized method of moments with a mix of analytic and simulated moments (Section 5.2 and Appendix C.2). Andrews, Gentzkow, and Shapiro (2017) defines the sensitivity matrix Λ for any estimator $\hat{\theta}$ that minimizes a criterion function $\hat{g}(\theta)'\hat{W}\hat{g}(\theta)$, where $\hat{g}(\theta)$ is a vector of moments or other statistics and $\hat{W}$ is a weight matrix.

Assume that $\sqrt{n}\hat{g}(\theta_0)$ converges in distribution to $\tilde{g}$ such that $\mathbb{E}[\tilde{g}] = 0$ under the model. For alternative specifications of the model, this may not be the case. Andrews, Gentzkow, and Shapiro (2017) define local perturbations of the maintained model and show that, under these perturbations, $\sqrt{n}(\hat{\theta} - \theta_0)$ converges in distribution to a random variable $\tilde{\theta}$ such that $\tilde{\theta} = \Lambda\tilde{g}$. Then the estimator $\hat{\theta}$ has first-order asymptotic bias

$$\mathbb{E}[\tilde{\theta}] = \Lambda\mathbb{E}(\tilde{g}),$$

where $\Lambda = -(G'WG)^{-1}G'W$, W is the probability limit of the weight matrix $\hat{W}$, and G is the Jacobian of the probability limit of $\hat{g}(\theta)$ at θ_0. We can estimate Λ using the standard plug-in estimates of W and G at little extra computational cost.

Because the moments are in different units, we scale Λ so that it can be read as the effect of a 1 standard-deviation violation of the given moment condition on the asymptotic bias of the given parameter. The value of entry Λ_{jk} is, therefore, measured in the units of parameter k per standard deviation of moment j.

Sensitivity has two equivalent interpretations. Formally, sensitivity is defined as the asymptotic bias of the estimator under local misspecification. One can also think of sensitivity as approximating how a change in one data moment, such as would be generated by an alternative model, would locally affect the parameter estimates.

D.3. *Sensitivity of Selected Parameters*

Table S.XIV presents sensitivities for selected parameters (columns) with respect to the estimation moments. Each column of the table gives the sensitivity of the column parameter to the estimation moments listed down the rows. Table S.XV condenses this information into the four moments to which each main parameter of interest is most sensitive.

D.3.1. *Sensitivity of Inspection Treatment Coefficient*

To illustrate the intuition for the sensitivity measure, consider the sensitivity of the inspection treatment coefficient in the targeting equation (Table S.XIV, column 1). The sensitivity of the treatment coefficient to the product of the mean of inspections and the treatment dummy is 1.89 (the row "Inspection mean × T"). This sensitivity means that if the product of inspections and the treatment dummy were higher by 1 standard deviation, the estimated inspection coefficient β_2 would increase by 1.89 (within the argument

TABLE S.XIV

SENSITIVITIES FOR SELECTED PARAMETERS[a]

Moment	β_2 (1)	ϕ_2 (2)	λ_1 (3)	λ_2 (4)	ϕ_0 (5)	σ_1 (6)	σ_2 (7)	μ_c (8)
Pollution resid. $\times T$	−0.360	3.114	−0.938	41.689	0.758	−0.362	0.444	1.579
Pollution resid.	−0.873	−1.858	−2.406	105.770	3.901	−0.927	−0.119	−1.800
Pollution resid. $\times X_1$	0.026	0.819	0.087	−3.703	0.272	0.033	0.106	0.402
Pollution resid. $\times X_2$	−0.047	−1.464	−0.151	6.477	−0.458	−0.058	−0.186	−0.817
Pollution resid. $\times X_3$	−0.185	0.081	−0.505	22.258	−0.055	−0.195	0.045	0.080
Pollution resid. $\times X_4$	0.439	−0.206	1.195	−52.658	−2.651	0.461	−0.081	0.160
Pollution resid. $\times X_5$	0.383	−0.015	1.042	−45.914	−2.829	0.402	−0.047	0.216
Inspection mean $\times T$	1.888	−0.411	1.779	−112.948	−0.357	0.688	−0.075	0.254
Inspection mean	−1.210	0.223	0.292	44.942	0.402	−0.583	0.027	0.223
Inspection mean $\times X_1$	0.297	0.168	−0.061	−4.052	0.013	0.029	0.017	0.063
Inspection mean $\times X_2$	−0.335	−0.133	0.102	3.278	−0.007	−0.027	−0.013	−0.074
Inspection mean $\times X_3$	0.128	−0.080	0.260	−12.134	−0.004	0.033	−0.014	0.278
Inspection mean $\times X_4$	−0.182	−0.283	−0.622	16.114	−0.083	−0.017	−0.020	−0.312
Inspection mean $\times X_5$	−0.172	0.023	−0.644	14.989	−0.031	−0.014	0.011	−0.191
Insp. squared mean $\times T$	−2.085	1.746	−2.992	171.161	0.880	−1.205	0.220	0.336
Insp. squared mean	1.469	−2.163	1.283	−95.999	−1.003	1.044	−0.232	−0.742
Prob(run)	−0.005	1.026	−0.014	0.603	0.172	−0.005	0.107	−1.356
$E[c_j\|\text{run}]$	−0.049	10.317	−0.138	6.066	1.729	−0.053	1.072	7.256
Var(pollution)	−0.000	0.090	−0.001	0.053	0.015	−0.000	0.886	0.067
Cov(pollution, insp.)	0.872	−0.731	2.376	−104.703	0.185	0.917	−0.255	0.806

[a]This table presents values from the estimated sensitivity matrix Λ, scaled so that the entries can be read as the effect of a 1-standard-deviation violation of the given moment condition on the asymptotic bias of the given parameter. For the column corresponding to a given parameter, we can interpret the entries as the sensitivity of the estimated parameter to beliefs about the degree of misspecification of each moment, expressed in standard deviations. Alternatively, we can interpret the estimated sensitivities as a measure of how estimates will respond to changes in the underlying data. While sensitivity is computed with respect to the complete set of variables, the table shows only a subset of particular interest.

TABLE S.XV

MOST IMPORTANT MOMENTS FOR SELECTED PARAMETERS[a]

Variable	Most Influential Moment (1)	Value (2)	2nd Most Influential Moment (3)	Value (4)	3rd Most Influential Moment (5)	Value (6)	4th Most Influential Moment (7)	Value (8)
Insp. treatment (β_2)	Insp. squared mean$\times T$	−2.085	Insp. mean$\times T$	1.888	Insp. squared mean	1.469	Inspection mean	−1.210
Run (ϕ_2)	$E[c_j\|\text{run}]$	10.317	Pollution resid.$\times T$	3.114	Insp. squared mean	−2.163	Pollution resid.	−1.858
Inspection shift (λ_1)	Insp. squared mean$\times T$	−2.992	Pollution resid.	−2.406	Cov(pollution, inspection)	2.376	Insp. mean$\times T$	1.779
Inspection level (λ_2)	Insp. squared mean$\times T$	171.161	Insp. mean$\times T$	−112.948	Pollution resid.	105.770	Cov(pollution, inspection)	−104.703
Constant (ϕ_0)	Pollution resid.	3.901	Pollution resid.$\times X_5$	−2.829	Pollution resid.$\times X_4$	−2.651	$E[c_j\|\text{run}]$	1.729
SD observed shock σ_1	Insp. squared mean$\times T$	−1.205	Insp. squared mean	1.044	Pollution resid.	−0.927	Cov(pollution, inspection)	0.917
SD unobserved shock (σ_2)	$E[c_j\|\text{run}]$	1.072	Var(pollution)	0.886	Pollution resid.$\times T$	0.444	Cov(pollution, inspection)	−0.255
Mean maintenance cost (μ_c)	$E[c_j\|\text{run}]$	7.256	Pollution Resid.	−1.800	Pollution resid.$\times T$	1.579	Prob(run)	−1.356

[a]Table 2 presents the names and values of the four moments with the largest absolute value of sensitivity for each parameter. The sensitivities are scaled so that the entries can be read as the effect of a one-standard-deviation violation of the given moment condition on the asymptotic bias of the given parameter. The magnitudes are thus comparable across moments for a given parameter, and can be interpreted as the degree of influence of a moment on the estimated parameter.

of the targeting function; the marginal effect of this change on mean inspections therefore depends on covariates). In this simple case, sensitivity is easy to understand. The higher are the mean inspections for treated firms, relative to control, the higher is the estimated treatment effect. Table S.XV shows that the four most influential moments for β_2 are (i) the mean of the product of treatment with inspections squared, (ii) the mean of the product of treatment with inspections, (iii) mean inspections squared, and (iv) mean inspections.

We now turn to consider the sensitivity of several parameters of interest to the estimation moments. We focus on the parameters of regulatory information, abatement cost, and the efficacy of pollution abatement

D.3.2. *Sensitivity of Inspection Targeting Parameters λ_1 and λ_2*

The inspection targeting equation relates unobserved (by the econometrician) pollution shocks and plant observables to inspections. A higher level parameter λ_2 gives the maximum number of inspections a plant may receive and a higher shift parameter λ_1 increases the argument of the targeting function, so that all plants receive a higher level of inspections.

Table S.XV shows that interactions of the treatment with the inspection distribution moments are influential for the targeting parameter estimates. In the λ_1 row, the column 2 sensitivity of -2.99 means that a 1-standard-deviation higher mean product of inspections and treatment, conditional on other moments, would decrease the inspection shift parameter λ_1 by 2.99. If squared inspections in the treatment group were relatively higher, without changing mean inspections, the estimated inspection shift parameter λ_1 would decrease. This shift downward would allow the model, with a steep targeting function, to match a *relatively* higher volatility of inspections in the treatment, since plants in the treatment group are more likely to be shifted out onto the steep part of the targeting function. The targeting parameters are also sensitive to other interactions of the treatment with the inspection distribution.

The units of sensitivity are not comparable across parameters, since they are measured in the units of each parameter. The sensitivities for λ_2 are so large because λ_2 gives the maximum number of inspections for an arbitrarily large observable pollution shock. Given the estimated shape of the targeting function, this maximum is not reached in the sample. Movements in λ_2 have a lesser effect on the changes in inspections for plants with in-sample covariates and more likely shocks.

The parameters λ_1 and λ_2 are sensitive to many of the same moments, but in opposite directions. The sensitivities of these parameters have different signs for 17 of 20 moments (Table S.XIV) and for all of the 4 most influential moments (Table S.XV). In addition to the moments of the inspection distribution, the inspection shift parameter is sensitive to the pollution residual and the covariance of the pollution and inspection residuals (Table S.XV, columns 4 and 6). This relationship arises since λ_1 is the constant in the argument of the targeting function, and u_1, the observed pollution shock, enters the same argument. The level of λ_1 must adjust to match the observed level and dispersion of inspections given that the observed pollution shock has mean zero. To put it another way, if the pollution residual increased and the inspection shift parameter did not change, the model would overpredict inspections on average.

D.3.3. *Sensitivity of Abatement Efficacy ϕ_2*

The sensitivities show that several moments affect the estimated effect of running abatement equipment on pollution, ϕ_2. The most obvious ex ante are that the ϕ_2 estimate is sensitive to the pollution residual times the treatment and the pollution residual. If treatment pollution were higher by 1 standard deviation, then $\hat{\phi}_2$ would increase from -0.71 to -0.51, indicating a decline in abatement efficacy of 11 percentage points (efficacy being $1 - \exp(\phi_2)$). If the treatment had reduced pollution less than observed, the model would infer that abatement was less effective.

The efficacy of abatement depends on the pollution equation, as it would exclusively in a single-equation model, but also the cost of maintenance. Table S.XV shows that pollution resid.$\times T$ is only the second most influential moment for ϕ_2, after the moment giving the abatement maintenance cost conditional on running equipment. If the mean maintenance cost conditional on running increased by 100 dollars, the $\hat{\phi}_2$ coefficient would increase from -0.71 to -0.53 (panel ϕ_2; moment $\mathbb{E}[c_j|\text{run}]$).

D.3.4. *Variances of Observed and Unobserved Pollution Shocks, σ_1 and σ_2*

A goal of the model is to understand what the regulator knows about pollution. Sensitivity analysis can tell us how the model separates the pollution shock into an observed component with standard deviation σ_1 and an unobserved component with standard deviation σ_2.

Table S.XV, row σ_2 shows that the unobserved component of pollution is most sensitive to maintenance cost, the variance of pollution, and the pollution residual interacted with treatment, in that order. Directly, if the variance of observed log pollution were higher, then the unobserved pollution shock would be estimated to be higher also (column 4). More subtly, if expected maintenance cost is higher conditional on running, for example, then it must be that there is greater variance in unobserved pollution, so as to induce plants to be willing to run their costly equipment (given a fixed inspection targeting function, set by parameters λ).

The moments important for the observed component σ_1 are notably distinct from those that are important for σ_2. The moment to which σ_1 is most sensitive is not a pollution moment but the product of squared inspections and treatment. Since the observed pollution shock enters the targeting function and since treatment plants have a higher targeting function argument (i.e., higher inspections), the observable pollution shock is estimated to be larger if the dispersion of inspections is greater for treatment plants (column 2) and for control plants (column 4). The squared inspections moments are the uncentered second moments of the targeting function and so capture how much variation in pollution there is on which the regulator can target. If the covariance of pollution and inspection residuals is higher, this increases the observable part of the pollution shock (row σ_1, column 8) but decreases the unobservable part (row σ_2, column 8). The model can, therefore, separate the two pollution shocks, since only the observable shock influences the distribution of inspections, through targeting, and since they have effects of opposite sign on the covariance between pollution and inspections.

REFERENCES

ANDREWS, I., M. GENTZKOW, AND J. M. SHAPIRO (2017): "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *The Quarterly Journal of Economics*, 132 (4), 1553–1592.[21,24]

BANERJEE, A. V., AND E. DUFLO (2014): "Do Firms Want to Borrow More? Testing Credit Constraints Using a Directed Lending Program," *The Review of Economic Studies*, 81 (2), 572–607.[16]

DUFLO, E., M. GREENSTONE, R. PANDE, AND N. RYAN (2013): "Truth-Telling by Third-Party Auditors and the Response of Polluting Firms: Experimental Evidence From India," *The Quarterly Journal of Economics*, 128 (4), 1499–1545.[4]

Co-editor Liran Einav handled this manuscript.

Manuscript received 30 September, 2014; final version accepted 15 March, 2018; available online 18 May, 2018.