American Economic Review 2021, 111(7): 2101–2126
https://doi.org/10.1257/aer.20181346

Unwatched Pollution: The Effect of Intermittent Monitoring on Air Quality†

By Eric Yongchen Zou*

Intermittent monitoring of environmental standards may induce strategic changes in polluting activities. This paper documents local strategic responses to a cyclical, once-every-six-day air quality monitoring schedule under the federal Clean Air Act. Using satellite data of monitored areas, I show that air quality is significantly worse on unmonitored days. This effect is explained by short-term suppression of pollution on monitored days, especially during high-pollution periods when the city's noncompliance risk is high. Cities' use of air quality warnings increases on monitored days, which suggests local governments' role in coordinating emission reductions. (*JEL* K32, Q35, Q58, R11)

Enforcement of environmental regulation relies on accurate monitoring of compliance behavior. In practice, limited budgets often force monitoring to be conducted on an intermittent basis, creating opportunities for polluters to comply when they are monitored, but to increase polluting activities when they are not. The potential for such strategic responses grows when polluters can anticipate the regulator's monitoring schedules, as demonstrated by the Volkswagen emissions scandal.[1] Nonetheless, strategic responses to intermittent monitoring are generally difficult to detect. There are two main challenges. First, independent measurement of polluting behavior during unmonitored times is usually unavailable. Second, the timing of monitoring is often nonrandom. Thus, a simple comparison of monitored and unmonitored polluting activities likely confounds strategic responses with latent factors (such as pollution leaks) that may have triggered monitoring or inspections in the first place.

This paper documents strategic responses to a broad-scale federal air pollution regulation. I explore a unique empirical context in which monitoring is based

* Department of Economics, University of Oregon and NBER (email: ericzou@uoregon.edu). Henrik Kleven was the coeditor for this article. I am grateful to Dan Bernhardt, Tatyana Deryugina, Don Fullerton, Nolan Miller, and David Molitor for invaluable guidance and support. This paper has also benefited from comments and suggestions by anonymous referees, David Albouy, Soren Anderson, Max Auffhammer, Wesley Blundell, Mark Borgschulte, Karen Brandon, Olivier Deschenes, Eliza Forsythe, Cynthia Giles, Daniel Goldberg, Corbett Grainger, Michael Greenstone, Koichiro Ito, Henrik Kleven, Shanjun Li, Ben Marx, Joseph Shapiro, Wolfram Schlenker, Rebecca Thornton, and seminar and conference participants at University of Illinois at Urbana-Champaign, Cornell University, UC Santa Barbara, University of Oregon, UC Berkeley, Wharton Risk Management Center, University of Tokyo, the MEA Meeting, the SEA Meeting, the WEAI Conference, the NBER Summer Institute, the Heartland workshop, and the Environmental Protection Agency. All errors are mine.

† Go to https://doi.org/10.1257/aer.20181346 to visit the article page for additional materials and author disclosure statement.

[1] See Guilbert Gates et al., "Explaining Volkswagen's Emission Scandal," *New York Times,* July 19, 2016.

on a publicly available, quasi-random schedule. Under the Clean Air Act, the US Environmental Protection Agency (EPA) is charged with enforcing a national safety level for outdoor particulate matter air pollution (henceforth, PM) that all counties are required to achieve. A network of monitoring sites tracks compliance with the standard. Due to high operating costs, the EPA grants permission for many sites to monitor pollution on an intermittent basis. To balance between the goal of obtaining representative monitoring results and the administrative costs that would be incurred by a random monitoring scheme, the EPA announces a cyclical, once-every-six-day ("1-in-6-day") monitoring schedule that will be followed each year. I leverage this monitoring rule as a policy experiment to examine whether higher concentrations of pollution occur during days when monitoring is not scheduled.

I begin by constructing an indirect measure of particulate pollution using 13 years of satellite observations. This measure allows me to observe air quality both during monitors' "on-days" and "off-days." Because the incentive to avoid monitoring is plausibly the only factor that changes on a 1-in-6-day basis, differences in pollution levels during on-days and off-days provide evidence of strategic responses to the monitoring schedule. In the baseline analysis, I use the satellite measure to compare on-day and off-day pollution levels around monitoring sites that follow the 1-in-6-day schedule. The results suggest a significant pollution gap between on-days and off-days. The satellite detects 1.6 percent less particulate pollution during on-days than during off-days, while pollution levels during off-days do not differ significantly from each other. The effect size is comparable to the average difference in air quality observed between weekdays and weekends. At the same time, placebo tests show no detectable pollution gap in the absence of an incentive to avoid monitoring. For example, the effect disappears when a monitor retires; furthermore, no pollution gap surfaces around sites that monitor pollution every day.

I then explore underlying mechanisms in three ways. First, I examine when gaming occurs. I find that gaming most likely reflects short-term cutbacks of polluting activities during "critical" times, for example, when counties' noncompliance risk is high. I show that the average pollution gap masks substantial heterogeneity, featuring large (over 7 percent) pollution gaps when a county's recent PM levels had approached the regulatory standard, and almost zero pollution gaps when the same county had experienced recent good air quality. In addition, I leverage the retirement of 1-in-6-day monitors, finding evidence that a decline of on-day pollution, rather than a substitution between on- and off-days, drives the pollution gap. In a way, this pattern resembles the key feature seen in the case of vehicle emission cheating, with monitored pollution levels lower than otherwise "normal" pollution levels when unmonitored. This pattern suggests occasional short-term suppression of polluting activities, rather than planned, routine substitution of pollution across all days.

Second, I examine the question of who is gaming. The EPA uses monitoring results to determine if *counties* (rather than any particular factories) are complying with the air quality standards. State and local governments are thus subject to the regulatory penalties, and they often exhibit strong aversion to noncompliance risks. To examine the possibility of local governments' strategic responses to intermittent monitoring, I analyze local governments' issuance of public air quality advisories ("Action Days"). Often issued during high pollution periods, these advisories call for citizens to reduce polluting activities. Analyzing the history of Action Day advisories in 346 cities and metro areas, I find a significant, 10 percent higher likelihood that an advisory is issued on PM monitoring days. The strategic use of Action Day advisories provides yet another example of gaming during critical times. More importantly, this analysis points to a role of coordination: although gaming of air quality monitoring is unlikely to arise solely from cooperative polluting activities, local government coordination may play a role.

Third, I use a cross-sectional analysis to characterize regulatory and industrial properties of counties that exhibit large (top 10 percent) pollution gaps. Data suggest that incentives matter: areas with large pollution gaps have higher pollution levels, a history of receiving penalties for violations, and the presence of intermittent monitoring sites. Moreover, areas with large pollution gaps also tend to be those in which the composition of local industry allows for strategic responses. For example, having a high concentration of wood product manufacturers, a highly polluting industry that often operates at partial production capacity, stands out as a strong and robust predictor of large pollution gaps. Together, evidence suggests that the regulatory *incentive* to avoid monitoring underlies strategic responses, and a *capacity* to respond also matters.

This paper is the first to introduce the idea of "space-truthing" into the context of intermittent monitoring. While *in situ* monitoring data are often considered as the "ground truth" in remote sensing science (e.g., Liu et al. 2007), satellite data can conversely provide validation of the integrity of ground monitoring data, especially when there are concerns about biases in how ground monitoring is carried out. A few economics papers have found, in various regulation contexts, discrepancies between satellite and ground monitoring results (Karplus, Zhang, and Almond 2018; Fowlie, Rubin, and Walker 2019; Sullivan and Krupnick 2019). My paper highlights satellites' unique ability to directly detect *strategic responses* that are otherwise difficult to observe.[2]

This paper is among the first to report evidence on local governments' strategic actions in response to federal environmental regulations. A misalignment of regulatory objectives can motivate local governments to achieve hyper-localized pollution reductions within noncompliant jurisdictions (Auffhammer, Bento, and Lowe 2009; Bento, Freedman, and Lang 2015), or even to obscure true pollution levels from the federal regulator (Greenstone et al. 2020). My paper is most related to Grainger, Schreiber, and Chang (2017), who use satellite data to investigate state governments' strategic *location* choices in placing air quality monitors. I focus on gaming in *time*, and propose specific channels, such as strategic Action Day advisories, through which the local governments can game intermittent monitoring. The enforcement challenge highlighted in my paper echoes prior literature on medium- and long-term regional or sectoral substitution of polluting activities in the face of incomplete regulations (e.g., Becker and Henderson 2000; Hanna 2010; Fowlie, Reguant, and Ryan 2016). By introducing short-term responses, my paper

[2] In rare cases, researchers conduct independent environmental monitoring and compare directly with the regulatory monitoring results. See, for example, Duflo et al. (2013, 2018) and Thompson et al. (2014).

contributes to the nascent yet developing understanding of environmental regulation monitoring and enforcement (Gray and Shimshack 2011, Shimshack 2014).[3]

On the policy front, this paper provides the first retrospective evaluation of the long-standing practice of intermittent air quality monitoring (Akland 1972, Gilbert 1987). While the monitoring network is shifting toward advanced, continuous technologies, the transition has been surprisingly slow.[4] Though my analysis does not touch on political complexities that regulators might face in designing the monitoring system, my results demonstrate an observable and significant margin along which a continuous system could improve the enforcement of national air quality standards.

The rest of the paper is organized as follows. Section I provides a brief background on particulate pollution regulation and monitoring in the United States. Section II presents the main identification of the pollution gap between monitored and unmonitored days. Section III explores characteristics of the pollution gap that shed light on mechanisms. Section IV concludes.

I. Background

A. *Particulate Matter Regulation*

Regulation of particulate matter pollution in the United States is coordinated under the Clean Air Act, which provides oversight of PM pollution at various industry levels. The regulation on *ambient* PM can be viewed as a policy lever to achieve the Act's ultimate goal: to maintain outdoor pollution concentrations below the established safety levels, known as the National Ambient Air Quality Standards (NAAQS). Three effective PM NAAQS were in place during my study period: the three-year average of daily "fine" particulate matter (PM2.5) level had to be below $15\,\mathrm{ug/m^3}$; the three-year average of annual ninety-eighth percentile PM2.5 level had to be below $35\,\mathrm{ug/m^3}$; the maximum "coarse" particulate matter (PM10) level had to be below $150\,\mathrm{ug/m^3}$. Every year, the EPA categorizes counties into "attainment" and "nonattainment" groups based on monitoring results.[5]

A nonattainment status results in substantial regulatory costs for state and local governments and factories. As a part of NAAQS provisions, state and local governments are required to develop State Implementation Plan (SIP) for maintaining or improving air quality in their jurisdictions. A SIP specifies how states design and enforce emission permitting programs to regulate existing and new polluting sources. Per NAAQS, plans for factories in nonattainment jurisdictions must be much more stringent, incorporating measures such as more frequent inspections and greater fines (e.g., Blundell, Gowrisankaran, and Langer 2018). Factories planning

[3] Increasingly, enforcement challenges have been revealed in various other contexts. Examples include the gaming of vehicle exhaust testing (Oliva 2015, Reynaert and Sallee 2016, Alexander and Schwandt 2019), and emission ramp-ups after sunset (Vollaard 2017; Agarwal, Qin, and Zhu 2020). Outside of the context of regulation, this paper is linked to the growing use of remote sensing data in economic and policy research in general (Donaldson and Storeygard 2016).

[4] Online Appendix Section A reports that if the current trend continues, it will take until year 2035 before the entire PM2.5 monitoring system becomes continuous.

[5] From 2001 to 2013, roughly 30 percent of monitored counties had ever been assigned a nonattainment status. Among them, roughly 60 percent of county $\times$ years are associated with the violation of the PM2.5 annual standard.

new production capacity in nonattainment jurisdictions must adopt technologies that achieve the lowest possible emission rates, irrespective of the cost of doing so. Nonattainment penalties have been shown to cause significant losses in firm productivity. Greenstone, List, and Syverson (2012) find that a county's nonattainment status is associated with significant reduction in manufacturing industry productivity, amounting to about \$20 billion of revenue annually. Walker (2013) shows a labor displacement cost in the order of \$8 billion for firms that locate in newly designated nonattainment counties. The stakes are high for state and local governments, too. In addition to fiscal losses from lower firm productivity, sustained nonattainment can give rise to direct penalties from the EPA. These penalties include financial sanctions that prohibit the approval of almost any highway project or grant in the nonattainment jurisdictions. A separate strand of literature finds evidence that local governments direct regulatory resources to areas in violation, and have been able to achieve localized air quality improvements near noncompliant monitors (e.g., Auffhammer, Bento, and Lowe 2009; Bento, Freedman, and Lang 2015).

B. *Particulate Matter Monitoring*

A network of more than 1,200 sites across the country monitor ambient PM concentrations. These monitoring sites are usually placed in areas with high population density to ensure a reasonably representative measure of population exposure. During my study period, the PM monitoring network spanned more than 600 counties that included over 70 percent of US population.

Unlike monitoring of many gaseous air pollutants (such as ozone) that uses automated, laser-based methods, PM monitoring mostly relies on filter-based methods involving substantial manual operation and maintenance tasks such as field sample collection and laboratory analysis.[6] EPA data show that the annualized per-site cost of PM monitoring associated with monitor procurement, operation, and maintenance is roughly \$21,000 with a 1-in-6-day schedule and \$41,000 with every-day (1-in-1-day) monitoring (US EPA 1993). With roughly 600 sites operating on a 1-in-6-day schedule, the cost savings from intermittent monitoring aggregates to about \$12 million per year, vis-à-vis the status quo spending of \$48 million per year on the entire PM monitoring network.

The cyclical 1-in-6-day monitoring method was introduced by Akland (1972). During the 1980s, the EPA introduced more frequent, 1-in-3 day and every-day sampling schedules. These more frequent schedules are often deployed in areas with higher levels of pollution (US EPA 1985).[7] Over time, the EPA has been tending toward more frequent monitoring. With the initiation of PM2.5 monitoring in the 1990s, the EPA began to stipulate that all PM2.5 monitors to operate on a minimum of 1-in-3-day basis (40 CFR 58.13 1997, US EPA 1998a). However, in response to

[6] In online Appendix Section A, I document a growing availability of continuous PM monitoring technologies since mid-2000s. These continuous technologies typically rely on laser and microbalance techniques to infer PM concentration, and they have been commonly used in states' public informational program (such as AirNow). Continuous monitors are not deployed in NAAQS nonattainment designation until 2009, when a growing number of technologies were classified by the EPA as "equivalent" to manual methods. In my analysis, I treat continuous monitors as every-day (1-in-1-day) monitors. Continuous monitors that are not considered in NAAQS regulation are not included in my analysis.

[7] Online Appendix Section A provides more details on determinants of schedule assignment.

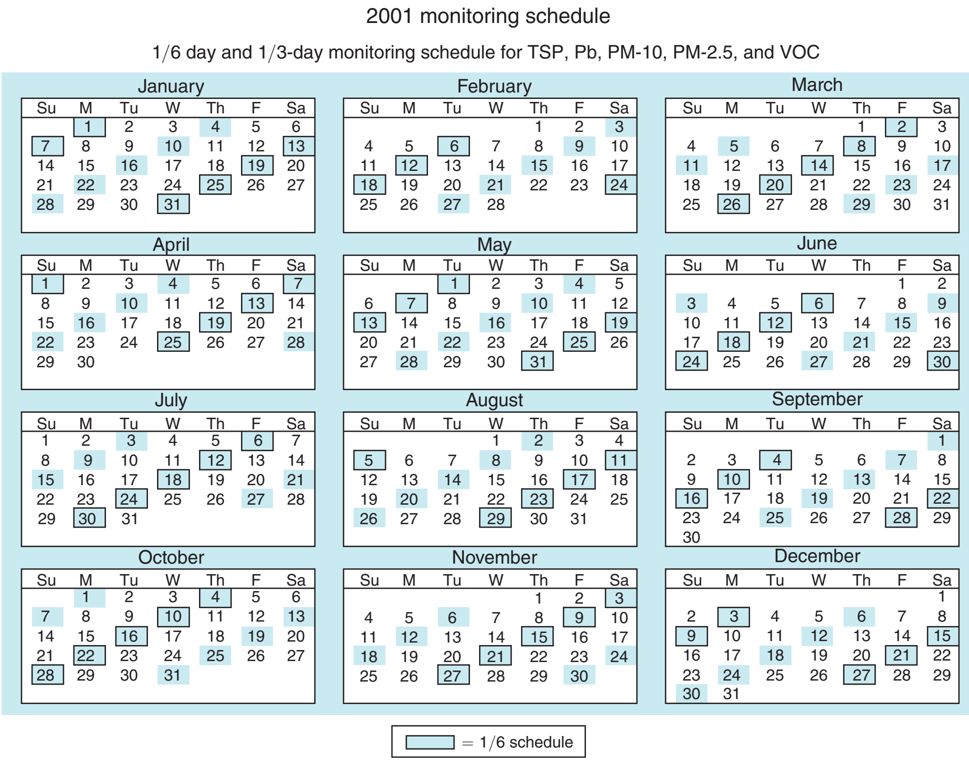

FIGURE 1. EPA'S AMBIENT POLLUTION MONITORING SCHEDULE, 2001

Notes: This figure shows the EPA's 2001 monitoring schedule calendar. Full archives of all calendars can be found at www3.epa.gov/ttnamti1/files/ambient/pm25/calendar.html (accessed May 2021).

states' concerns over cost burdens, the EPA allows exemptions for lower monitoring frequencies on a case-by-case basis (US EPA 1997a, US EPA 1998b). From 2001 to 2013, the vast majority of PM monitors followed either a 1-in-6-day (42 percent of monitors), 1-in-3-day (33 percent), or every-day (22 percent) schedule. While my main analysis focuses on the 1-in-6-day schedule in which gaming is most likely to occur, I test for a potential pollution gaps at all monitored sites (including the 1-in-3-day and the every-day monitoring sites). The sites that are monitored everyday serve as a "placebo" test where no strategic responses are expected.

The monitoring schedule is public. The EPA publishes a monitoring calendar on its website at the end of each calendar year, informing states of the monitoring dates for all intermittent monitoring sites for the next calendar year. Figure 1 presents the calendar for 2001.

II. The 1-in-6-Day Pollution Gap

A. *Data and Summary Statistics*

Monitor Data.—I obtain PM monitor characteristics from the EPA's Air Quality System (AQS) for the years 2001 to 2013. The annual summary data files of the

AQS are the source of monitor-level information on scheduled number of monitored days, actual monitored days, latitude and longitude location, and annual PM concentration statistics, such as the mean and the maximum. I identify 1-in-6-day (1-in-3 day, 1-in-1 day) monitors by finding monitors that are required by the EPA to sample 60 or 61 (121 or 122, 365 or 366) days a year.

Satellite Data.—I construct a measure for atmospheric particle pollution ("aerosol") using satellite data from the National Aeronautics and Space Administration (NASA) Moderate Resolution Imaging Spectroradiometer (MODIS) algorithm (Kaufman et al. 1997, Remer et al. 2005). Equipped with a flexible set of spectral radiance instruments, MODIS retrieves atmospheric aerosol concentrations by measuring the extinction of sunlight, based on knowledge of aerosol's ability to scatter and absorb light at different spectral wavelengths. MODIS summarizes aerosols in a dimensionless index called aerosol optical depth, which has a theoretical range of -0.05 to 5, with a smaller value corresponding to lower level of aerosol concentration. In the United States, the index's value mostly falls within the range of 0 to 1, with a mean of roughly 0.12.[8] To simplify language, in subsequent sections I refer to this measure as "aerosol" or "aerosol concentration."

Two distinctions are worth noting between the satellite-based aerosol measure and the traditional ground monitor-based PM measure. First, while PM monitors measure pollution concentration at a specific point on the ground, the satellite measure captures aerosol conditions for the entire "column" of air from its viewpoint 700 km above the ground. Also, the satellite measures pollution at a spatial resolution of 10 km $\times$ 10 km (about one-half of the size of the average US ZIP code), rather than pollution at any exact point. Second, most ground-monitoring data provide the average pollution concentration within a given time frame, usually a 24-hour period. By contrast, the satellite data come from a polar-orbiting satellite (Terra), providing snapshots of pollution in a given area at approximately 10:30 a.m. local time every day.

Despite the difference in measurement approaches, these ground and satellite checks have a similar *target* of measurement. PM monitors retrieve the concentration of small air particles (e.g., nitrates, sulfates, and black carbons) by measuring the amount of particles deposited when air is passed through a size-discriminating filter media; when in the atmosphere, these particles interact with sunlight, and therefore, they are picked up by the satellite measure. Such an overlap in the measurement target has been the foundation of a large body of atmospheric science literature that documents a strong correlation between the satellite-based and ground-monitoring data (e.g., Liu et al. 2007).

The key dataset I construct is a daily panel of aerosol concentration linking each 10 km $\times$ 10 km grid cell from 2001 to 2013. The geographic unit of my analysis is a monitoring site, and the aerosol level around a site is defined as the aerosol level within the 10 km $\times$ 10 km grid cell that corresponds to the site's location. The

[8] I use the "MOD04_L2" product which provides the simplest cloud-screened, quality-controlled product developed out of the raw imageries. MODIS provides more processed versions of the aerosol data that involve interpolation and smoothing in time and/or in space, such as the "L3" gridded files. Because my research design exploits monitoring frequency variations in time and/or in space, I use the least-processed data to minimize the impact of ex post modeling on my findings.

Table 1—Summary Statistics

	AOD			Exceedance per year		
Year	1/1d (1)	1/3d (2)	1/6d (3)	1/1d (4)	1/3d (5)	1/6d (6)
2001	17.9	17.6	15.2	7.0	2.4	2.0
2002	21.6	22.2	18.9	6.3	2.2	1.9
2003	18.4	18.8	16.0	4.2	1.7	1.3
2004	15.8	16.3	14.0	3.7	1.4	1.1
2005	16.7	19.5	16.5	4.6	2.6	1.7
2006	16.0	18.5	16.3	3.5	1.1	1.5
2007	19.6	22.1	18.5	3.9	1.6	1.9
2008	16.8	18.5	17.0	2.8	0.9	1.3
2009	14.8	16.2	13.7	2.2	0.7	1.0
2010	14.0	15.6	13.7	1.8	0.7	0.7
2011	15.9	19.1	16.3	2.6	0.4	1.1
2012	15.9	17.9	16.1	1.8	0.3	0.8
2013	14.1	15.3	14.3	2.4	0.6	1.2
	Number of sites			Population in million		
Year	1/1d (7)	1/3d (8)	1/6d (9)	1/1d (10)	1/3d (11)	1/6d (12)
2001	250	593	776	94.2	145.2	142.6
2002	277	660	774	108.2	162.3	143.3
2003	263	576	725	100.7	144.4	142.4
2004	274	628	737	100.8	145.9	149.6
2005	279	611	679	89.8	146.4	141.2
2006	286	596	690	95.9	148.9	141.2
2007	311	559	600	109.4	145.4	138.8
2008	313	520	538	103.4	149.0	133.9
2009	347	543	537	116.2	154.1	132.2
2010	376	513	522	126.9	144.9	132.0
2011	394	441	458	112.7	128.9	124.8
2012	436	461	426	119.0	132.0	114.5
2013	489	471	406	132.7	136.8	112.9

Notes: Each row represents statistics for a calendar year. Columns 1 to 3 show grid-level mean aerosol levels around 1-in-1-day, 1-in-3-day, and 1-in-6-day sites. Columns 4 to 6 show monitor-level total number of days that any PM NAAQS standards were exceeded in the year. Columns 6 to 9 count number of monitoring sites falling in each category. Columns 10 to 12 report population (in millions) of counties with each type of monitoring.

analysis is done at the site level (rather than at the monitor level) because there is no effective difference in the locations of different monitors within the site. To be conservative in aggregating monitor-level schedules to the site level, I define a site to be a 1-in-6-day site if any monitor in that site follows the 1-in-6-day schedule. Defining the sample this way is expected to work against finding evidence of strategic responses. For example, some monitoring sites may have an every-day monitor *and* a 1-in-6-day monitor, where the less frequent monitor is used to provide quality assurance data.[9]

[9] In online Appendix Table D.1, I confirm that strategic response is stronger if I restrict to sites with a standalone 1-in-6-day monitor. I find no evidence of strategic response when a 1-in-6-day monitor is co-located with an every-day monitor.

Summary Statistics.—Table 1 presents summary statistics by calendar year and by monitoring frequency. Columns 1 to 3 show average aerosol levels around monitoring sites. Over the period of study, the aerosol level stayed relatively stable, declining by an average of 1.5 percent per year. Columns 4 to 6 show average number of days during the year that exceeded any NAAQS standards.[10] Columns 7 to 9 report number of monitoring sites. The number of 1-in-6-day sites decreased over time due to a NAAQS revision in 1997 that initiated PM2.5 monitoring and redirected sources from monitoring of PM10, which are often operated on a 1-in-6-day basis (US EPA 1997b). In the analysis, I exploit monitor retirement events to show that the pollution gap between off- and on-days closes as monitors retire. Columns 10 to 12 calculate population of counties with 1-in-6-day, 1-in-3-day, and every-day monitoring. On average, about 200 million people live in counties with any form of PM monitoring. Among this population, 67 percent live in counties with at least one 1-in-6-day monitoring site.

B. *Empirical Framework*

Estimation Equation.—The strictly 1-in-6-day design of the monitoring schedule motivates a straightforward identification strategy that estimates the causal effect of the schedule on pollution by simply comparing levels of air pollution across days of a 1-in-6-day monitoring cycle. The primary estimation equation is

$$(1) \qquad Aerosol_{st} \;=\; \beta \cdot \mathbf{1}(Offdays_t) + Time_t + \alpha_s + X_{st}\gamma + \varepsilon_{st},$$

where $Aerosol_{st}$ is the logged satellite aerosol concentration at monitoring site s at time t, and $\mathbf{1}(Offdays_t)$ is a dummy variable that indicates days when monitoring is scheduled off (five out of six days of the monitoring cycle). The key coefficient of interest is β which represents the gap in pollution levels between an average off-day and an average on-day. The strictly 1-in-6-day cyclicality in the $\mathbf{1}(Offdays_t)$ variable implies that very few confounders may bias the OLS estimate $\hat{\beta}$ from identifying the causal effect of the monitoring schedule. To confirm this point, I report results from two types of specifications. In the first, I report regressions conditional on *no* covariates, so that $\hat{\beta}$ is simply the raw difference between off-days and on-days. Second, I report regressions that include a rich array of controls including time fixed effects $Time_t$ (year, month-of-year, and day-of-week fixed effects), monitoring site fixed effects (α_s), as well as weather controls (X_{st}) including daily temperature categorized into ten 10-degree bins, daily wind speed quartiles, and quadratic daily precipitation. Because pollution observed at a site is likely driven by emissions elsewhere that also affect nearby sites, all inferences allow for correlations in errors across different monitoring sites within the same county, clustering standard errors at the county level. The results are robust to a range of specification changes (discussed in detail in Section IID).

[10] Online Appendix Table D.2 reports the odds of any exceedance over the course of a year in my sample (50.9 percent for 1-in-1-day sites, 33.4 percent for 1-in-3-day sites, and 29.5 percent for 1-in-6-day sites), among other related statistics.

For $\hat{\beta}$ to be the causal estimate of the monitoring schedule's impact, the identification assumption must hold that no pollution gap would have been observed in the absence of the monitoring schedule. While this assumption is not directly testable, in Section IID I implement placebo tests based on the idea that a strategic response is not expected to occur in areas that lack incentives to reduce pollution on monitored days, such as in regions in which monitors sample air quality every day.

The "Strategic Response" Interpretation.—A reduction in polluting activities when enforcement is stronger does not necessarily indicate *strategic* responses. For example, a permanent reduction in pollution levels after the initiation of monitoring can arise from compliance behavior. However, the fact that pollution levels differ across monitored and unmonitored days in places that do have a monitoring system in place indicates that the nature of the response is strategic. In this paper, I interpret $\hat{\beta}$ as an estimate of strategic behavior, or "gaming" of the monitoring schedule.

A related point is that β represents the causal effect of intermittent monitoring on the *differences* in pollution levels on off-days relative to on-days. The empirical framework, however, does not identify the effect of intermittent monitoring on the *levels* of pollution. Section IIIA analyzes changes in pollution levels, which could be important when thinking about policy counterfactuals and mechanisms.

C. *Main Results*

I begin by estimating an event study version of equation (1), replacing the $\mathbf{1}(Offdays_t)$ indicator variable with five event-day indicator variables running from three days before to two days after the on-day. The on-day, marked as day 0, is the omitted category in the regression.[11] Therefore, the coefficients are interpreted as percentage changes in air pollution during the off-days relative to the on-days. I estimate the event study using all monitoring sites containing at least one 1-in-6-day PM monitor during the year. The sample includes 1,193 monitoring sites that span 563 counties in the lower 48 states from 2001 to 2013. Figure 2 reports the results. I do not condition the regression on any covariates; the solid line thus simply represents the time path of air pollution in a 1-in-6-day monitoring cycle, averaged across all cycles. The results reveal a striking pattern of air quality differences: within a monitoring cycle, air pollution on average exhibits a flat path, except for a sharp drop-off during the on-day. Table 2 reports the average 1-in-6-day off-/on-day pollution gap using equation (1). Results in column 1 correspond to Figure 2 and show that air pollution is on average 1.6 percent higher on an off-day relative to an on-day. Column 2 reports that adding the full set of controls does not change the estimates.

Columns 3 and 4 of Table 2 perform the same analysis for 1-in-3-day sites. Note in this case $\mathbf{1}(Offdays_t)$ indicates unmonitored days in a 3-day monitoring cycle. I find that for 1-in-3-day PM sites, the pollution gap is less than 0.3 percent and not statistically significant. The estimates have enough precision to reject an effect size

[11] Due to the cyclical nature of monitoring, the choice of the omitted category is inconsequential for estimation. For example, using "day −3" as the omitted category would simply cause a parallel downward shift on all event days, with no effect on the pollution gap estimate.

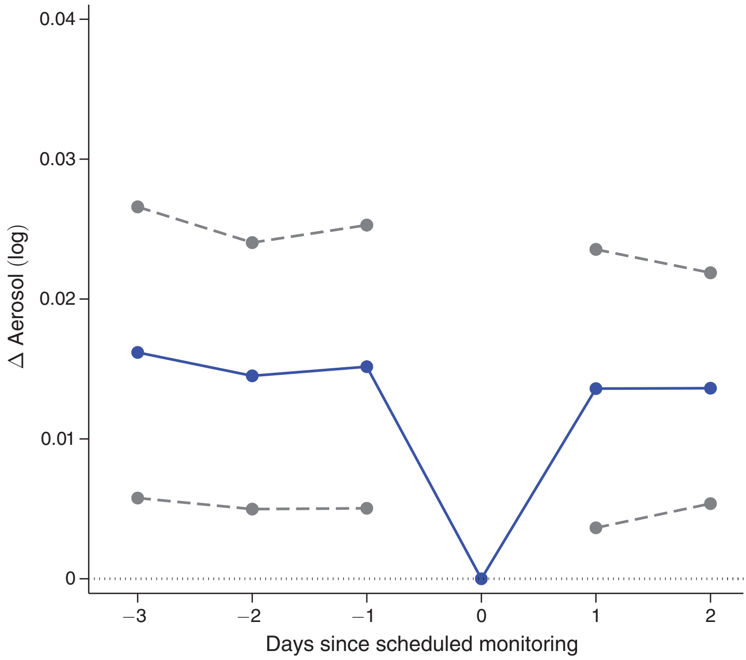

FIGURE 2. THE POLLUTION GAP: OFF-DAYS VERSUS ON-DAYS POLLUTION DIFFERENCES

Notes: This figure plots the path of pollution concentration by days of 1-in-6-day monitoring cycle. The sample includes all sites that contain at least one 1-in-6-day PM monitor. Pollution is measured by satellite-based aerosol concentration within the 10km × 10km area that contains the monitoring site. Day 0 corresponds to the scheduled monitoring day, which is normalized to 0. The regression is not conditional on any covariates. Dashed lines represent 95 percent confidence intervals constructed using standard errors clustered at the county level.

TABLE 2—REGRESSION ESTIMATES OF THE POLLUTION GAP

Monitoring frequency:	1-in-6-day		1-in-3-day		1-in-1-day	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: aerosol concentration (log)						
1(*Offdays*)	0.0160	0.0162	0.0028	0.0029	−0.0013	0.0023
	(0.0040)	(0.0035)	(0.0026)	(0.0020)	(0.0107)	(0.0080)
Controls		✓		✓		✓
Sites	1,193	1,193	1,064	1,064	556	556
Observations	685,060	685,060	598,859	598,859	231,532	231,532

Notes: Each column reports a separate regression. Estimation samples are sites with any 1-in-6-day monitors (columns 1 and 2), sites with any 1-in-3-day monitors (columns 3 and 4), and sites with any 1-in-1-day monitors and no collocated intermittent monitors (columns 5 and 6). **1**(*Offdays*) indicates days when PM monitoring is not scheduled. *Controls* include site, year, month-of-year, and day-of-week fixed effects, and weather covariates (Section IIB). Standard errors are clusterd at the county level.

that is one-half of that observed around 1-in-6-day site. The null effect at 1-in-3-day carries a potentially important policy implication: doubling the monitoring frequency may be a feasible policy option to substantially reduce strategic responses. One caveat to this conclusion is that, while strategic responses near 1-in-3-day sites appear smaller in magnitude, there is evidence that a 1-in-3-day monitoring frequency is still "gameable." For example, in unreported analysis, I find a significant response can be detected around sites that are far away from every-day monitors.

D. *Placebo Tests and Robustness Checks*

The identification assumption states that no pollution gap would have been observed in the absence of the 1-in-6-day monitoring schedule. To assess this assumption, I consider two types of placebo tests in situations where gaming is not expected.

In the first test, I estimate the 1-in-6-day pollution gap near every-day sites (1-in-1-day sites without co-location of intermittent monitors). Columns 5 and 6 of Table 2 document the "pollution gap" at these sites. Reassuringly, I find no evidence of differential levels of pollution across off-days and on-days. In online Appendix Table D.3, I further examine over 800 hazardous air pollutants (HAPs) monitoring sites that mostly follow the 1-in-6-day schedule, but are not subject to any regulatory standards. Again, I find no evidence of pollution gaps around these "unregulated" monitors.

The second placebo test explores the retirement of 1-in-6-day monitoring sites. If gaming is indeed targeting the 1-in-6-day schedule, then one should expect the pollution gap to "close" after sites are removed. I begin by identifying 490 cases of 1-in-6-day monitoring site retirement events. Using the satellite measure to track air quality in the areas where these sites are located, I compare the "off-days" an "on-day" pollution gap before and after sites' retirement. Note that I can estimate the pollution gap even after a site retires because the monitoring calendar applies universally; hence, even in the absence of a working monitoring site, I know what the monitoring dates would have been. Figure 3 reports the results, showing the pollution gap as a function of years relative to sites' retirement. The gap is about 2.1 percent for the time frame in which the site operated. For the exact same area, the gap closes after monitor retires.

Online Appendix Table D.4 reports additional robustness checks. I show that the results are robust to the inclusion of more flexible fixed effects controls, such as monitoring site-specific time controls, as well as a difference-in-differences design that uses every-day sites as the control group.

III. Characteristics of the Pollution Gap

A. *Changes in Levels of Pollution*

Section II focuses on *differences* in pollution on off-days relative to on-days. It is also be of interest to understand the effect of intermittent monitoring on the *levels* of pollution. The pollution gap could be explained by a pure reduction of pollution on the on-days, a pure substitution of polluting activities between off-days and on-days, or somewhere in between. Such distinctions are important for two reasons. First, it is relevant for the regulator to know the counterfactual levels of pollution in the absence of intermittent monitoring. Second, information on level changes may shed light on the underlying nature of strategic behavior, as discussed further below.

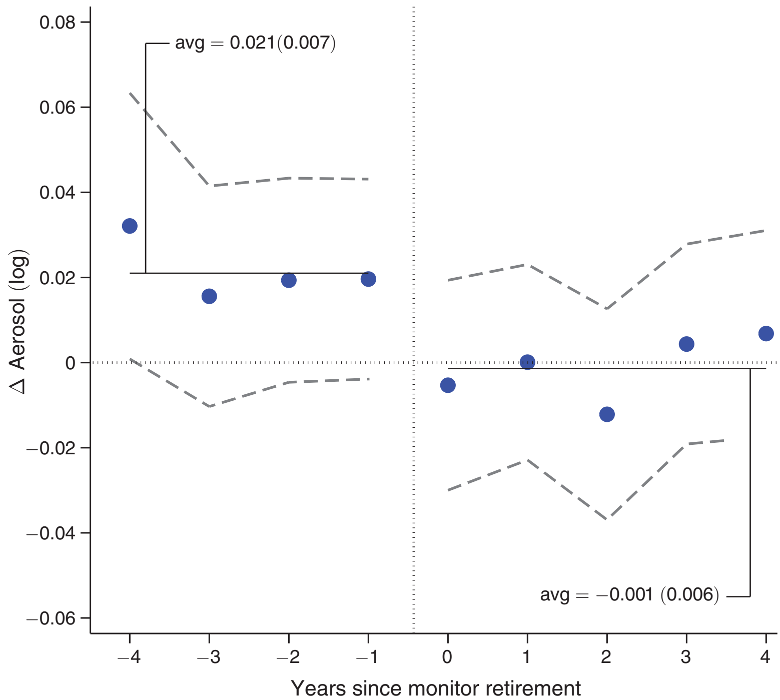

Figure 3. Pollution Gap by Years Relative to Monitoring Site Retirement

Notes: This figure plots the pollution gap between monitored (on-days) and unmonitored days (off-days) as a function of years relative to the retirement of a site that conducted monitoring every six days. Average estimates show off-days effect separately estimated before and after the retirement of the monitoring site. The sample includes retirement of 490 sites from 2001 to 2013. The regression is not conditional on any covariates. Dashed lines represent 95 percent confidence intervals constructed using standard errors clustered at the county level.

The key empirical challenge in estimating level changes is that one cannot observe the same monitor under counterfactual monitoring frequencies. Instead, I exploit the monitoring retirement events (Section IID) again. I examine changes in local-area pollution levels overall, as well as separately for on- and off-days, before and after the 1-in-6-day monitor retirements. To allow for trends in pollution that might correlate with monitor retirement (for example, unobserved factors that might drive monitor retirement in the first place), I use pollution levels around *non-retiring* 1-in-6-day sites in the same state as the control group. For each 1-in-6-day site that retires in year t, I match the site with all 1-in-6-day monitors in the same state that are not retiring in year t. An average retiring site is matched to 31 non-retiring sites in the same state.

Figure 4 provides an initial look at the raw trends. In panel A, the solid (dashed) lines with markers show pollution levels on off-days (on-days) around 1-in-6-day sites by years relative to retirement. In panel B, the solid (dashed) lines without markers show pollution levels on off-days (on-days) around non-retiring 1-in-6-day sites. Three sets of difference-in-differences-style evidence emerge. First, pollution levels around retiring sites rise during the pre-treatment period (event years −4 to −1), suggesting that the timing of retirement is not exogenous. However, non-retiring sites

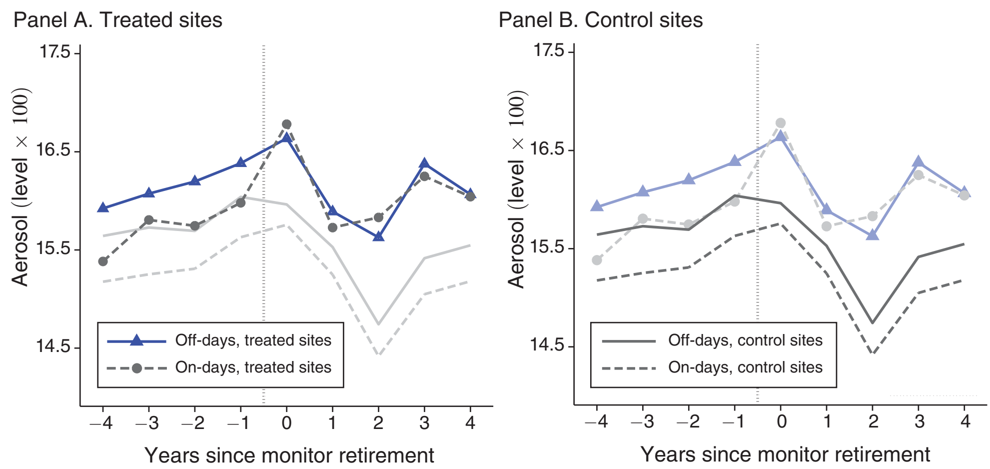

Figure 4. Pollution Levels by Years Relative to Monitoring Site Retirement

Notes: This figure plots the pollution levels on monitored (on-days) and unmonitored days (off-days). Trends are plotted for treated sites (retired 1-in-6-day sites) in panel A and for control sites (non-retired 1-in-6-day site within the same state as the retired site) in panel B. Event year 0 is the year when the corresponding treated site retires.

exhibit a similar trend in pollution *and* a similar gap in pollution across off-days and on-days. The parallel trends, for both on-days and off-days, make the non-retiring sites a reasonable control group. Second, starting at the retirement year and throughout the post-treatment period (event years 0 to 4), the pollution gap at the retiring sites closes, while the gap at the non-retiring sites remains open. This evidence provides further support that monitor retirement removes strategic response only at the treated sites (Section IID). Third, throughout the entire study window (event years -4 to 4), all four lines, indicating on-day and off-day pollution level at retiring and non-retiring sites, move in fairly parallel manner; the lone exception is the permanent jump that occurs at event year 0, showing the surge of pollution levels at retired sites on what were previously on-days. Altogether, evidence suggests that the retirement of intermittent monitoring is associated with a rise of pollution on what had previously been on-days, and little effect on levels of pollution on what had previously been off-days.

The graphical analysis motivates a difference-in-differences strategy to identify the effect of intermittent monitoring on pollution levels. I fit the following estimation equation:

$$(2) \qquad Aerosol_{st} = \beta \cdot \mathbf{1}(treat_s) \cdot \mathbf{1}(after_y) + \delta \cdot \mathbf{1}(treat_s) + \theta \cdot \mathbf{1}(after_y) + \eta_{group} + Time_t + \alpha_s + X_{st}\gamma + \varepsilon_{st},$$

where $\mathbf{1}(treat_s)$ indicates retiring sites, $\mathbf{1}(after_y)$ indicates post-retirement years. Note a non-retiring site can serve as the control to multiple retiring sites, thus it

TABLE 3—THE EFFECT OF MONITORING SITE RETIREMENT ON POLLUTION LEVELS

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: aerosol concentration (*level* × *100*)						
1(*treat*) × **1**(*after*)	0.137		0.178		0.154	
	(0.096)		(0.094)		(0.173)	
1(*treat*) × **1**(*after*) × **1**(*Ondays*)		0.366		0.374		0.350
		(0.144)		(0.141)		(0.199)
1(*treat*) × **1**(after) × **1**(*Offdays*)		0.090		0.138		0.115
		(0.097)		(0.095)		(0.174)
Group fixed effects	✓	✓	✓	✓	✓	✓
Controls			✓	✓	✓	✓
Trend break model					✓	✓
Mean dependent variable	15.7	15.7	15.7	15.7	15.7	15.7
Observations	9,854,996	9,854,996	9,854,996	9,854,996	9,854,996	9,854,996

Notes: Each column reports a separate regression. **1**(*treat*) indicates retired sites. **1**(*after*) indicates years after monitoring retirement. **1**(*Offdays*) indicates days when PM monitoring is not scheduled. **1**(*Ondays*) indicates days when PM monitoring is scheduled. *Group fixed effects* indicate a retired site with all its matched non-retired sites. *Controls* include site, year, month-of-year, and day-of-week fixed effects, and weather covariates (Section IIB). *Trend break model* include additional terms on linear event time trends interacted with the treatment dummy and the after dummy, separately. All regressions control for lower-order interaction terms and main effect terms. Standard errors are clustered at the county level.

may appear multiple times in the estimation sample. I include group fixed effects (η_{group}) to conduct comparison between retiring site and its matched non-retiring sites. The rest of the specifications are the same with equation (1). For each site, I apply weights derived using the synthetic control method so that the weighted average control site closely tracks the trends in retiring sites' pre-treatment pollution levels (Abadie, Diamond, and Hainmueller 2010).[12]

Table 3 presents the difference-in-differences estimates. Beginning with column 1, note that the coefficient suggests that the level of aerosol pollution around retiring sites increases by 0.137 unit (relative to synthetic controls) from a baseline aerosol of 15.7 units, thus representing a 0.87 percent increase. The point estimate is not statistically significant at the conventional level. Perhaps not surprisingly, it is difficult to capture an overall increase in pollution driven by an increase in pollution on the on-days, which represent only one-sixth of the sample. Column 2 decomposes this increase into on-days and off-days, and finds a significant increase on the on-days. Columns 3 and 4 repeat the same analysis, adding control variables of the preferred specification: county, year, month-of-year, day-of-week fixed effects, and weather controls. In columns 4 and 5, I estimate a "trend break" model in which I include a flexible set of event-time trends for treatment and control sites, before and after treatment year. In the spirit of a parametric regression discontinuity design, the trend break model allows for flexible time trends but detects changes from a break in the trends that occur exactly at the retirement year. Overall, results are more precise with models with simpler controls, but the qualitative findings remain robust.

[12] Online Appendix Section C.1 documents full details on the synthetic control diagnosis statistics. In particular, I compute synthetic weights based on data in periods $t = -4$ and $t = -3$ only, leaving $t = -2$ and $t = -1$ out to serve as a validation sample of the method's performance.

Evidence suggests that the pollution gap is likely driven by a fall of pollution levels on the on-days. This finding has two potential implications on the characteristics of strategic responses. First, it suggests that, in the absence of strategic behavior, the counterfactual levels of pollution are likely close to those of off-days. Second, to the extent that there is little evidence of pollution substitution across on-days and off-days, the evidence indicates that short-term responses occur only occasionally, for example, taking place during periods when a suppression of polluting activities can alleviate risk of noncompliance, rather than through planned, routine responses that occur often. In the next subsections, I present examples of such "critical" periods when strategic responses might be occurring.

B. *High Pollution Periods*

A natural example of a "critical" time is when the county's nonattainment hazard is raised. I begin with a simple heterogeneity analysis that examines whether the pollution gap is larger when high levels of PM2.5 have been recorded in the county's recent past. I augment equation (1) by interacting the $\mathbf{1}(Offdays_t)$ dummy variable with average PM2.5 in the past six months (i.e., six lags of monthly PM2.5 in the county). As a placebo-style exercise, I also include interaction terms with six leads of monthly PM2.5. The estimation equation is

$$(3)\quad Aerosol_{st} = \sum_{\substack{m\in[-6,6]\\ m\neq 0}} \beta_m \cdot \mathbf{1}\big(Offdays_t\big) \cdot PM2.5_{cm} + \sum_{\substack{m\in[-6,6]\\ m\neq 0}} \delta_m \cdot PM2.5_{cm} + \theta \cdot \mathbf{1}\big(Offdays_t\big) + Time_t + \alpha_s + X_{st}\gamma + \varepsilon_{st}\,,$$

where c indicates county that contains the monitoring site s. The key coefficients β_m represent changes in the magnitude of pollution gap per unit increase of PM2.5 in relative month m. More intuitively, β_m coefficients capture the correlation between pollution gap today and realized PM2.5 in month m relative to today.

Figure 5 plots the $\hat{\beta}_m$ estimates. The chart features a strong correlation between the pollution gap and realized PM2.5 in the previous month. This correlation is not detected in earlier or later periods, either just one month earlier or in any future months.[13] In online Appendix Figure D.1, I report additional evidence that the correlation between strategic response and previous-month PM2.5 is mainly driven by months in which the PM2.5 level is above $10\,\mathrm{ug/m^3}$. I find that the pollution gap exceeds 7 percent when the previous month's PM2.5 exceeded the annual regulatory standard value of $15\,\mathrm{ug/m^3}$. In contrast, no pollution gap is detected at times when the pollution level is far below the standard.

The heterogeneity analysis leads to several expected and unexpected implications. Consistent with prior literature, evidence suggests that resources are allocated to achieve localized air quality improvements, in my setting, reduced pollution on monitored days, that would help jurisdiction's NAAQS compliance. At the same

[13] Note that this is different from mean-reverting, as the result indicates a larger percentage *difference* in off-day versus on-day pollution levels following a high level of pollution realization.

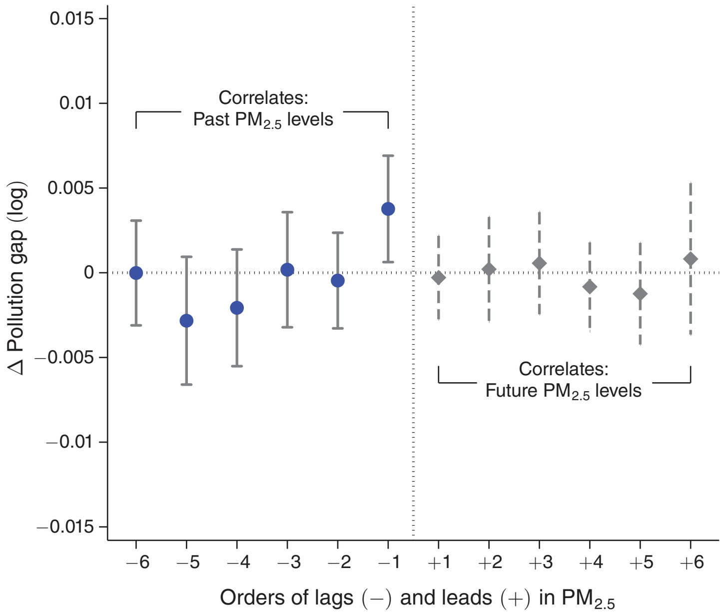

Figure 5. Heterogeneous Pollution Gap by County's Realized PM2.5 in the Past and Future

Notes: This figure reports the regression coefficients on the interaction of 1-in-6-day pollution gap with six lags and six leads of the county's monthly average PM2.5 concentration. The underlying regressions include main effect terms, site, year, month-of-year, and day-of-week fixed effects, and weather controls (Section IIB). Range bars represent 95 percent confidence intervals constructed using standard errors clustered at the county level.

time, however, the short-term nature of the strategic response indicates that the magnitude of the pollution gap depends on recorded pollution in the immediate past, which in itself reveals the existence of the capacity to game the schedule. The evidence casts some initial light on the question of who is gaming. Gaming is unlikely to arise from factories' self-coordinated, cooperative emission behavior, especially in regions with many polluters. Factories also typically do not have ready access to air monitoring results from the previous month. From a plausibility perspective, local government agencies, who monitor, analyze, and forecast air quality on a day-to-day basis, are better informed about when "critical" air quality periods surface the local areas. In the next subsection, I examine the potential role of local governments in strategic responses.

C. *Air Pollution "Action Day" Advisories*

I assess a rare situation that provides an opportunity to directly observe local government coordination actions. I examine the pattern of issuances of air pollution "Action Day" advisories, which many local air pollution control agencies use to provide warnings to the public when air pollution is expected to reach unhealthy levels. On an Action Day, governments urge citizens to help improve air quality by "taking actions," such as reducing energy and automobile use.

Two features make the Action Day program an opportunity to study local government coordination. First, Action Day advisories significantly affect outdoor activities and transportation decisions, as shown by previous research (Cutter and Neidell 2009, Graff Zivin and Neidell 2009, Neidell 2009). For example, Cutter and Neidell (2009) find that San Francisco Bay Area's "Spare the Air" warning causes 3–3.5 percent reduction in daily traffic.[14] Second, local agencies have discretion about when to issue warnings. Warnings are based on real-time air quality forecasts. As shown in online Appendix Figure D.2, many warnings occurred when forecasted air quality index (AQI) crosses the "yellow-to-orange" line (forecasted AQI = 100) or the "orange-to-red" line (forecasted AQI = 150). However, warnings are also observed across a wide range of AQIs, which suggests that the issuance decision is not formulaic, or determination made by a matter-of-fact calculation. I examine whether Action Day advisories are more likely to be issued on days when PM monitoring is scheduled, with the presumption that warnings help pollution fall.[15]

I obtain a comprehensive database of Action Day records from 346 reporting areas from 2004 to 2013. This database is maintained by the EPA's AirNow program (airnow.gov). The reporting areas are a mix of cities, counties, metro areas, and states that together cover roughly 50 percent of the US population. To avoid double counting jurisdictions, I aggregate the data to the core-based metro area (CBSA) level. Action Day warnings often span multiple days (an average episode lasts 2.4 days). Thus, to better target issuance timing I focus on the first day of warning for any period with consecutive Action Day warnings. The final estimation sample includes 6,232 Action Days at the CBSA $\times$ daily level. I test whether Action Day warnings are more likely to occur on monitored days again using equation (1), but changing the outcome variable to an indicator of whether an Action Day advisory is issued on the CBSA $\times$ day.

Figure 6 summarizes the main result. The graphical pattern provides evidence of a significant excess of advisories on pollution sampling days. On average, an Action Day is 0.108 percentage points more likely to be issued on an on-day than on an off-day. Given an average daily issuance probability of roughly 1 percent, this effect represents a 10 percent difference in the odds of issuance. Put differently, out of the 6,232 warnings in the data, monitoring avoidance influences about 100 warning issuances. Online Appendix Section C.2 presents a state-level heterogeneity analysis, showing that strategic Action Day warnings more often surface in states with specific political characteristics, such as a history of receiving and contesting the EPA's previous nonattainment designation decisions. Overall, evidence points to the role of "administrative capacity" underlying states' strategic Action Day warnings, in addition to their bureaucratic incentives to avoid NAAQS violation.

The Action Day provides another example of gaming in critical, high-pollution periods. The analysis also suggests that gaming need not arise from factories' individual behavior. The Action Day analysis raises the possibility that local government

[14] In unreported analysis, I have extended the Cutter and Neidell (2009) findings, linking daily traffic volume to Action Day warnings using all CBSAs with available data. I find that an Action Day warning is associated with a robust and statistically significant 1.5–3 percent reduction in daily traffic.

[15] Online Appendix Table D.5 reports that monitoring cycles that contain Action Day warnings feature a pollution gap of 5 percent to 7 percent, compared to the average pollution gap of 1.6 percent.

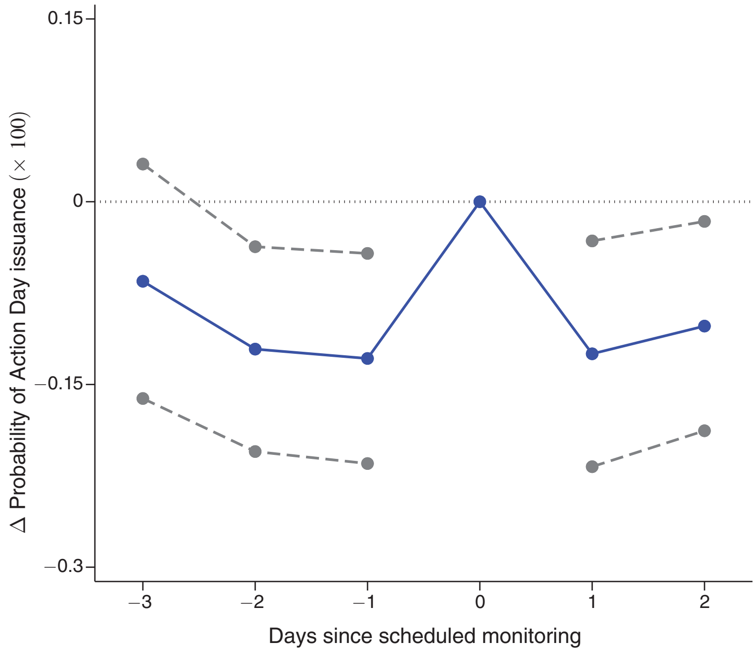

FIGURE 6. EVENT STUDY: POLLUTION "ACTION DAY" WARNINGS ON OFF-DAYS VERSUS ON-DAYS

Notes: This figure plots the probability of local governments' issuance of pollution "Action Day" advisories by days of 1-in-6-day monitoring cycle. The outcome variable is a core-based statistical area (CBSA) × daily dummy for whether any Action Day is issued. In cases of issuances that span a consecutive number of days, only the first day of issuance is counted. The estimation sample spans 2004–2013 and includes 14,945 issuances across 171 CBSAs. Day 0 corresponds to the scheduled monitoring day, which is normalized to 0. The regression is not conditional on any covariates. Dashed lines represent 95 percent confidence intervals constructed using standard errors clustered at the CBSA level.

coordination might occur in industrial settings as well. While it is difficult to directly measure coordination through industrial activities, in the next subsection, I leverage rich geographic variations in the observed effect sizes to document industrial characteristics of areas with large pollution gaps.

D. *Industrial Correlates*

County-Level Estimates of Pollution Gaps.—I begin by estimating off-day and on-day pollution gaps separately for each county. For a county c, the following estimation equation is fitted:

$$Aerosol_{ict} = \beta_c \cdot \mathbf{1}(Offdays_t) + Time_{ct} + \varepsilon_{ict} \tag{4}$$

where $Aerosol_{ict}$ denotes logged aerosol level in grid i inside county c on date t. Seasonality controls $Time_{ct}$ include year, month-of-year, and day-of-week dummies. The average county-level regression contains 35,236 observations (median = 21,086 observations) at the 10 km × 10 km-by-daily level. To account for sampling variation, especially for small counties that contain fewer grid cell

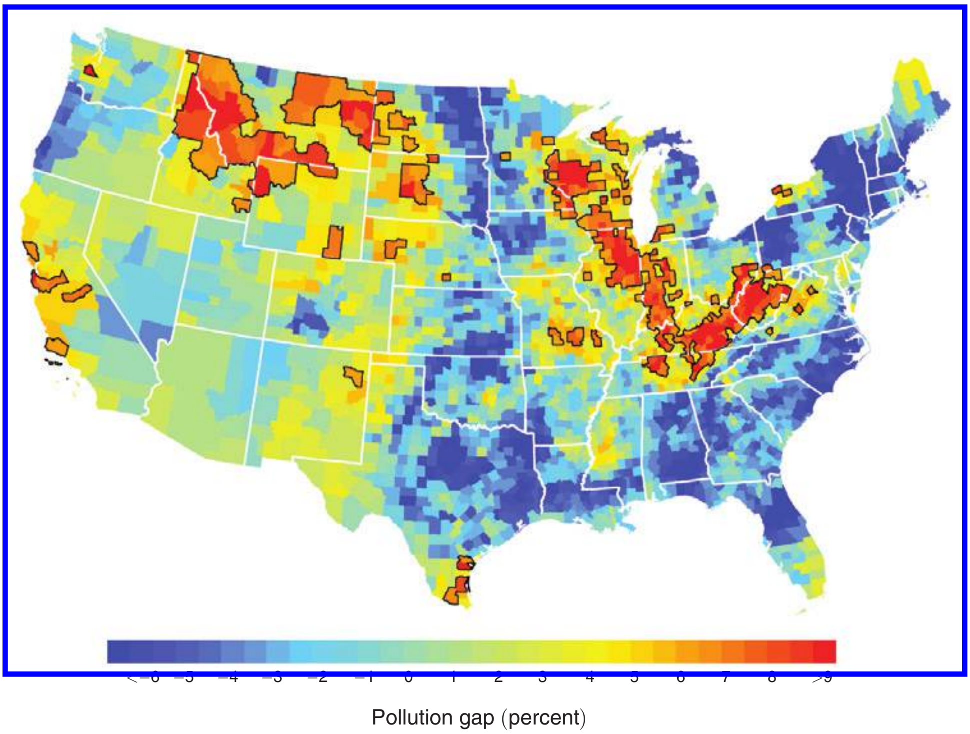

Figure 7. County-Level 1-in-6-Day Pollution Gap Estimates

Notes: This map plots Empirical Bayes-adjusted county-level 1-in-6-day pollution gap estimates. Red (blue) areas correspond to higher (lower) aerosol levels on off-days relative to on-days. For all counties, off-days and on-days are defined using EPA's monitoring schedule. Counties with outlines are those with the highest 10 percent pollution gap estimates ("hotspots").

observations, I adjust the $\hat{\beta}_c$ estimates using an Empirical Bayes-style procedure (Chetty and Hendren 2018; Finkelstein, Gentzkow, and Williams 2019). This procedure generates an MSE-minimizing prediction of the true β_c through a combination of the (unbiased but high variance) raw pollution gap estimate $\hat{\beta}_c$ and a (biased but low-variance) "regionwide" estimate that uses data from all grid cells within 150 miles of the county of interest.[16] Online Appendix Section C.3 provides full details on the estimation.

Figure 7 plots the county-level Empirical Bayes-adjusted estimates of the pollution gap. The map is drawn so that warmer colors indicate areas with larger pollution gaps. The map exhibits two features. First, large pollution gaps surface in several area clusters, such as parts of California, Montana, southern Texas, and a group of states in the Midwest. In the following analysis, I define pollution gap "hotspot" counties as those with a top-decile pollution gap estimate. Conclusions are not sensitive to alternative definitions of hot spots, such as defining hot spots as counties with pollution gaps in the top quintile. Second, some counties exhibit a *negative*

[16] The 150-mile radius is chosen to approximate the size of an average state. I do not use actual state boundaries to avoid substantial difference in the size of states. The average region-level pollution gap estimation contains 2.2 million observations.

pollution gap, which indicates that aerosol levels are lower during off-days in comparison to on-days. In online Appendix Section C.3, I provide an analysis that suggests that the spatial pattern of the positive and negative pollution gap is consistent with wind transport dynamics. In the analysis below, I focus primarily on characteristics of hot spots, and set aside the variation in the magnitude of the pollution gap, because a gap's magnitude is more likely to reflect a complex mix of local strategic responses and a blown-in pollution pattern from neighboring areas.

Correlates of the Pollution Gap Hot Spots.—I estimate a cross-sectional regression of the indicator variable for a hotspot county to a host of county-level characteristics. A natural beginning point is a county's regulatory and monitoring status. I include an indicator for the presence of 1-in-6-day monitoring sites, whether the county has ever been assigned a PM-related nonattainment status, and a continuous variable of the county's average aerosol level. I then include a series of dummy variables for whether a county has a high (top 10 percent) employment share of each of the 34 polluting industries.[17] I then estimate both multivariate OLS and post-LASSO regressions that link pollution hotspot counties to the regulatory and industrial characteristics.[18]

Figure 8 summarizes the coefficient estimates. The top panel shows that large pollution gaps tend to emerge in counties that have high levels of aerosol levels, have experienced nonattainment, and are subject to 1-in-6-day monitoring. These characteristics turn out to be the strongest correlates, both in terms of magnitude and precision, of large pollution gaps. Going down the panel, I organize polluting industries into sector blocks, ranked from sectors that emitted the highest share of PM (Utilities) to the lowest (Administrative Services, and Waste Management, and Remediation Services). Two specific polluting industries, wood product manufacturing and mining, emerge as significant and positive correlates of large pollution gaps. The post-LASSO estimation results largely agree with the OLS findings. Industries with low emission contributions tend not to be selected as relevant predictors; on the other hand, coefficient estimates remain largely unchanged for industries that are recognized by the OLS as consistent and significant correlates. In unreported analysis, I find that the correlation of wood product manufacturing is particularly robust across a wide range of specifications.

A potential explanation of this heterogeneity relates to the capacity to respond. For example, while the utility sector is a major PM emitter, power plants, especially coal plants, often run around the clock; for them, ramping production up or down in response to a short-term monitoring schedule would be very costly. By contrast, wood product manufacturers also contribute to a significant portion (3 percent) of total PM emissions among point sources, but they usually run at low capacity. In fact, wood plants, which operate at roughly 60 percent of capacity, have the lowest capacity utilization rate among all polluting manufacturers examined in this study.

[17] I define the polluting sector as all two-digit NAICS industries that contribute to at least 1 percent of national total PM10 emissions in the EPA's 2011 National Emissions Inventory. For each three-digit NAICS industry of the polluting sector, I use Census County Business Pattern to compute its employment share in each county, and flag counties in the top decile of the distribution.

[18] The post-LASSO regression is the same OLS regression where the right-hand-side variables are restricted to those selected by a first-step LASSO estimation.

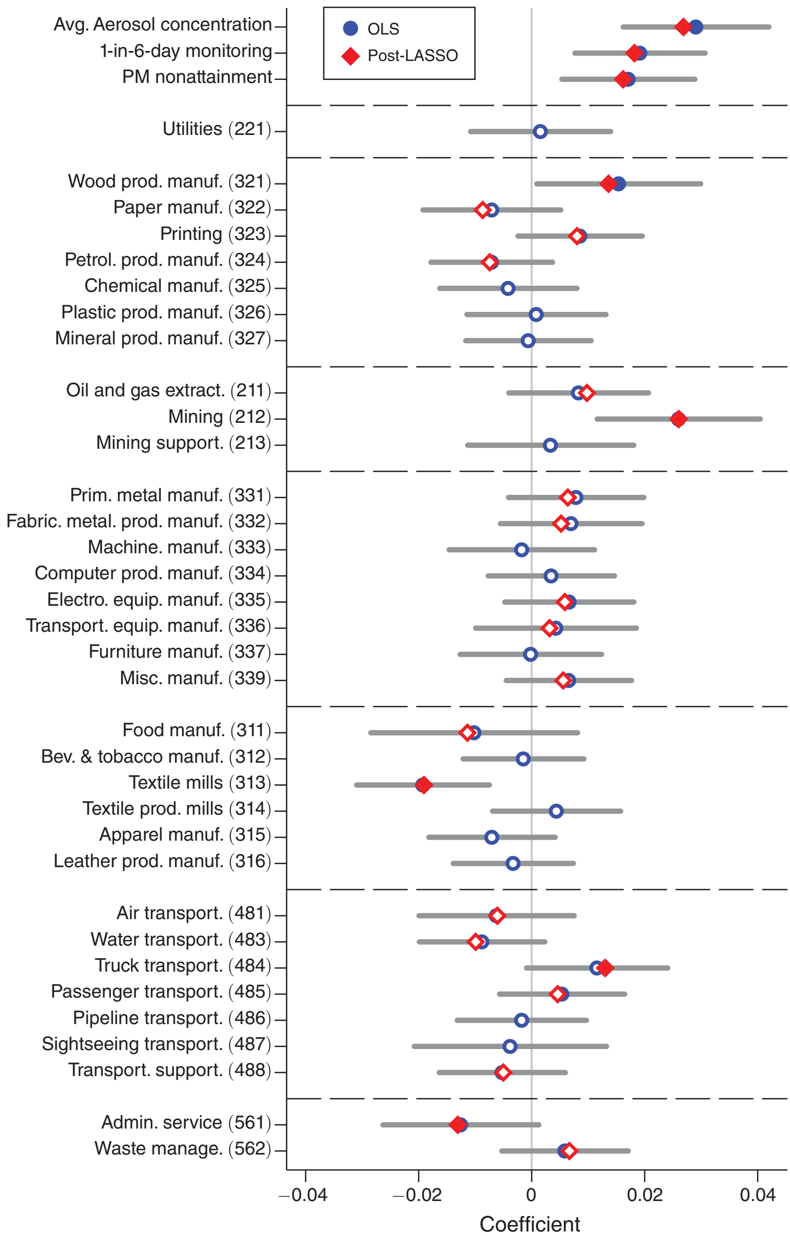

FIGURE 8. REGULATORY AND INDUSTRIAL CHARACTERISTICS OF POLLUTION GAP HOTSPOTS

Notes: This figure plots coefficients from two regressions: (i) a cross-section OLS regression of an indicator for pollution gap hotspots (counties with highest 10 percent pollution gap estimates) on county-level regulatory (rows 1 to 3) and industrial characteristics (row 4 onward), and (ii) the post-LASSO version of the OLS regression in (1). Two-digit NAICS industry blocks are ordered by its share of national total PM10 emissions according to EPA's 2011 National Emissions Inventory. In the interest of space, the graph only reports coefficients for polluting sectors that contribute at least 1 percent of total PM10. Solid dots indicate coefficients that are individually significant at the 5 percent level. Range bars represent 95 percent confidence intervals from the OLS regression.

The unused capacity might enable these plants to shift production activities around in avoidance of the pollution monitoring schedule.

IV. Conclusion

Fiscal constraints often motivate environmental regulators to monitor polluting behavior on an intermittent basis. A largely overlooked issue with intermittent monitoring is its vulnerability to polluters' strategic responses, as highlighted by recent evidence on the vehicle emission scandals. This paper reinforces these recent findings and extends the literature to a broader setting of ambient air quality regulations. I have presented evidence that a widely used once-every-six-day monitoring schedule for outdoor particulate pollution leads to significantly poorer air quality on unmonitored days than on the days that the regulator observes through the monitoring. The combined evidence, based on satellite measure of air quality, and analysis of various characteristics of regions with wider pollution gaps between monitored and unmonitored days, shows that gaming of the monitoring schedule can and does occur. My results reveal the possibility that strategic responses may arise through coordination by local governments, which, under the federal regulations, share noncompliance costs with industries. As an illustration of this mechanism, I show that some local governments issue air quality advisories that are strategically timed in response to the monitoring schedule; such actions are likely intended to manipulate public behavior, such as transportation decisions that affect pollution emissions, to reduce pollution levels recorded on monitored days. The evidence shows that the intensity of strategic responses differs substantially across geographic regions, depending strongly on the characteristics of local area economies and industries. The results are consistent with the view that strategic responses are more likely to arise in regions with greater incentives and/or capacity to avoid monitoring. Thus, the likelihood of gaming behavior is greater in regions that risk noncompliance penalties, and in areas with high-emission industries that often operate below capacity, a feature that makes it possible for them to make adjustments timed with the monitoring schedule.

The results raise several possible policy solutions. A natural possible fix would be to use a random monitoring schedule. Though a randomized schedule (with short notice to state and local governments) has conceptual appeal, such a monitoring system may be problematic in practice because filter-based PM methods require intensive field work involving sampling and subsequent laboratory analyses. One might instead consider using an ex ante generated random schedule set, for example, at the beginning of a calendar year. In light of this paper's findings on local government coordination, however, it is unclear whether the ex ante schedule would be strictly hidden from local polluters. However, randomization is not the only potential solution. My results suggest that increased monitoring frequency reduces gaming and improves air quality. In fact, evidence shows that strategic responses are much weaker in regions under schedules that impose monitoring once every third day, rather than once every sixth day. Though a comprehensive evaluation comparing total health benefits with the administrative costs of increased monitoring frequency is beyond the scope of this paper, my findings do support two related, emerging themes in environmental policymaking. First, advanced continuous monitoring technologies should be promoted to replace discrete, sampling-based monitoring of the environment (e.g., Giles 2013). Second, retrospective and independent evaluations, in addition to ex ante cost-benefit calculations, are important to ensure the integrity and efficacy of regulation design (Greenstone 2013; Auffhammer 2015; Cropper, Fraas, and Morgenstern 2017).

REFERENCES

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490): 493–505.

Agarwal, Sumit, Yu Qin, and Hongjia Zhu. 2020. "Disguised Pollution: Industrial Activities in the Dark." Unpublished.

AirNow and Sonoma Technology. 2004–2015. "AirNow Action Day." https://www.airnow.gov (accessed June 8, 2015).

Akland, Gerald G. 1972. "Design of Sampling Schedules." *Journal of the Air Pollution Control Association* 22 (4): 264–66.

Alexander, Diane, and Hannes Schwandt. 2019. "The Impact of Car Pollution on Infant and Child Health: Evidence from Emissions Cheating." Unpublished.

Auffhammer, Maximilian. 2015. "VW's Deepwater Horizon?" UC Berkeley Energy Institute Blog.

Auffhammer, Maximilian, Antonio M. Bento, and Scott E. Lowe. 2009. "Measuring the Effects of the Clean Air Act Amendments on Ambient PM 10 Concentrations: The Critical Importance of a Spatially Disaggregated Analysis." *Journal of Environmental Economics and Management* 58 (1): 15–26.

Becker, Randy, and Vernon Henderson. 2000. "Effects of Air Quality Regulations on Polluting Industries." *Journal of Political Economy* 108 (2): 379–421.

Bento, Antonio, Matthew Freedman, and Corey Lang. 2015. "Who Benefits from Environmental Regulation? Evidence from the Clean Air Act Amendments." *Review of Economics and Statistics* 97 (3): 610–22.

Blundell, Wesley, Gautam Gowrisankaran, and Ashley Langer. 2018. "Escalation of Scrutiny: The Gains from Dynamic Enforcement of Environmental Regulations." NBER Working Paper 24810.

Chetty, Raj, and Nathaniel Hendren. 2018. "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates." *Quarterly Journal of Economics* 133 (3): 1163–228.

County Business Patterns. 2001–2013. "Complete County File." https://www.census.gov/programs-surveys/cbp/data/datasets.html (accessed August 21, 2019).

Cropper, Maureen, Arthur Fraas, and Richard Morgenstern. 2017. "Looking Backward to Move Regulations Forward." *Science* 355 (6332): 1375–76.

Cutter, W. Bowman, and Matthew Neidell. 2009. "Voluntary Information Programs and Environmental Regulation: Evidence from 'Spare the Air.'" *Journal of Environmental Economics and Management* 58 (3): 253–65.

Donaldson, Dave, and Adam Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives* 30 (4): 171–98.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2013. "Truth-Telling by Third-Party Auditors and the Response of Polluting Firms: Experimental Evidence from India." *Quarterly Journal of Economics* 128 (4): 1499–545.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2018. "The Value of Regulatory Discretion: Estimates from Environmental Inspections in India." *Econometrica* 86 (6): 2123–60.

Esri. 2010. "U.S. Major Highways." https://www.arcgis.com/home/item.html?id=fc870766a3994111bce4a083413988e4 (accessed February 6, 2017).

Federal Geographic Data Committee. 2016. "U.S. National Grid (USNG) 10000m." https://gis-fema.hub.arcgis.com/datasets/d96095fb637846889fb0e46ce69e3967_2 (accessed June 17, 2016).

Finkelstein, Amy, Matthew Gentzkow, and Heidi L. Williams. 2019. "Place-Based Drivers of Mortality: Evidence from Migration." NBER Working Paper 25975.

Fowlie, Meredith, Mar Reguant, and Stephen P. Ryan. 2016. "Market-Based Emissions Regulation and Industry Dynamics." *Journal of Political Economy* 124 (1): 249–302.

Fowlie, Meredith, Edward Rubin, and Reed Walker. 2019. "Bringing Satellite-Based Air Quality Estimates Down to Earth." *AEA Papers and Proceedings* 109: 283–88.

Gallup. 2006. "Partisanship in U.S. States Gallup Polls." https://news.gallup.com/poll/26308/democratic-edge-partisanship-2006-evident-national-state-levels.aspx (accessed September 15, 2019).

Gilbert, Richard O. 1987. *Statistical Methods for Environmental Pollution Monitoring*. New York: John Wiley & Sons.

Giles, Cynthia. 2013. "Next Generation Compliance." *Environmental Forum* 30 (5): 22–26.

Graff Zivin, Joshua, and Matthew Neidell. 2009. "Days of Haze: Environmental Information Disclosure and Intertemporal Avoidance Behavior." *Journal of Environmental Economics and Management* 58 (2): 119–28.

Grainger, Corbett, Andrew Schreiber, and Wonjun Chang. 2017. "Do Regulators Strategically Avoid Pollution Hotspots when Siting Monitors? Evidence from Remote Sensing of Air Pollution." Unpublished.

Gray, Wayne B., and Jay P. Shimshack. 2011. "The Effectiveness of Environmental Monitoring and Enforcement: A Review of the Empirical Evidence." *Review of Environmental Economics and Policy* 5 (1): 3–24.

Greenstone, Michael. 2013. "Eliminating Unnecessary and Costly Red Tape Through Smarter Regulations." Testimony to US Congress Joint Economic Committee. June 26, 2013.

Greenstone, Michael, Guojun He, Ruixue Jia, and Tong Liu. 2020. "Can Technology Solve the Principal-Agent Problem? Evidence from China's War on Air Pollution." NBER Working Paper 27502.

Greenstone, Michael, John A. List, and Chad Syverson. 2012. "The Effects of Environmental Regulation on the Competitiveness of US Manufacturing." NBER Working Paper 18392.

Hanna, Rema. 2010. "US Environmental Regulation and FDI: Evidence from a Panel of US-Based Multinational Firms." *American Economic Journal: Applied Economics* 2 (3): 158–89.

Karplus, Valerie J., Shuang Zhang, and Douglas Almond. 2018. "Quantifying Coal Power Plant Responses to Tighter SO_2 Emissions Standards in China." *Proceedings of the National Academy of Sciences* 115 (27): 7004–09.

Kaufman, Y. J., D. Tanré, L. A. Ramer, E. F. Vermote, A. Chu, B. N. Holben. 1997. "Operational Remote Sensing of Tropospheric Aerosol over Land from EOS Moderate Resolution Imaging Spectroradiometer." *Journal of Geophysical Research* 102 (D14): 17051–67.

League of Conservation Voters. 2001–2013. "League of Conservation Voters Congressional Scorecard Data." https://scorecard.lcv.org/scorecard/archive (accessed October 8, 2016).

Liu, Yang, Meredith Franklin, Ralph Kahn, and Petros Koutrakis. 2007. "Using Aerosol Optical Thickness to Predict Ground-Level PM2.5 Concentrations in the St. Louis Area: A Comparison between MISR and MODIS." *Remote Sensing of Environment* 107 (1–2): 33–44.

Menne, Matthew J., Imke Durre, Bryant Korzeniewski, Shelley McNeal, Kristy Thomas, Xungang Yin, Steven Anthony, et al. National Oceanic and Atmospheric Administration National Centers for Environmental Information. 2001–2013. "Global Historical Climatology Network." ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/by_year (accessed February 16, 2016).

National Aeronautics and Space Administration. 2001–2013. "EarthData, The MODIS Level-2 Atmospheric Aerosol Product." http://dx.doi.org/10.5067/MODIS/MOD04_L2.006 (accessed February 29, 2016).

National Cancer Institute Surveillance, Epidemiology, and End Results Program. 2001–2013. "U.S. Population Data County-Level Population Files: 19 Age Groups 1969–2018 White, Black, Other, All States Combined." https://seer.cancer.gov/popdata/download.html (accessed June 8, 2015).

National Centers for Environmental Prediction. 2001–2013. "North American Regional Reanalysis U-wind at 10 m and V-wind at 10 m." ftp://ftp.cdc.noaa.gov/Datasets/NARR/Dailies/monolevel. (accessed October 19, 2015).

Neidell, Matthew. 2009. "Information, Avoidance Behavior, and Health: The Effect of Ozone on Asthma Hospitalizations." *Journal of Human Resources* 44 (2): 450–78.

Oliva, Paulina. 2015. "Environmental Regulations and Corruption: Automobile Emissions in Mexico City." *Journal of Political Economy* 123 (3): 686–724.

Public Integrity Section Criminal Division. 2001–2013. "Report to Congress on the Activities and Operations of the Public Integrity Section." https://www.justice.gov/criminal-pin/annual-reports (accessed January 17, 2017).

Remer, Lorraine A., Y. J. Kaufman, D. Tanré, S. Mattoo, D. A. Chu, J. Vanderlei Martins, R-R. Li, et al. 2005. "The MODIS Aerosol Algorithm, Products, and Validation." *Journal of the Atmospheric Sciences* 62 (4): 947–73.

Reynaert, Mathias, and James M. Sallee. 2016. "Corrective Policy and Goodhart's Law: The Case of Carbon Emissions from Automobiles." NBER Working Paper 22911.

Shimshack, Jay P. 2014. "The Economics of Environmental Monitoring and Enforcement." *Annual Review of Resource Economics* 6 (1): 339–60.

Sullivan, Daniel M., and Alan Krupnick. 2019. "Using Satellite Data to Fill the Gaps in the US Air Pollution Monitoring Network." Resources for the Future Working Paper 18-21.

Thompson, Gregory J., Daniel K. Carder, Marc C. Besch, Arvind Thiruvengadam, and Hemanth K. Kappanna. 2014. "In-use Emissions Testing of Light-duty Diesel Vehicles in the United States." Unpublished.

United States Bureau of Economic Analysis. 2001–2013. "Table CAEMP25N Total Full-Time and Part-Time Employment by NAICS Industry." https://apps.bea.gov/regional/downloadzip.cfm (accessed August 21, 2019).

United States Census Bureau. 2010. "TIGER/Line Geodatabases County TIGER/Line Shapefiles tl_2010_us_county10." https://www2.census.gov/geo/tiger/TIGER2010/COUNTY/2010/tl_2010_us_county10.zip (accessed July 2, 2016).

United States Census Bureau. 2013. "TIGER/Line Geodatabases Core-based Statistical Area TIGER/Line Shapefiles tl_2013_us_cbsa." https://www2.census.gov/geo/tiger/TIGER2013/CBSA/tl_2013_us_cbsa.zip (accessed July 7, 2014).

United States Department of Transportation Federal Highway Administration. 2001–2013. "Travel Monitoring Analysis System." https://www.fhwa.dot.gov/policyinformation/tables/tmasdata/ (accessed March 25, 2019).

United States Environmental Protection Agency. 1993. "Guidance for Estimating Ambient Air Monitoring Costs for Criteria Pollutants and Selected Air Toxic Pollutants." EPA-454/R-93-042.

United States Environmental Protection Agency. 1997a. "Guidance for Network Design and Optimum Site Exposure for PM2.5 and PM10." EPA-454/R-99-022.

United States Environmental Protection Agency. 1997b. "Guidance for Network Design and Optimum Site Exposure for PM2.5 and PM10." EPA-454/R-99-022.

United States Environmental Protection Agency. 1998a. "Memorandum: Guidelines for Granting Exemptions for Daily PM2.5 Monitoring."

United States Environmental Protection Agency. 1998b. "Memorandum: PM2.5 Site Types and Sampling Frequency During CY-99."

United States Environmental Protection Agency. 2001–2013a. "Air Data, Annual Summary Data." https://aqs.epa.gov/aqsweb/airdata/download_files.html (accessed October 17, 2015).

United States Environmental Protection Agency. 2001–2013b. "Air Data, Daily Summary Data PM2.5 FRM/FEM Mass (88101)."

United States Environmental Protection Agency. 2001–2013c. "Green Book, Whole or Part County Nonattainment Status by Year for All Criteria Pollutants." https://www.epa.gov/green-book/green-book-national-area-and-county-level-multi-pollutant-information (accessed October 10, 2015).

United States Environmental Protection Agency. 2001–2013d. "Toxic Release Inventory Basic Data Files: Calendar Years 1987–2018." https://www.epa.gov/toxics-release-inventory-tri-program/tri-basic-data-files-calendar-years-1987-2018? (accessed November 11, 2015).

United States Environmental Protection Agency. 2012. "Area Designations for the 2012 Annual Fine Particle (PM2.5) Standard State Recommendations and EPA Responses for Area Designations." https://www3.epa.gov/pm/designations/2012standards/staterec.htm (accessed October 8, 2016).

Vollaard, Ben. 2017. "Temporal Displacement of Environmental Crime: Evidence from Marine Oil Pollution." *Journal of Environmental Economics and Management* 82 (March): 168–80.

Walker, W. Reed. 2013. "The Transitional Costs of Sectoral Reallocation: Evidence from the Clean Air Act and the Workforce." *Quarterly Journal of Economics* 128 (4): 1787–835.

Zou, Eric Yongchen. 2021. "Replication Data for: Unwatched Pollution: The Effect of Intermittent Monitoring on Air Quality." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E124621V1.