Unwatched Pollution:
The Effect of Intermittent Monitoring on Air Quality

Eric Zou*

August 2018

Abstract

Intermittent monitoring of environmental standards may induce strategic increases in polluting activities during unmonitored times. 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. Correspondingly, cities' use of air quality warnings increases on monitored days, which suggests local governments' role in coordinating emission reductions. Higher levels of pollution on unmonitored days lead to measurable changes in outcomes, including lower school test scores and elevated crime.

* Dyson School of Applied Economics and Management, Cornell University. Email: ericzou@cornell.edu. 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, 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, 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.

1. Introduction

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. A consequence of intermittent monitoring is that it creates opportunities for polluters to show compliance during monitoring but to increase polluting activities when no monitoring takes place. The potential for such strategic responses is exacerbated when the polluters can anticipate the regulator's monitoring schedules, as demonstrated by the Volkswagen emissions scandal (Gates, Ewing, Russell, and Watkins, 2016). Strategic responses to intermittent monitoring, however, 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 non-random. 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 where monitoring is based on a publicly available, quasi-random schedule. Under the Clean Air Act, the U.S. Environmental Protection Agency (EPA) is charged with enforcing a national safety level for outdoor particle air pollution (known as particulate matter, henceforth PM) that all counties are required to achieve.[1] A network of monitoring sites tracks compliance with the standard. Due to high operating and maintenance 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 pre-announces a cyclical, once-every-sixth-day ("1-in-6-day") monitoring schedule each year to be followed by intermittent monitors. I leverage this monitoring policy experiment to examine whether higher concentrations of pollution are observed during days when monitoring is not scheduled. I begin by constructing an indirect measure of particle pollution using 13 years of satellite observations. This measure allows me to observe air quality both during monitor "on-days" and monitor "off-days". Because

[1] For example, the EPA requires counties to maintain concentrations of fine particulate matter ($PM_{2.5}$) at levels below 15 ug/m^3 (see Section 2.1). Failure to comply with the EPA's outdoor air quality regulation results in large economic costs for a county. For example, Walker (2013) estimates that the Clean Air Act's 1990 Amendments for particulate matter pollution costs workers about $8 billion in earnings in the years following non-attainment designations; Greenstone, List, and Syverson (2012) show that ambient pollution regulations led to reduced productivity representing a loss of about $20 billion in revenues annually among U.S. manufacturers. Notably, air quality often improves in non-compliance areas, as substantial resources are spent to limit polluting activities (e.g., Chay and Greenstone, 2003; Auffhammer, Bento, and Lowe, 2009).

the incentive to avoid monitoring is plausibly the only factor that changes on a 1-in-6-day basis, differences in pollution levels during off-days and on-days provide evidence of strategic responses to the monitoring schedule.

In the baseline analysis, I use the satellite measure to compare pollution levels on off-days and on-days 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 particle pollution during on-days than during off-days, while pollution levels between 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; and, around sites that monitory pollution daily, no pollution gap is found.

The paper proceeds by answering three key questions about the underlying mechanisms. The first question is: why would polluters engage in costly actions to change polluting activities every six days? I provide evidence that the pollution gap is driven not by pollution cycles (i.e., swaying polluting activities *every* six days), but instead by large pollution gaps during periods when reducing pollution helps avoid regulatory punishment. Consistent with the incentive structure of the regulation, I find pollution gaps over 7 percent when a county's recorded PM level approached the regulatory standard in the previous month; by contrast, I find no pollution gap when a county experienced recent good air quality.

The second question is: how could outdoor air pollution – a consequence of collective polluting activities – be effectively manipulated on such a high-frequency basis? I pursue this inquiry from several angles. First, I show that the conventional "market power"-style reasoning (that manipulation should be easier when counties have few major polluters) applies, but cannot fully explain the pollution gap. For example, an interaction with a Herfindahl-Hirschman index (HHI) of emission concentration indicates a significantly larger pollution gap in high HHI areas. However, I cannot reject the existence of a moderate yet individually significant gap in relatively low HHI regions. To further illustrate why an HHI-explanation is not sufficient, I examine the extreme case of air pollution near highways. I find that the pollution gap around PM monitors increases with the proximity of a monitor to the nearest highway. Because traffic responses to the monitoring schedule are unlikely to arise without coordination at a central level, this finding prompts a different direction of investigation. Specifically, I investigate whether local governments might play a role in coordinating the avoidance of pollution on monitoring days. I test the plausibility of this mechanism in the context of local governments' strategic issuance of air quality advisories. These

advisories call for citizens to reduce outdoor activities and vehicle use to prevent air quality deterioration. I show that these advisories are 10 percent more likely to be issued on days when pollution monitoring is scheduled. Although these results only speak to coordination of public behavior, they raise the possibility that local government coordination might occur in industrial settings as well.

I then ask: are other polluters responsible for the pollution gap? While the exercise is eventually constrained by the satellite data's resolution, I examine robust industry-level correlates that explain the geographic distribution of pollution gaps. I begin by mapping out substantial cross-sectional variation in the magnitude of the 1-in-6-day pollution gap across counties. This exercise identifies several regional clusters where the pollution gap is exceptionally large. I then present evidence on observable characteristics of these "hot-spot" regions. Using a simple statistical learning framework that combines traditional regression analysis and machine learning, I show that hot spots are more common in regions with high concentrations of certain polluting industries, such as mining and wood product manufacturing. While these correlational findings need not reflect causal effects of industries, complementary evidence suggests that the 1-in-6-day pollution gap is, in fact, observed around facilities in relevant industries.

While this paper focuses on strategic monitoring and the role of incentives, it also briefly examines the consequences of strategic polluting activities to shed light on potential benefits of implementing an alternative, every-day monitoring rule. I discuss three examples in line with the literature on the cognitive, behavioral, and health consequences of short-term pollution exposure (e.g., Ebenstein, Lavy, and Roth, 2016; Herrnstadt, Heyes, Muehlegger, and Saberian, 2016; Schlenker and Walker, 2015). First, I use test score data from the California High School Exit Examination to show that standardized exam performance is worse when the test date falls on a monitor off-day. Second, using crime records data, I find that criminal activity, especially violent crime, increase on off-days relative to on-days. Third, in a companion paper, Zou, Miller, and Molitor (2018) study the health consequences of the 1-in-6-day monitoring schedule. Using Medicare administrative data on the universe of elderly beneficiaries from 2001 to 2011, the authors show that higher pollution on the monitor off-day coincides with a significant increase in the elderly mortality rate the following day. The general conclusion of this analysis is that, outcomes would be better if pollution levels during off-days were consistent with those observed on on-days.

The idea that underpins cyclical pollution monitoring in use dates back at least to the 1970s (Akland, 1972). The methodology passed all subsequent evaluations of its *statistical* efficacy and was implemented

in the first regulation of particulate matter under the Clean Air Act.[2] This paper presents the first retrospective evaluation of the monitoring method's *enforcement* efficacy. I show how economic incentives to avoid regulation, coupled with a non-continuous enforcement scheme, can lead to significant deviations in pollution levels from the levels observable to the regulator – even if the monitoring method itself is statistically unbiased. My findings support two emerging themes in environmental regulatory policy making: that advanced continuous monitoring technologies should be promoted to replace discrete, sampling-based monitoring of the environment (e.g., Giles, 2013), and that regulation design needs to have retrospective and independent evaluations built in, in addition to traditional elements such as ex-ante cost-benefit effectiveness calculations (e.g., Greenstone, 2013; Auffhammer, 2015; Cropper, Fraas, and Morgenstern, 2017).

This paper adds to the growing literature on the economics of environmental regulation monitoring and enforcement, as reviewed by Gray and Shimshack (2011) and Shimshack (2014). With unprecedented access to better data, researchers have begun to reveal enforcement challenges that previously were largely theoretical in nature. Recent examples of analysis of monitoring avoidance include Oliva (2015) and Reynaert and Sallee (2017) in the context of vehicle exhaust testing, and Vollaard (2017) in the context of illegal dumping of oil waste. This paper extends this literature to a setting of a broad scale, national ambient pollution regulation; strong responses are found in various regions across the country, suggesting that strategic responses are far more pervasive than previously documented. Moreover, this paper suggests that polluters' ability to respond to environmental regulations on a short-term basis may be stronger than appreciated by previous literature that focuses on medium- and long-term regional or sectoral substitution of polluting activities (e.g., Becker and Henderson, 2000; Hanna, 2010; Fowlie, Reguant, and Ryan, 2016).

This paper also contributes to a developing understanding of the value of satellite data in regulatory decision making. Partly due to an inherent difference in their missions of operation, cooperation between environmental regulation agencies such as the U.S. EPA and agencies that operate satellite surveillance of atmospheric pollutants, including the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Association (NOAA), remains in a nascent stage. While regulators increasingly recognize the value of satellite data in certain fields such as wildfire surveillance (Ruminski et

[2] The typical approach in this literature is to compare summary statistics (such as mean and dispersion) of an every-day pollution sample to the same statistics computed from a subsample of days drawn using the cyclical method. See Akland (1972), Nehls and Akland (1973), Gilbert (1987), and Rumburg, Alldredge and Claiborn (2001).

al., 2006; Ichoku, Kahn and Chin, 2012) and air quality forecasting (Kittaka et al., 2004), the potential of satellite data is far from being fully exploited (see e.g., Duncan et al., 2014). This paper presents a specific example where the satellites' unique ability to complement intermittency of ground-based pollution monitoring can be leveraged to inform the desirability of a pollution monitoring policy. More broadly, this paper is linked to the growing use of remote sensing data in economic and policy research, reviewed by Donaldson and Storeygard (2016). An excellent and closely related study is Grainger, Schreiber, and Chang (2017), who rely on satellite data to investigate state governments' strategic decisions on the placement of ozone and NO_2 monitors.

The remainder of the paper is organized as follows. Section 2 provides a brief background on particulate matter (PM) regulation and monitoring in the United States. Section 3 presents the main identification of the off-days vs. on-days pollution gap. Sections 4 and 5 explore sources and mechanisms underlying the pollution gap. Section 6 examines the consequences of intermittent monitoring. Section 7 concludes.

2. Background on PM Regulation and Monitoring

2.1. PM Regulation

Regulation of ambient PM ($PM_{2.5}$ and PM_{10}) pollution in the United States is coordinated under the Clean Air Act. The act contains multiple provisions that oversight PM emissions at various industry levels. The regulation on *ambient* PM can be viewed as a policy lever to achieve one of the Act's ultimate goals: to maintain outdoor PM concentrations below the established safety levels of the National Ambient Air Quality Standards (NAAQS).[3]

Three effective PM standards were in place during my study period: 1) The 3-year average of daily "fine" PM (i.e., $PM_{2.5}$) level had to be below 15 ug/m3 ("$PM_{2.5}$ annual standard"), 2) the 3-year average of annual 98th percentile $PM_{2.5}$ level had to be below 35 ug/m3 ("$PM_{2.5}$ 24-hr standard"), and 3) the maximum "coarse" PM (i.e., PM_{10}) level had to be below 150 ug/m3. From 2001 to 2013, roughly 30% of monitored counties had ever been assigned a non-attainment status. Among them, roughly 60% of county × years are associated with the violation of the $PM_{2.5}$ annual standard.

[3] Separate standards have been established for PM_{10} and $PM_{2.5}$, and for four other outdoor pollutants: carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), and lead.

The Clean Air Act authorizes the EPA to enforce these standards. To make sure that the standards are met nationally, every year the EPA categorizes counties into "attainment" and "non-attainment" groups based on monitoring results. Non-attainment counties face substantially elevated regulation costs. In cases of non-attainment, the parent state is required to develop a State Implementation Plan that details how plant-specific regulations will be implemented in order to achieve compliance. These regulations typically involve the adoption of pollution abatement technologies and emission limits that affect both existing and new polluters. Local governments and individual polluters occasionally receive direct penalties from the EPA in cases of sustained non-attainment. These include financial sanctions that prohibit the approval of almost any highway projects or grants, as well as emission sanctions that require reduced emissions from existing pollution sources for any new or modified emission sources, where the reductions from existing sources must be at least twice the increases from new sources.

Non-attainment regulation poses real regulatory threats. Existing literature has shown that PM non-attainment leads to significant losses in employment and earnings (Walker, 2013), reductions in factories' productivity (Greenstone, List, and Syverson, 2012), and notably, reductions in ambient PM pollution (Auffhammer, Bento, and Lowe, 2009).[4]

2.2. PM Monitoring

Ambient PM concentration is monitored through a network of more than 1,200 monitoring sites across the country. These sites are usually placed in areas with high population density to ensure a reasonably representative measure of population pollution exposure. During my study period, 2001 to 2013, the PM monitoring network spanned more than 600 counties, accounting for roughly 70 percent of U.S. population.

Unlike monitoring of many gaseous air pollutants (such as ozone) that uses automated laser-based methods, PM monitoring is mostly filter-based and is associated with substantial manual operation and maintenance tasks such as field sample collection and laboratory analysis. By the EPA's data on the estimated cost break-down of PM monitoring (U.S. EPA, 1993), the annualized per-site cost of PM monitoring associated with monitor procurement, operation, and maintenance is estimated to be roughly $21,000 with a 1-in-6-day schedule and $41,000 with daily (1-in-1 day) monitoring (both estimates in 2013 dollars). With roughly 600 sites operating on a 1-in-6-day schedule, the cost savings from intermittent

[4] A rich literature documents the significant effects of other provisions of the NAAQS targeting at different air pollutants, such as Total Suspended Particulate and Ozone, on air quality (Henderson, 1996; Chay and Greenstone, 2005) and industrial activities (Becker and Henderson, 2000; Greenstone, 2002).

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.

High operation and maintenance costs prompt the practice of intermittent monitoring, which has been adopted since the initiation of atmospheric particle pollution sampling in the 1950s (U.S. Public Health Service, 1957). The cyclical 1-in-6-day monitoring method was introduced by Akland (1972) in the wake of regulatory monitoring of total suspended particulates in the 1970s. This practice was subsequently adopted for PM monitoring starting in the 1980s. By that time, the EPA also introduced more frequent 1-in-3 day and daily sampling schedules. These more frequent schedules are used at sites with higher levels of pollution where higher data capture rates are desired (U.S. EPA, 1985). Since the initiation of $PM_{2.5}$ monitoring in the 1990s, the EPA has been tending toward more frequent monitoring. For example, in the initiation of $PM_{2.5}$ monitoring, the EPA requires all $PM_{2.5}$ monitors to operate on a minimum of 1-in-3 day basis (40 CFR 58.13, 1997; U.S. EPA, 1998a). However, in response to states' concerns over cost burdens and a continuing interest in less frequent monitoring, the EPA allows exemptions for lower monitoring frequencies on a case-by-case basis (U.S. 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 daily (22 percent) schedule.[5] While the main analysis focuses on the 1-in-6-day schedule in which gaming is most likely to occur, below I also report "placebo" tests around 1-in-1-day (daily monitoring) sites, where no strategic responses are expected. The Appendix details responses to the more frequent, 1-in-3-day monitoring; these strategic responses are found to be much weaker.

As a guideline for states to schedule their monitoring routine, the EPA publishes a monitoring calendar on its website at the end of each calendar year, informing states of the monitoring schedule for the next calendar year. Figure 1 presents the calendar for 2001. Notice that the monitoring schedule is not staggered across 1-in-6-day monitors; put differently, all 1-in-6-day monitors are scheduled to sample pollution on Jan 1st, 2001, followed by Jan 7th, 2001, and so forth. Also, starting from year 2001, the 1-in-6-day monitoring schedule follows *strictly* six-day cycles beginning with the first monitoring date of Jan 1st, 2001. Therefore, monitoring status of any day post Jan 1st, 2001 is predetermined. In this sense, the

[5] A small number of monitors were granted the exemption to conduct seasonal sampling or, in rare cases, to follow a once-every-12-day schedule. The appendix includes a more detailed discussion of the history of intermittent PM monitoring and how monitoring frequencies are assigned. In general, more frequent sampling rates are assigned to areas with a higher chance of violating the NAAQS. For a comprehensive study on ambient monitoring network design, see Muller and Rudd (2017).

annual calendar publication can be viewed as only providing reminders to states of the monitoring schedule.

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

3.1. 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 max. 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 NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) algorithm. Equipped with a flexible set of spectral radiance instruments, MODIS retrieves atmospheric aerosol concentration 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 smaller value corresponding to lower level of aerosol concentration (Kaufman, et al., 1997; Remer et al., 2005; Voiland, 2010). In the United States, the vast majority of the index's value falls within the range of 0 to 1, with a mean of roughly 0.12. Specifically, I use the "MOD04_L2" product which provides "Level 2" files drawn from MODIS installed on satellite *Terra*'s platform. "Level 2" files are the cloud-screened, quality-controlled product developed out of the, raw, unscreened imageries from lower "Levels."[6] There are other sources that provide more "processed" versions of the aerosol data (such as the MODIS Level 3 global daily gridded file), but most would involve certain degrees of interpolation and smoothing in time and/or in space. Because much of the research design I use below exploits monitoring frequency variations in time and/or in space, here I

[6] To assure quality of the aerosol measure, MODIS employs infrared channels to help delete pixels with cloud presence. Because cloud coverage is not expected to change systematically on a 1-in-6-day basis, I do not expect cloud-screening to be a significant source of bias in my estimates.

use the least-processed data to minimize the impact of ex-post modeling on my results. To simplify language, in subsequent sections I refer to this measure as "aerosol" or "aerosol concentration."

To facilitate interpretation, it is perhaps helpful to proceed with three connections between the concept of the satellite-based aerosol measure and the traditional ground monitor-based PM measure. First, despite the difference in measurement approaches, ground monitor and satellite 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 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 measure and ground-monitoring data.[7]

Second, unlike PM monitors, which measure pollution concentration at the ground level, the satellite orbits at a height of roughly 700 km; when the satellite passes over a given point on the ground, it captures aerosol conditions for the entire column of air from its viewpoint. Therefore, the aerosol measure should be interpreted as the amount of particle pollution in the entire atmosphere corresponding to a given location, rather than a precise measure of population exposure at that location. For example, while the satellite may see a tall smoke stack generating pollution, the population in the local area is probably exposed to that pollution to a lesser extent, as the wind blows.

Finally, most ground-monitoring data provide the average pollution concentration within a given time period (usually within a 24-hour period) on days when monitoring is conducted; in contrast, satellites provide daily snapshots of pollution in a given area at an (almost) fixed time of day. Aboard the polar-orbiting satellite *Terra*, MODIS continuously observes the earth's atmosphere with a 2,330km-wide sweeping swath, scanning each point on the planet every day at approximately 10:30am local time.[8] The swaths are then merged together to form daily imagery of aerosol concentration with a spatial resolution

[7] See e.g., Liu, Franklin, Kahn, and Koutrakis (2007), Lee, Coull, Bell, and Koutrakis (2012), Zhang and Li (2015). Previous economic research has also used the aerosol measure as the proxy for air pollution in developing-country contexts where monitoring data are sparse, see e.g. Foster, Gutierrez, and Kumar (2009), Chen, Jin, Kumar, and Shi (2013), and Bombardini and Li (2016). For a recent review, see Donaldson and Storeygard (2016).

[8] MODIS is also installed on the satellite *Aqua*, *Terra*'s sister satellite, although data from *Aqua*'s platform is available only since 2004. For that reason, this paper uses data from *Terra*.

of 10km×10km (i.e. half of the size of the average U.S. ZIP Code). My key outcome dataset is a daily panel of aerosol concentration linking each 10km×10km grid cell from 2001 to 2013.[9]

Summary Statistics. Table 1 presents satellite aerosol and ground monitor summary statistics by calendar year. Columns 1 to 4 show 10km×10km grid × daily-level satellite aerosol statistics. Over the period of study, the aerosol level stayed relatively stable, declining by an average of 0.5 percent per year from 2001 to 2013.[10]

Columns 5 to 7 of Table 1 show monitor-level statistics, including, for each year, the total number of monitors, number of 1-in-6-day monitors, and number of 1-in-6-day monitors exceeding PM NAAQS.[11] Columns 5 and 6 show that the total number of monitors decreased over time. This trend began in 1997, with a NAAQS revision that initiated $PM_{2.5}$ monitoring and redirected sources from monitoring of PM_{10} (U.S. EPA, 1997a). As will be described in further detail below, I exploit monitor retirement events to show that the pollution gap between off- and on-days disappears as monitors retire. Column 7 shows that, every year, about 7 percent of 1-in-6-day monitors exceed NAAQS.[12]

Columns 8 and 9 present counts of monitoring site. A monitoring site is a geographic unit that may contain multiple monitors. The main analysis is done at the monitoring site level because the locations of different, individual monitors within the site are not distinguished among one other. To be conservative in aggregating individual monitor-level schedules to the site level, I define a site to be a 1-in-6-day site if any PM monitor in that site follows the 1-in-6-day schedule. Defining the sample this way is expected to bias against finding strategic responses to the six-day monitoring schedule. For example, some monitoring sites may have a daily monitor *and* a 1-in-6-day monitor, where the latter is used to provide quality

[9] Pixel arrays in the original satellite imagery do not identically overlap on a daily basis. To create the panel dataset, the imagery is mapped onto a fixed map of 10km×10km grid cells, which I obtain from the U.S. National Grid Information Center. This procedure ensures that the grid dataset preserves the original resolution of the satellite imagery, while each grid tracks aerosol concentrations for the same area over time. See the Appendix for more details on the satellite data.

[10] The rate of decline of aerosol is about 1.42 percent per year around $PM_{2.5}$ monitors and about 1.05 percent around PM_{10} monitors. These are roughly on par with trends of pollution measured using ground monitor data, which show an annual decline rate of 2.3 percent for $PM_{2.5}$ and 1.35 percent for PM_{10}.

[11] In all following analyses, I restrict to NAAQS-eligible monitors, i.e., monitors that obtain at least 75 percent of required samples for each quarter of the year. Monitors that fail to satisfy such requirement are not eligible to be used by states to demonstrate NAAQS compliance. Results are not sensitive to the inclusion of ineligible monitors.

[12] The number of NAAQS-exceeding monitors decreases over the sample period, with a temporary increase around 2006 due to a tightening of the $PM_{2.5}$ maximum standards starting October 2006.

assurance data for the former.[13] In the analysis below, I confirm that the evidence of schedule gaming is stronger if I restrict to sites with a standalone 1-in-6-day monitor.

Columns 10 to 13 present monitor-level statistics aggregated to the county level. Column 10 shows the number of counties that have PM monitors. Column 11 shows that a population of roughly 200 million people live in these counties in the lower 48 states, and column 13 shows that about 65 percent of them live in counties that have at least one 1-in-6-day monitor. Note that, although total population coverage of PM-monitoring counties has stayed almost constant over time, population that live in 1-in-6-day counties has declined significantly during the study period, driven by the substantial decrease in the number of 1-in-6-day monitors (column 6).

3.2. Empirical Framework

Empirical Specification. 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 estimation equation is

$$Aerosol_{st} = \sum_{d=-3,d\neq 0}^{2} \beta_d \cdot 1(t = d) + Time_t + \alpha_s + X_{st}\gamma + \varepsilon_{st} \qquad (1)$$

where $Aerosol_{st}$ is the logged satellite aerosol concentration at monitoring site s at time t, measured by the daily aerosol level within the 10km×10km grid cell that corresponds to the land area containing the site.[14] The key coefficients of interest are the β's (β_{-3}, β_{-2}, β_{-1}, β_1, β_2) that represent air pollution on each day of cycle, running from three days before to two days after the on-day. The on-day is marked as day 0, which is the omitted category in the regression, so the β's should be interpreted as percentage changes in air pollution during the off-days relative to the on-day. The strictly 1-in-6-day cyclicality of the off-days treatment implies that very few confounders may bias β's from identifying the causal effect of

[13] For quality assurance purpose, the EPA requires a certain percentage of 1-in-1 day and 1-in-3 day monitors to be "collocated" with a 1-in-6-day monitor in each state. See more discussion in the Appendix.

[14] For the sake of description, in the main analysis I ignore the fact that a 10km×10km grid may contain multiple monitoring sites. In fact, during the study period the most populous grid contains 13 monitoring sites. However, more than 85 percent of grids contain a single site, and less than 1 percent of grids contain more than three sites. I confirm that dropping duplicative grid-day observations has negligible impacts on the results.

the monitoring schedule, especially given the extensive length of the panel dataset. To confirm this point, I report results from two types of specifications. In the first, I report estimates of β's conditional on no covariates, so that β_d simply shows the raw difference between pollution on day *d* of a cycle relative to the on-day. 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 X_{st}, which is a matrix of time-variant weather controls including daily temperature categorized into ten 10-degree bins, daily wind speed quartiles, and quadratic daily precipitation.[15] Since 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.

I also estimate a more parsimonious version of equation (1) which takes the following form

$$Aerosol_{st} = \beta \cdot 1(Off\text{-}days_t) + Time_t + \alpha_s + X_{st}\gamma + \varepsilon_{st} \qquad (2)$$

All components in this estimation equation are the same with equation (1), with the only difference being that, rather than having five dummies separately indicating days of a 1-in-6-day monitoring cycle, equation (2) includes the $1(Off\text{-}days_t)$ dummy which indicates all five off-days. The coefficient β therefore represents the gap in pollution levels between an average off-day vs. an average on-day.

To interpret β as the causal impact of the monitoring schedule on air pollution, the identification assumption must hold that no differential pollution levels would have been observed between on- and off-days in the absence of the monitoring schedule. Put differently, I assume that the only reason that ambient air quality might show a significant pattern once every six days is because polluters react to the incentive of monitoring avoidance generated by the 1-in-6-day sampling schedule. While this assumption is not directly testable, in section 3.4 I implement placebo tests; these tests are based on the idea that a pollution gap is not expected to occur in areas that lack incentives to reduce pollution during monitored days, such as would be the case in regions in which monitors sample air quality every day.

[15] In unreported results, I confirm that the results are robust to including more stringent time fixed effects controls such as month-of-sample fixed effects.

Interpretation: changes versus levels. The estimation framework above identifies the causal effect of intermittent monitoring on strategic responses by looking at *changes* in pollution levels on off-days relative to on-days ("Is off-days pollution higher than on-days?"). The framework, however, does not identify the effect of intermittent monitoring on the average *level* of pollution. For example, the research design does not speak to the counterfactual levels of air pollution if places that currently undertake 1-in-6-day monitoring conduct 1-in-1 day monitoring instead. While this paper primarily focuses on studying strategic responses, knowledge of counterfactual pollution levels is important when thinking about the social damages, such as health costs, due to intermittent monitoring. I discuss this issue further in section 6.

3.3. Baseline Results: 1-in-6-day Pollution Gap

I first estimate equation (1) using my preferred sample, which includes all monitoring sites containing at least one 1-in-6-day PM monitor from 2001-2013. The sample includes 1,193 monitoring sites that span 563 counties in the lower 48 states. Figure 2 reports the results. I do not condition the regression on any covariates, so the solid line simply represents the time path of air pollution in a 1-in-6-day monitoring cycle, averaged across all cycles in the sample. The results reveal polluters' striking ability to manipulate ambient air quality at the monitoring sites on a short-term basis: within a typical monitoring cycle, air pollution exhibits a flat path, except for a sharp drop during the on-day.

Table 2 reports the average 1-in-6-day off-days vs. on-days pollution gap using equation (2). 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. In column 3, I restrict the estimation sample to sites with a standalone 1-in-6-day PM monitor. This action is expected to reduce the diluting impact from sites where the 1-in-6-day monitor is collocated with high-frequency monitors, such as ones that sample every day. In this case, the pollution gap rises to about 1.8 percent. The effect persists if I further restrict the sample to counties with only 1-in-6-day monitors (column 4). The fact that gaming on average appears stronger in standalone sites provides suggestive confirmation that gaming is in fact targeting 1-in-6-day monitors.[16]

[16] I perform the same analysis for sites that follow the 1-in-3-day monitoring schedule, with details reported in the appendix. I find that for 1-in-3-day PM sites, the pollution gap is less than 0.3 percent and not statistically significant, suggesting that doubling the monitoring frequency at 1-in-6-day sites may be a feasible policy option to reduce strategic responses. In unreported analysis, I find that there is no statistically detectable effect around 1-in-6-day monitors that "collocate" with 1-in-1-day monitors (as explained in the appendix, such 1-in-6-day monitors tend to serve the purpose of data quality control).

3.4. Placebo Tests

The identification assumption states that no pollution gap would have been observed in the absence of the 1-in-6-day monitoring schedule. To boost the confidence in the internal validity of the empirical design, I provide two types of placebo tests that establish a null pollution gap in places where gaming is not expected.

The first 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 disappearance of the pollution gap after sites are removed.[17] To operationalize this test, I first draw upon information in the EPA's monitor listing file and identify 490 cases of 1-in-6-day monitoring sites retirement events. The analysis then uses the satellite measure to track air quality in the areas where these sites are located, and compares the off-days vs. on-days pollution gap before and after sites' retirement. Note that I can estimate the pollution gap even after the site was removed because the monitoring calendar is universally applied, and, hence, even in the absence of a working monitoring site, I know what the sampling dates would have been. Figure 3 reports the results, where the pollution gap is shown as a function of years relative to sites' retirement. The gap is about 2.1 percent for the time frame when the site was still operating; for the exact same area, the gap closes immediately after monitor retires.

In the second type of placebo check, I apply the same methodology to estimate the 1-in-6-day pollution gap near sites where gaming is either not feasible or not necessary. These include about 560 monitoring sites that follow the 1-in-1 day schedule, and about 800 hazardous air pollutants (HAPs) monitoring sites that also follow a 1-in-6-day schedule for the monitoring of other pollutants not subject to any regulatory standards.[18] Table 3 reports that no significant 1-in-6-day pollution pattern is detected near these sites. These findings again support the identification assumption that no 1-in-6-day pollution gap would have been observed in the absence of the 1-in-6-day monitoring schedule.[19]

---

[17] Monitoring sites retire for two main reasons: (1) budgets are reallocated to the monitoring of other pollutants; (2) very occasionally, pollution sources in the monitored area exit, leaving little need to continue monitoring. These reasons are not distinguished in the test which simply exploits the discrete removal of the incentive to respond to 1-in-6-day monitoring.

[18] These sites monitor a total of 734 different toxic air pollutants among which the five most commonly monitored are Benzene, Toluene, Ethylbenzene, o-Xylene, and Styrene.

[19] Simple power calculations show that the placebo tests have over 80 percent statistical power to detect a 1.5 percent effect (similar to the effect size in Table 2) at a 5 percent significance level.

4. Interpretation of the Pollution Gap

This section examines two interpretational aspects of the pollution gap finding. Section 4.1 explores the role of regulatory incentives, and shows that the 1-in-6-day pollution gap is likely driven by large monitored-day pollution reductions during generally high pollution periods, rather than by routinely cyclical polluting activities. Section 4.2 considers a series of indirect tests of ways in which the pollution gap may be coordinated.

4.1. Evidence on the Role of Incentives

Although the monitoring schedule follows strictly six-day cycles, avoidance actions are unlikely to arise *every* six days. Naturally, we expect these actions to concentrate in time periods or regions where gaming the monitoring schedule is highly rewarding, e.g., when counties come close to or exceed regulatory air quality standards (Auffhammer, Bento, and Lowe, 2009). Here I test for a heterogeneous pollution gap by the county's measured pollution levels in the recent past. I begin by augmenting equation (2) with an interaction between the $Offdays_t$ dummy and the county's average $PM_{2.5}$ level in the previous month. In this exercise, I include the control matrix (containing geographic and time fixed effects, and time-variant weather controls, etc.) as specified in section 3.2 in order to leverage the idiosyncratic variation of monthly $PM_{2.5}$ levels within the same county.

Figure 4, panel A plots pollution gap estimates separately by bins of the county's previous month $PM_{2.5}$ level (< 6 ug/m^3, 6 – 10 ug/m^3, 11 – 15 ug/m^3, and > 15 ug/m^3). I find pollution gaps of over 7 percent when the previous month's pollution level exceeded the regulatory standard of 15 ug/m^3. In contrast, no pollution gap is detected at times when pollution level is far below the standard (< 6 ug/m^3). This result indicates a strikingly short-term gaming strategy: monitoring avoidance efforts are allocated to "pull down" the measured pollution level when it is approaching or has recently exceeded the regulatory guideline.[20] To extend this exercise further, I examine the dynamics of the "pollution gap – past pollution levels" correlation. I estimate a specification that correlates the magnitude of the pollution gap with six lags of monthly average $PM_{2.5}$ levels, and, as a placebo exercise, six leads of monthly pollution, as well. Figure 4, panel B shows that neither longer-term lags in pollution non-compliance in the county nor leads of pollution predict a pollution gap in the current month; only pollution concentration in the immediate past seems to matter.

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

These results indicate a high level of sophistication in monitoring avoidance. Even though counties do not fall into non-compliance for just a month's exceedance, close attention appears to be paid to pollution levels in the recent past, and monitoring avoidance efforts are allocated to target those most polluted time periods. While individual polluters may have the capacity to carry this out, that probably requires a set of very conscientious factory managers who understand their relative contributions to ambient air quality, the federal monitoring schedule, and what pollution levels have recently been recorded in the county. From a plausibility perspective, this information is much more readily available to the local government agencies who oversight air quality on a day-to-day basis. I explore coordination issues in the following subsection.

4.2. Evidence on Local Government Coordination

The Role of Emission Concentration. One possibility is that the pollution gap is organized entirely through self-coordinated avoidance actions by polluters. Because ambient pollution levels are jointly determined by individual polluters' emissions, under the self-coordination hypothesis, we expect a pollution gap to occur only in regions with few major polluters. I test this hypothesis by documenting heterogeneous pollution gaps by county's emission Herfindahl-Hirschman index (HHI) constructed from the EPA's Toxic Release Inventory (TRI) data which contain annual observations of plants' reported total air emissions. The HHI ranges from 0 to 1 and takes larger values in counties where fewer polluters contribute to total emissions. The Appendix provides more detail of the HHI construction.

Figure 5 reports heterogeneous 1-in-6-day pollution patterns by high (≥ 0.9) vs. low (< 0.9) HHI. Results show significantly stronger gaming in areas with high levels of emission concentration, where the pollution gap averages 3.1 percent. However, I cannot reject the existence of a significant pollution of about 1.2 percent in low HHI regions as well.

Pollution Gap Near Highways. The view that pollution gap is driven by polluters' self-coordinated avoidance does not appear to be supported by the data. To extend this exercise to the extreme, I examine evidence on pollution gap related to the traffic sector, where self-coordinated gaming by individual drivers is improbable. To operationalize the test, I augment equation (2) of the paper by allowing the pollution gap estimate to vary flexibly by the monitor's distance to nearest highway segment. I focus on monitors located within 3 miles of the highway (< 0.5 miles, 0.5-1 miles, … , 2.5-3 miles) and all other monitors are pooled into a single group (> 3 miles).

Figure 6 suggests a distance gradient. I find that monitors within 1 mile of highways exhibit a strong response (roughly 2.5 percent pollution gap), while the pollution gap is not detectable for monitors that are between 1 mile and 3 miles away from highways. The results also suggest that the pollution gap observed near highways is likely to only partially explain the main finding on the average pollution gap, as a significant pollution gap is also precisely estimated for all monitors that fall more than 3 miles away from highways.

Strategic Issuance of Pollution "Action Day" Advisories. The highway exercise suggests the possible role of local government coordination in strategic responses to the EPA's PM monitoring schedule. To further test such possibility, I focus on an empirical setting where a specific form of coordination behavior is observable. Throughout the country, many local air pollution control agencies adopt air pollution "Action Day" programs that issue public warnings when ambient pollution concentration is expected to reach unhealthy levels.[21] When an action day is announced, citizens are advised to "take actions," such as reducing energy and automobile use, to prevent deterioration of air quality. Previous research shows that these advisories affect outdoor activities and transportation decisions (Neidell, 2009; Cutter and Neidell, 2009; Graff Zivin and Neidell, 2009). I obtain all Action Day records from the EPA's AirNow program. From 2004 to 2013, 346 reporting areas voluntarily adopted Action Day programs. These areas are a mix of cities, counties, metro areas, and states covering roughly 50 percent of the U.S. population. To avoid double counting issuances in cases of overlapping or nested jurisdictions, I aggregate the data to the core-based metro area (CBSA) level.[22] This gives me a total of 14,945 issuances at the CBSA × daily level from 2004 to 2013.

I examine whether Action Day advisories are more likely to be issued on days when PM monitoring is scheduled. The hypothesis is tested using the same estimation framework outlined in equation (1), with the outcome variable changed to a dummy that indicates whether an Action Day advisory is issued on the CBSA × day. Since Action Days are often issued on an "episode" basis spanning multiple days, to better

[21] Prediction of future pollution is based on previous days' continuous monitoring data. These data are obtained from proxy (i.e. non-regulatory) monitors that are able to provide pollution estimates in real-time. See the Appendix for more detailed discussion of these continuous monitors.

[22] CBSAs are urban-centered geographic units representing county groups. Each CBSA has a population size of at least 10,000 and has commuting patterns tied to the urban center. Aggregation of pollution advisories to the CBSA level is motivated by the EPA's rule which specifies that public broadcast of the Air Quality Index (AQI), which usually serves as the base for Action Day issuances, should be implemented at the CBSA level (U.S. EPA, 2013). I have confirmed that results are not sensitive to using disaggregated reporting agency level data.

capture issuance timing I also estimate a variant of the model where only the first issuance (6,232 out of 14,945 total Action Days) is count for a period of consecutive Action Days.

The main result is summarized in Figure 7, panel A, which tests for strategic timing of Action Day declarations adjusted for consecutive issuance. The graphical pattern provides evidence of a significant excess of advisories that are issued on pollution sampling days. On average, an action day is 0.108 percentage points more likely to be issued on an on-day, based on an average daily issuance probability of roughly 1 percent at the CBSA level.

Next, returning to the main pollution gap specification as outlined in equation (2), I test whether Action Day warnings are in fact effective in generating a difference in air quality on off- an on-days. For each monitor, I group its 1-in-6-day cycle into three broad categories: (1) those that are located in which the metropolitan area has received a warning; (2) those that are located in areas that have not received warnings but have adopted Action Day programs; and (3) those that are located in metropolitan areas with no Action Day programs. Results in Table 4 appear to suggest that warnings are effective in generating pollution reduction on monitored days: cycles that contain warnings realize pollution reductions of 5 percent to 7 percent during on-days relative to off-days. On the other hand, strategic warnings are unlikely to fully explain the pollution gap finding. For example, I find a statistically significant, albeit smaller, pollution gap when no warnings are issued. There is also some evidence that a pollution gap exists in areas that did not adopt warning infrastructure. Overall, these findings on coordination in the specific setting of public behavior manipulation makes it possible to expect that coordination also exists in industrial settings. I explore industry characteristics of the pollution gap in the following section.

5. Industrial Characteristics of the Pollution Gap

Having documented that local government coordination in the traffic sector may not fully explain the pollution gap, I now turn to the correlational analysis of industry responses. It is important to note that the same set of mechanisms that I discussed in previous analyses may be at work in industrial settings as well: polluting facilities that face the threat of air quality regulation may self-initiate strategic responses; alternatively, local governments may coordinate the avoidance of federal pollution monitoring by encouraging reductions of pollution during monitored days. While the data are not equipped to distinguish between these two mechanisms, they can potentially point us to industries that coincide with the strongest occurrence of pollution gap. Section 5.1 begins by documenting the variation in the

magnitude of the pollution gap across counties. In section 5.2, I present a data-driven exercise to explore industry characteristics of regions with large pollution gaps.

5.1. Identifying Pollution Gap "Hot Spots"

Pollution Gaps across Counties. I begin by estimating one off-day vs. on-day pollution gap per county in the contiguous U.S. For each county *c*, the following estimation equation is fitted:

$$Aerosol_{i(c)t} = \beta_c \cdot 1(Off\text{-}days_t) + Time_t + \alpha_{(c)} + \varepsilon_{i(c)t} \qquad (3)$$

where $Aerosol_{i(c)t}$ denotes logged aerosol level in grid *i* inside county *c* on date *t*. Seasonality controls $Time_t$ include year, month-of-year, and day-of-week dummies. County fixed effects $\alpha_{(c)}$ are not actually included in the estimation, because regressions are run separately by county. The average county-level regression contains 35,236 observations (median = 21,086 observations) at the 10km×10km-by-daily level.[23]

Figure 8 plots $\hat{\beta}_c$'s, the county-level estimates of the pollution gap from equation (3). The map is drawn so that warmer colors indicate areas where pollution is higher during monitor off-days relative to on-days. The map exhibits two features. First, areas with large pollution gaps exhibit a strong "clustering" pattern that is evident in parts of California, parts of Montana, Southern Texas, as well as a group of states in the Midwestern U.S. In the following analysis, I define pollution gap "hot-spot" counties as those with a top-decile pollution gap estimate. However, due to the spatial nature, most conclusions I present in this section are not qualitatively 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* pollution gap, which indicates that aerosol levels are *lower* during off-days in comparison to on-days. These areas are usually observed in between hot spots, and are more common in the eastern United States.

In the Appendix, I show that wind transport provides a potential explanation of negative pollution gaps. I find that the timing when pollution drops within a six-day monitoring cycle shifts as a function of a

[23] Alternatively, one could employ a county fixed effects approach and obtain the same point estimates of β_c in a single regression where the off-days dummy and the time fixed effects are allowed to vary flexibly by county; such a procedure, however, will substantially increase the computational burden. Results are very similar in estimating equation (3) with alternative levels of stringency, such as removing all controls or further including for grid cell-specific effects in addition to seasonality controls.

county's distance to the nearest pollution hot-spot county; the speed of pollution gap shifting is consistent with the average wind speed; shifting primarily occurs along the direction in which prevailing winds blow. In other words, the spatial pattern of positive and negative pollution gap is consistent with a simple dispersion dynamic in which 1-in-6-day pollution gaps generated by the hot-spot counties were transported to the rest of the counties. In subsequent sections, I focus primarily on characteristics of hot spots and set aside the variation in the magnitude of the pollution gap, because the latter is more likely to reflect a complex mix of local strategic responses and a blown-in pollution pattern from neighboring areas.

5.2. Industry Characteristics of the Pollution Gap Hot Spots

Because industries differ across various dimensions such as PM emission intensity, location, production technology and so forth, they may respond differently to the intermittent monitoring schedule. Here I explore industry determinants by linking pollution gap hot spots to variation in industry concentration. The conceptual exercise is to compare the maps of pollution gap hot spots and industry concentrations. A stronger industry correlation is therefore indicated by a better line-up between hot spots and counties with a high concentration of that particular industry. I implement this exercise in a simple statistical learning framework. Begin with the following estimation equation:

$$1(Hotspot_c) = \sum_j \alpha_j \cdot \underbrace{1\left(Industry_c^j\right)}_{\substack{\text{High concentration} \\ \text{of industry } j \\ \text{in county } c}} + \overbrace{X_c}^{\text{other controls}} \cdot \gamma + \varepsilon_c \quad (4)$$

The key explanatory variables are a set of dummies $1\left(Industry_c^j\right)$, each indicating whether county *c* is in the top decile of industry *j*'s county-level concentration distribution, where concentration is measured by the share of industry-*j* employment relative to the county total. In other words, for each industry *j*, I compute its employment share in each county *c*, and $1\left(Industry_c^j\right)$ indicates counties in the top decile for that industry. To construct the employment share measures, I use 3-digit NAICS-level employment data from the Census County Business Pattern. All measures are computed as averages from 2001 to 2013. The key coefficients of interest are α_j's, which tell us if a high concentration of industry *j* is predictive of hot spots, conditional on other industry profiles in the county. Other county-level

characteristics X_c include an indicator for the presence of 1-in-6-day monitoring sites, whether the county has ever been assigned PM non-attainment status, and other controls, such as state fixed effects dummies, described in more detail below. [24]

While the exercise does not aim to tease out the causal relationship between hot spots and industry concentrations, I examine which industry characteristics appear to be fundamental features of hot-spot counties by exploring robustness of industry correlates through a variety of specification changes. I begin by enriching the specifications to capture omitted variables that potentially correlate with observable industry characteristics. Several actions are taken. First, I present an augmented version of equation (4) where, in addition to "polluting" industries (defined 2-digit NAICS industries that contribute to at least 1 percent of national total PM_{10} emissions[25]), I control for dummies indicating high concentrations of all available 3-digit industries included in the County Business Pattern. To capture potential influence from mobile pollution sources, I also include dummies for whether highways and/or major (Class 1) railroads intersect the county. Second, I report models that use within-state variations in industrial concentrations by including state fixed effects. This action is expected to reduce the influence of unobserved geographic correlates that don't vary much within a confined area. Third, I estimate the model in subsamples that are restricted to counties that fall within 50 miles of the hot spots. In a related specification, I further restrict to hot-spot counties themselves, and use the same estimation model to predict the intensity of the pollution gap. This estimation uses $\hat{\beta}_c$ from equation (3) as the outcome variable while weighting observations by the precision measure $1/\mathrm{SE}_c^2$ where SE_c is the standard error for the pollution gap estimate $\hat{\beta}_c$.

Next, I simplify specifications by seeking sparse solution to the econometric modeling. While the dummy-variable approach enjoys flexibility, having too many explanatory variables can potentially make the estimation inefficient and reduce the generalizability of the results. I pair OLS estimation with the Least Absolute Shrinkage and Selection Operator (LASSO) which performs variable selection on the entire set of regulatory, industrial, and state dummies (Tibshirani, 1996). I employ a simple version of LASSO which regularizes the complexity of the model by imposing a penalty on the sum of absolute values of all

[24] In the Appendix, I report that both PM non-attainment and 1-in-6-day monitoring statuses are positive correlates of pollution gap hot spots.

[25] Emission shares are computed using the EPA's 2011 National Emissions Inventory which contains a near-census of all pollution sources known to the EPA from its various emission-reporting programs.

regression coefficients. This procedure yields a sparse solution of the original optimization problem by shrinking coefficients of certain explanatory variables to zero.[26]

Figure 9 offers a visualization of the results. In this heat chart, a cell represents *t*-statistics of a coefficient estimate; cells on a same column are obtained from a joint regression which is specified by the chart's header section. The color scheme is such that redness indicates positive correlation; blueness indicates negative correlation; and darker color indicates stronger correlation. I organize polluting industries into sector blocks, ranked from sectors that emitted the highest share of PM (Utilities) to the lowest (Administrative, Support, Waste Management, and Remediation Services). In the interest of space, I do not report coefficients for all other non-polluting industries that contribute less than 1 percent of total PM emissions.

Begin with the panel on the left-hand side, which corresponds to OLS estimation. There are two evident color patterns. First, strong signals (i.e., dark red) are observed from the upper part of the chart, while signals from the lower part of the chart are weak. This pattern is consistent with the view that strategic responses are expected to come from polluters that emit high volumes of particulate matter pollution and therefore have a chance to manipulate ambient air quality. Second, a number of polluting industries exhibit consistently strong positive correlations with hot spots. These include wood product manufacturing, chemical product manufacturing, and mining. Moreover, consistent with previous discussion on traffic manipulation, counties with highway segments are more likely to be pollution gap hot spots. The LASSO estimation results on the right hand side panel largely confirm 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 the Appendix, I report complementary evidence of industrial responses with finer geographic detail. Linking the satellite pollution measure to the EPA's Toxic Release Inventory data that contain information on polluter location, I find that a 1-in-6-day pollution gap is indeed observed near neighborhoods that contain wood product, chemical manufacturing, and coal mining facilities. This effect also shows a distance-gradient in which the response is stronger the closer the facility is to the nearest 1-in-6-day PM monitor.

[26] Conceptually, this imposes a constraint of $\sum_p |\gamma_p| \leq \lambda$ when solving the original OLS problem for equation (4). This would yield a path of solution depending on the degree of penalization λ, and I choose the optimal λ using a 10-fold cross validation.

A capacity-related explanation seems reasonable regarding why some industries may respond strategically to the monitoring schedule while others may not. For example, while the utility sector is a major PM emitter, power plants, especially coal plants, often run around the clock; for them, ramping up and down production in response to short-term monitoring schedule is very costly. By contrast, wood product manufacturers also contribute to a significant 3 percent of total PM emissions among point sources, but usually run on a low-capacity factor. In fact, wood plants, which operate at roughly 60 percent of capacity, have the lowest utilization rate among all polluting manufacturers examined in this study. The unused capacity might enable these plants to shift production activities around in avoidance of the pollution monitoring schedule.[27]

6. Consequences of the Pollution Gap

Thus far, I have shown that the intermittent monitoring policy induces strategic increases in pollution on unmonitored days. In this section, I present evidence that such unintended pollution consequences affect well-being. Specifically, I ask whether health and other economic outcomes are worse on monitor off-days when pollution increases relative to monitor on-days. Put differently, I examine whether outcomes would be better if pollution levels during off-days were consistent with the levels observed on on-days.[28]

The most natural beginning point is to examine the effect of the pollution gap on health. This topic is treated in a companion paper by Zou, Miller, and Molitor (2018) linking the pollution gap to changes in daily health outcomes constructed from Medicare administrative records. The authors show that higher pollution on an unmonitored day coincides with a significant increase in elderly mortality rate the

[27] Capacity utilization statistics are drawn from the Census Bureau's 2008-2013 Quarterly Survey of Plant Capacity Utilization. The wood industry's low capacity utilization rate is partially explained by the fact that it is usually a hard hit during recessions. Capacity utilization for wood (entire manufacturing) dropped from 70 percent (78 percent) at the beginning of 2008 to a lowest 48 percent (67 percent) in the second quarter of 2009.

[28] As is mentioned in Section 3, the full cost of the pollution gap cannot be calculated without knowing the counterfactual level of pollution under daily monitoring. To see this, consider the pollution consequences of "upgrading" from 1-in-6-day monitoring to 1-in-1-day monitoring. It is tempting to think that when polluters face the same incentive to reduce pollution every day that they faced only once per six days before, pollution levels should fall to the previous on-day level. However, I cannot rule out the possibility that such pollution reductions are too costly for the polluters, so that when faced with 1-in-1-day monitoring, the equilibrium average level of pollution is slightly higher than the current on-day level. However, one may also speculate that increased monitoring frequency can induce polluters to adopt more efficient abatement technology or production practices that lead to ambient pollution levels even below the current on-day level.

following day. Similar to the pollution examination in this paper, the mortality effect shows up in counties with 1-in-6-day monitoring but not in counties with 1-in-1-day monitoring. There is also some evidence that emergency room visits for asthma exacerbation ailments respond promptly to the pollution gap. Zou, Miller, and Molitor (2018) show that, compared to the counterfactual scenario in which off-day pollution does not deviate from observed on-day levels, the value of the annual loss of life due to the intermittent monitoring policy is an order of magnitude larger than the roughly $12 million resources saved by reduced monitoring frequency. Below I present two additional examples on the effect of the pollution gap on standardized test scores and criminal activities, following recent literature on the cognitive and behavioral consequences of short-term air pollution fluctuations (e.g., Ebenstein et al., 2016; Herrnstadt et al., 2016).

6.1. The Effect of the Pollution Gap on Test Scores

Background and Data. The test scores examination uses the California High School Exit Exams (CAHSEE) school-level test performance data published by the California Department of Education. CAHSEE was designed and first offered in 2001 to volunteer students. Starting 2004, the passage of both CAHSEE math and English tests became an official diploma requirement. Every academic year, multiple test sessions are administered. Each session is comprised of two consecutive test days, with the English test taking place on the first day, and the math test on the second day. The exact administration (such as number of sessions and test months) varies across years, but most sessions took place in July, October, November, December, and February, March, and May in the next calendar year, and so forth. For both subjects, test performance is expressed as a scale score that ranges from 275 to 450, with 350 being the usual passing score. All students are required to make their first CAHSEE attempt in grade 10, usually during the February and March sessions (called "Grade 10 census"). Students who miss the census may take makeup sessions held in May. In cases of failure, students are allowed to retake the exams in future sessions until both math and English tests were passed. Students are not allowed to retake any exams that they have already passed. As a tradition, most of the tests are scheduled on the first Tuesday or Wednesday of the test month. This scheduling practice creates nice intersection with the EPA's 1-in-6-day monitoring schedule, which is exploited in the analysis below.

I obtain publicly available test performance data for the universe of tests taken from 2004 to 2013. A unit of observation in the data is a school-test, i.e., a CAHSEE math (English) test taken by all students in a given school on a given date. As a privacy protection measure, test scores are masked if they are averaged over the performance of fewer than 10 students. In cases of masked cells, however, I can observe the number of tests taken. The final estimation sample includes test data from about 2,800

schools spanning all 58 counties in California. These test score statistics are aggregated from more than 14 million individual tests taken on 91 different test dates from 2004 to 2013. Table 5 reports test performance statistics. The average math (English) scale score is 367.2 (370.3). As expected, February and March sessions have substantially higher average scores due to the fact that tests in other months are mostly taken by students who did not pass CAHSEE on their first attempt in the February or March census. In the analysis below, I control for compositional difference by including month-of-year fixed effects.

Estimation and Results. As mentioned earlier, because most CAHSEE tests are scheduled on the first Tuesday or Wednesday of the test month, 75 out of 91 tests, i.e., about 4.95 of every 6, coincide with the EPA's monitor 1-in-6-day off-days. However, the scarcity of test dates implies that the interaction between test dates and monitoring dates alone does not ensure strong balance with respect to observable characteristics such as seasonality. In the following, I report two types of specifications. In the first, I use a simple estimation strategy that compares test scores on off-days vs. on-days, including only month-of-year dummies to control for different student composition for tests administered in different months of the year. I then report a much richer specification to ensure balance across various dimensions, controlling for subject fixed effects, school × month-of-year fixed effects, and time period fixed effects (academic year, day-of-year, weekend). The fixed effects are further interacted with decile distance dummies for the schools' distance to the nearest non-attainment PM monitors.

The analysis starts with a simple comparison of off-days vs. on-days test performance across all schools in California. I then extend the analysis by exploiting geographic variations in the pollution gap, and examine whether test scores responses are stronger for schools that are located closer to non-attainment monitors. In all regressions, I weight observations by the number of tests taken in the school-test cell. Standard errors are two-way clustered at the school and the test levels. In addition to test scores, I also examine whether the number of tests taken decreases on monitor off-days, as would be the case if pollution causes sickness among students.

Table 6 reports the test performance results. The upper panel reports results where the outcome variable is standardized (i.e., mean 0 and standard deviation 1) test score. Column 1 reports a simple specification where test scores are regressed on an indicator for taking a test on a monitor off-day, conditional on 12 month-of-year dummies. This specification shows that taking an exam on an off-day reduces test score significantly, by 5.3 percent of a standard deviation. In column 2, I include the full set of controls as described earlier to ensure balance of various test characteristics across on-days and off-days. Although this specification also yields a negative coefficient estimate, it reduces the effect size to

about 2.5 percent of a standard deviation and it is not statistically significant. In columns 3 to 5, I repeat the estimation in column 2, but separately for schools that are close to (< 10 miles) or far away from (10-50 miles and > 50 miles) the nearest non-attainment PM monitor. Results suggest that the effect of off-days is driven by schools close to non-attainment monitors. Column 3 shows that for schools that are within 10 miles of the nearest non-attainment monitor, taking a test on an off-day reduces scores by a statistically significant 6.3 percent of a standard deviation. By contrast, columns 4 and 5 show that no significant responses are detected for schools that are beyond 10 miles from a non-attainment monitor, although the standard errors do not allow me to rule out small effects.

The impact of monitor off-days on test scores is moderate in size. Take the baseline estimate in column 1, for example. On an average off-day, a test score is about 5.3 percent of a standard deviation lower, which is roughly 1.3 points in scale score. The average white-black CAHSEE test score gap from 2004-2013 is 32.15 points. Thus, the effect size of an exam taken on an off-day is about 4 percent of the white-black test score gap. Notably, the effect size is comparable to that from Ebenstein, Lavy, and Roth (2016) who estimate that a standard deviation increase in fine particulate matter leads to a roughly 1.3 percent reduction in matriculation test performance among Isareli students. My pollution and test estimates imply that a standard deviation increase in aerosol pollution corresponds to 2.6 percent reduction in test performance.[29]

In the lower panel of Table 6, I repeat the same estimation, but now test whether fewer tests are being taken on monitor off-days. Note that across columns the sample size is larger than in the test scores regressions because number of tests taken can be observed even for cells where test performance is masked due to privacy protection. Point estimates in columns 1 and 2 suggest that 4.6 percent to 6.8 percent fewer tests are taken on off-days, but the effects are not precisely estimated. Columns 3 through 5 again present estimations separately for schools close to or far away from non-attainment monitors. The analysis provides suggestive evidence of a distance gradient. Column 3 shows that about 11.5 percent fewer tests are taken on an off-day at schools less than 10 miles away from the nearest non-attainment monitor, although the effect is marginally significant. In columns 4 and 5, schools that are farther away

[29] In the Appendix, I show that air quality during off-days is roughly 5 percent worse than during on-days near California high schools from 2001-2013. Because the comparison is ultimately done across days when tests took place, I further restrict the estimation sample to the 91 test days. I find that pollution is about 12 percent higher on an on-day test day than on an off-day test day. This latter estimate is used to compare my effect size with that from Ebenstein, Lavy, and Roth (2016).

show smaller and insignificant responses. Due to the lack of statistical precision, I cannot draw a strong conclusion that fewer students attend tests administered on monitor off-days.

6.2. The Effect of Pollution Gap on Crime

Background and Data. I use crime data from the Federal Bureau of Investigation's National Incident-Based Crime Reporting System (NIBRS) from 2001 to 2013. These data contain detailed crime incident-level information, such as the date, the location of the reporting jurisdiction, and the offense code, reported by jurisdictions that participate in the NIBRS program. Reporting jurisdictions are usually city (or county) law enforcement agencies. The number of participating jurisdictions in the NIBRS has grown over time. By 2013, NIBRS covered a population of about 92 million people in 33 states, and accounted for more than 28 percent of all crime reported to the FBI Uniform Crime Reporting Program. To alleviate concerns about compositional changes in the NIBRS-covered population over time, my analysis is restricted to jurisdictions that have participated in NIBRS for at least 10 years during the study period.[30]

I begin by first constructing daily crime rates at the county level using information on jurisdictions' county location and the population covered. The crime rate is defined as the reported number of crime incidents in the county divided by the population covered by NIBRS within the county.[31] These data are then merged with counties' PM monitoring frequencies. From 2001 to 2013, 356 counties follow the 1-in-6-day sampling schedule (defined as counties where all PM monitors follow the 1-in-6-day schedule) and crime data are available in 47 counties. These counties span 19 different states, and include a total population of 1.33 million people, representing 67 percent of the total population in those counties.

My analysis focuses on three broad groups of criminal activities. Following the FBI's categorization, I create crime rate variables for violent crime (aggravated assault, robbery, forcible rape, murder, and

[30] Despite the growing coverage, crime prevalence in NIBRS-participating jurisdictions is not considered to be representative of overall crime rates. For example, jurisdictions with fewer people are known to be disproportionately more likely to report data (James and Council, 2008). Differences in levels of crime are not necessarily a threat to the identification strategy which uses day-to-day variation in pollution monitoring status within the same area. In unreported results, I confirm that the findings in this section are similar if I relax the restriction on panel balance, or if I use a strictly balanced panel, i.e. counties that consistently report to NIBRS for the entire 2001-2013 period. However, I recognize that sensitivity to air pollution may differ among people living in reporting and non-reporting areas. This is an important caveat regarding external validity throughout the interpretation of results obtained in this section.

[31] Every year, about 10 percent of incidents of crime occur in jurisdictions that cross county borders. In these cases, NIBRS provides estimates on the population that is covered by the jurisdiction in each of the counties the jurisdiction spans. I use this information to assign a given incident of crime to each of the counties, with the assignment probability proportional to the county's population share within the jurisdiction.

nonnegligent manslaughter), property crime (burglary, larceny-theft, motor vehicle theft, and arson), and all other crime. Table 7 presents summary statistics for counties in the estimation sample. These statistics are compared to statistics of the entire NIBRS population.

Estimation and Results. The identification of the causal effect of the monitoring schedule on crime is once again a straightforward comparison of crime rates on monitor off-days vs. on-days. As before, I report results from two polar specifications: one with no controls, and the other with a full array of fixed effects controls (county, year, month-of-year, day-of-week, and day-of-month). Regressions are weighted by the population covered by NIBRS in the county. Standard errors are clustered at the county level.

Table 8, column 1 begins by repeating the pollution analysis. The estimation conceptually corresponds to column 4 of Table 2, whereas now pollution is measured at the county average level. For example, column 1, panel A shows that the off-days vs. on-days pollution gap is estimated to be 1.3 percent. In column 2, the pollution regression is repeated, again using the estimation sample, which yields a pollution gap of 2.7 percent without controls and 1.9 percent with controls. The 95 percent confidence intervals of these estimates overlap with the mean coefficients in column 1.

Columns 3 to 5 document crime effects of the monitoring schedule. I first focus on panel A, which presents raw comparison of off-days vs. on-days means. Column 3 shows that raw comparison across off-days vs. on-days means show that violent crime is about 0.257 per million (NIBRS population) higher on off-days. Based on the daily mean of 15.88 per million, the effect represents a 1.6 percent increase. Whereas previous literature documents little evidence on the effect of pollution on property crime, column 4 reports that property crime also increases by about 1 per million on an off-day. However, the relative size of this effect is smaller than the violent crime effect -- about 0.91 percent out of the daily mean of 110.31 per million. Column 5 continues the analysis with other categories of crime, in which I find no evidence of a precise increase. Panel B repeats the same analysis controlling for geographic and time fixed effects, which yields very similar results.

7. 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 is 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 regulation. I have presented evidence that a widely used once-every-six-day monitoring schedule for outdoor particle pollution causes significant deterioration in air quality on unmonitored days as compared to the levels observable to the air regulator on monitored days. My results reveal the possibility that strategic responses may arise through coordination by the local governments who, by design under the federal regulation, share non-compliance costs. As an illustration of this mechanism, I show that some local governments issue air quality advisories strategically in response to the monitoring schedule, an action that is likely intended to manipulate public behavior such as transportation decisions, and, ultimately, to reduce pollution levels recorded on monitored days.

The key finding of strategic responses is based on an atmospheric measure of air quality. Although I cannot precisely attribute the effect to individual polluting sources, I provide indirect tests of on the mechanisms by which strategic responses may arise. In particular, I show that the intensity of responses differs substantially across geographic regions, depending strongly on the characteristics of local areas' economies and industries. Overall, my results are consistent with the view that strategic responses are more likely to arise where the *incentives* to avoid monitoring are stronger, such as areas subject to non-compliance punishments, and where the *scope* for avoidance is larger, such as areas with high-emission industries that often do not operate on full capacity.

One policy option would be to advance to a monitoring system that operates more continuously. This option is based on the empirical evidence that: strategic responses are weaker in regions that face monitoring once every three days; no detectable strategic response occurs at sites that are monitored daily; and the administrative costs of extra monitoring are unlikely to exceed the resulting health benefits, as determined by the calculations of this paper as well those as in Zou, Miller, and Molitor (2018). The next-best policy option would be to use a randomly determined monitoring schedule. This approach has the potential to remove strategic responses while keeping a similar sampling frequency and, thus, similar administrative costs. But in actual implementation, this approach may be challenging. Because filter-based PM monitoring requires intensive field work ranging from sampling to subsequent laboratory analyses, it can be problematic for state and local agencies to carry out these activities with short notice. Alternatively, a random schedule can be generated ex-ante, e.g., at the beginning of the year, and be handed over to the state and local agency for implementation. However, in light of this paper's findings on local government coordination, it is unclear whether the ex-ante schedule would be strictly hidden from the polluters.

References

Akland, Gerald G. "Design of sampling schedules." *Journal of the Air Pollution Control Association* 22, no. 4 (1972): 264-266.

Auffhammer, Maximilian. “VW’s Deepwater Horizon?” UC Berkeley Energy Institute Blog. September 21, 2015

Auffhammer, Maximilian, Antonio M. Bento, and Scott E. Lowe. "Measuring the effects of the Clean Air Act Amendments on ambient PM10 concentrations: The critical importance of a spatially disaggregated analysis." *Journal of Environmental Economics and Management* 58, no. 1 (2009): 15-26.

Becker, Randy, and Vernon Henderson. "Effects of air quality regulations on polluting industries." *Journal of Political Economy* 108, no. 2 (2000): 379-421.

Bombardini, Matilde, and Bingjing Li. "Trade, pollution and mortality in china." No. w22804. National Bureau of Economic Research, 2016.

Chay, Kenneth Y., and Michael Greenstone. "Does air quality matter? Evidence from the housing market." *Journal of Political Economy* 113, no. 2 (2005): 376-424.

Chen, Yuyu, Ginger Zhe Jin, Naresh Kumar, and Guang Shi. "The promise of Beijing: Evaluating the impact of the 2008 Olympic Games on air quality." *Journal of Environmental Economics and Management* 66, no. 3 (2013): 424-443.

Cropper, Maureen, Arthur Fraas, and Richard Morgenstern. "Looking backward to move regulations forward." *Science* 355, no. 6332 (2017): 1375-1376.

Cutter, W. Bowman, and Matthew Neidell. "Voluntary information programs and environmental regulation: Evidence from ‘Spare the Air’." *Journal of Environmental Economics and Management* 58, no. 3 (2009): 253-265.

Donaldson, Dave, and Adam Storeygard. "The view from above: Applications of satellite data in economics." *Journal of Economic Perspectives* 30, no. 4 (2016): 171-198.

Duncan, Bryan N., Ana I. Prados, Lok N. Lamsal, Yang Liu, David G. Streets, Pawan Gupta, Ernest Hilsenrath et al. "Satellite data of atmospheric pollution for US air quality applications: Examples of applications, summary of data end-user resources, answers to FAQs, and common mistakes to avoid." *Atmospheric Environment* 94 (2014): 647-662.

Ebenstein, Avraham, Victor Lavy, and Sefi Roth. "The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution." *American Economic Journal: Applied Economics* 8, no. 4 (2016): 36-65.

Foster, Andrew, Emilio Gutierrez, and Naresh Kumar. "Voluntary compliance, pollution levels, and infant mortality in Mexico." *American Economic Review* 99, no. 2 (2009): 191-197.

Fowlie, Meredith, Mar Reguant, and Stephen P. Ryan. "Market-based emissions regulation and industry dynamics." *Journal of Political Economy* 124, no. 1 (2016): 249-302.

Gates, Guilbert, Jack Ewing, Karl Russell, and Derek Watkins. 2016. Explaining Volkswagen's emission scandal. New York Times. July 19, 2016.

Gilbert, Richard O. Statistical methods for environmental pollution monitoring. John Wiley & Sons, 1987.

Giles, Cynthia. "Next generation compliance." *The Environmental Forum*, pp. 22-26. 2013.

Graff Zivin, Joshua, and Matthew Neidell. "Days of haze: Environmental information disclosure and intertemporal avoidance behavior." *Journal of Environmental Economics and Management* 58, no. 2 (2009): 119-128.

Grainger, Corbett, Andrew Schreiber, and Wonjun Chang. "How sates comply with federal regulations: Strategic ambient pollution monitoring." Working paper. 2017.

Gray, Wayne B., and Jay P. Shimshack. "The effectiveness of environmental monitoring and enforcement: A review of the empirical evidence." *Review of Environmental Economics and Policy* 5, no. 1 (2011): 3-24.

Greenstone, Michael. "The impacts of environmental regulations on industrial activity: Evidence from the 1970 and 1977 clean air act amendments and the census of manufactures." *Journal of Political Economy* 110, no. 6 (2002): 1175-1219.

Greenstone, Michael. "Eliminating unnecessary and costly red tape through smarter regulations". Testimony to U.S. Congress Joint Economic Committee. June 26, 2013.

Greenstone, Michael, John A. List, and Chad Syverson. "The effects of environmental regulation on the competitiveness of US manufacturing." No. w18392. National Bureau of Economic Research, 2012.

Hanna, Rema. "US environmental regulation and FDI: evidence from a panel of US-based multinational firms." *American Economic Journal: Applied Economics* 2, no. 3 (2010): 158-189.

Henderson, Vernon. "Effects of air quality regulation". *American Economic Review* 86, no. 4 (1996): 789-813.

Herrnstadt, Evan, Anthony Heyes, Erich Muehlegger, and Soodeh Saberian. "Air Pollution as a Cause of Violent Crime: Evidence from Los Angeles and Chicago." Working paper. 2016.

Ichoku, Charles, Ralph Kahn, and Mian Chin. "Satellite contributions to the quantitative characterization of biomass burning for climate modeling." *Atmospheric Research* 111 (2012): 1-28.

Kaufman, Y. J., D. Tanré, L. Ao Remer, E. F. Vermote, A. Chu, and B. N. Holben. "Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer." *Journal of Geophysical Research: Atmospheres* 102, no. D14 (1997): 17051-17067.

Kittaka, Chieko, James Szykman, Brad Pierce, Jassim Al-Saadi, Doreen Neil, Allen Chu, Lorraine Remer, Elaine Prins, and John Holdzkom. "Utilizing MODIS satellite observations to monitor and analyze fine particulate matter, PM2. 5, transport event." In Conference on Atmospheric Chemistry, 6th: Air Quality in Megacities, Seattle, WA, pp. 11-15. 2004.

Lee, Hyung Joo, Brent A. Coull, Michelle L. Bell, and Petros Koutrakis. "Use of satellite-based aerosol optical depth and spatial clustering to predict ambient PM2.5 concentrations." *Environmental Research* 118 (2012): 8-15.

Liu, Yang, Meredith Franklin, Ralph Kahn, and Petros Koutrakis. "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, no. 1 (2007): 33-44.

Muller, Nicholas Z., and Paul A. Ruud. "What forces dictate the design of pollution monitoring networks?" *Environmental Modeling & Assessment* 23, no. 1 (2018): 1-14.

Neidell, Matthew. "Information, Avoidance behavior, and health the effect of ozone on asthma hospitalizations." *Journal of Human Resources* 44, no. 2 (2009): 450-478.

Nehls, Gerald J., and Gerald G. Akland. "Procedures for handling aerometric data." *Journal of the Air Pollution Control Association* 23, no. 3 (1973): 180-184.

Oliva, Paulina. "Environmental regulations and corruption: Automobile emissions in Mexico City." *Journal of Political Economy* 123, no. 3 (2015): 686-724.

Remer, Lorraine A., Y. J. Kaufman, D. Tanré, S. Mattoo, D. A. Chu, J. Vanderlei Martins, R-R. Li et al. "The MODIS aerosol algorithm, products, and validation." *Journal of the Atmospheric Sciences* 62, no. 4 (2005): 947-973.

Reynaert, Mathias, and James M. Sallee. "Corrective Policy and Goodhart's Law: The Case of Carbon Emissions from Automobiles." No. w22911. National Bureau of Economic Research, 2016.

Rumburg, Brian, Richard Alldredge, and Candis Claiborn. "Statistical distributions of particulate matter and the error associated with sampling frequency." *Atmospheric Environment* 35, no. 16 (2001): 2907-2920.

Ruminski, Mark, Shobha Kondragunta, Roland Draxler, and Jian Zeng. "Recent changes to the hazard mapping system." In Proceedings of the 15th International Emission Inventory Conference, pp. 15-18. 2006.

Schlenker, Wolfram, and W. Reed Walker. "Airports, air pollution, and contemporaneous health." *Review of Economic Studies* 83, no. 2 (2015): 768-809.

Shimshack, Jay P. "The economics of environmental monitoring and enforcement." *Annual Review of Resource Economics* 6, no. 1 (2014): 339-360.

Tibshirani, R., 1996. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B* (*Methodological*), pp.267-288.

United States Environmental Protection Agency. 1985. "Guideline for PM10 monitoring and data reporting."

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. "Memorandum: Guidelines for granting schedule exemptions for PM10 monitoring."

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. 1999. "Guideline on data handling conventions for the PM NAAQS. EPA-454/R-99-008."

United States Environmental Protection Agency. 2013. "Technical assistance document for the reporting of daily air quality - the Air Quality Index (AQI)." EPA-454/B-13-001.

United States Public Health Service. Air Pollution Measurements of the National Air Sampling Network." Analysis of Suspended Particulate Samples Collected 1061 (1957)

Voiland, Adam. "Aerosols: Tiny Particles, Big Impact: Feature Articles." (2010).

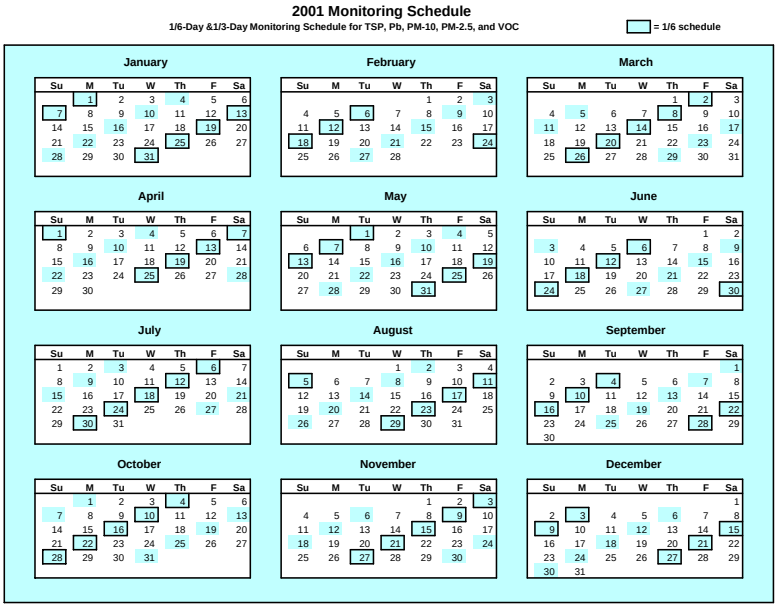
Vollaard, Ben. "Temporal displacement of environmental crime: Evidence from marine oil pollution." *Journal of Environmental Economics and Management* 82 (2017): 168-180.

Walker, W. Reed. "The transitional costs of sectoral reallocation: Evidence from the Clean Air Act and the workforce." *Quarterly Journal of Economics* 128, no. 4 (2013): 1787-1835.

Zhang, Ying, and Zhengqiang Li. "Remote sensing of atmospheric fine particulate matter (PM2.5) mass concentration near the ground from satellite observation." *Remote Sensing of Environment* 160 (2015): 252-262.

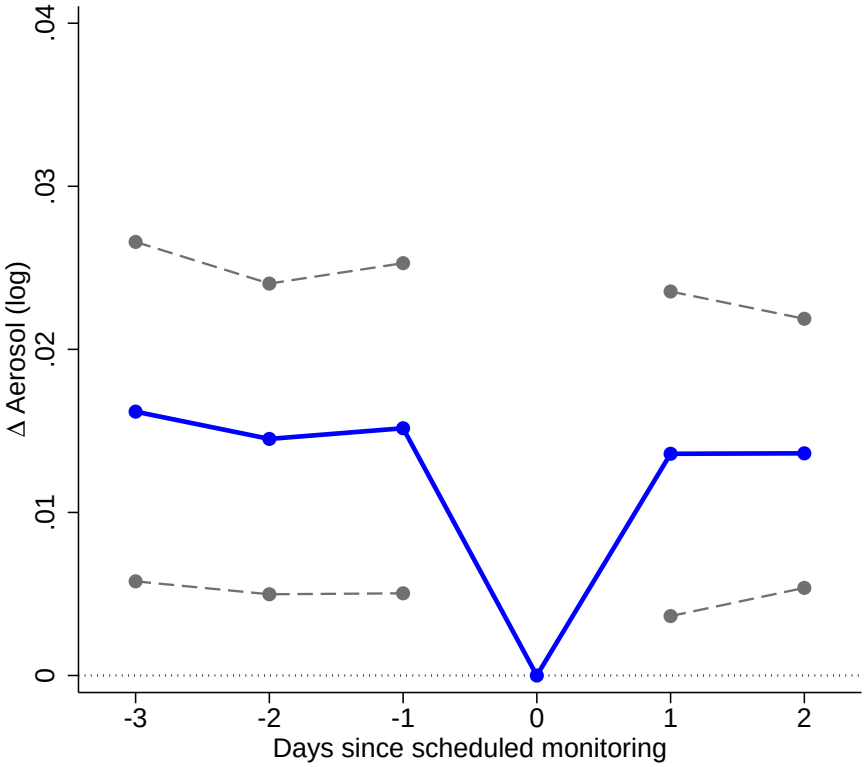
Zou, Eric, Nolan Miller, and David Molitor. "Sampling Meets Big Data: Measuring the Unintended Consequences of Intermittent Air Quality Monitoring." Work in progress. 2018.

Figure 1. EPA's Ambient Pollution Monitoring Schedule, 2001



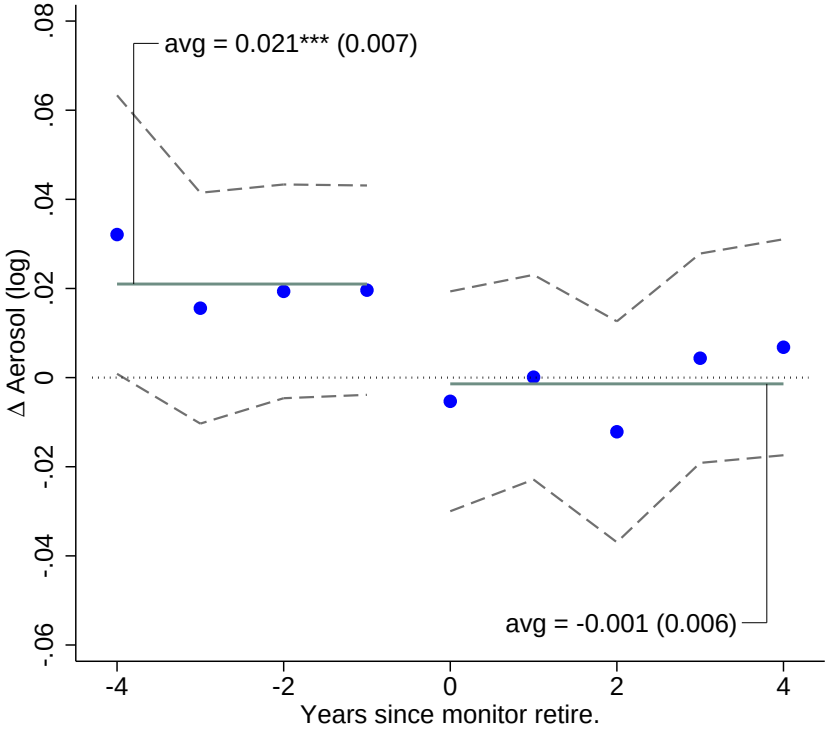
Notes: This figure shows the EPA's 2001 monitoring schedule calendar. Full archives of all calendars can be found here: <https://www3.epa.gov/ttn/amtic/calendar.html>.

Figure 2. Event Study: Off-days vs. On-days Pollution Gap



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 sampling day, which is normalized to 0. The regression is not conditional any covariates. Dashed lines represent 95% confidence intervals constructed using standard errors clustered at the county level.

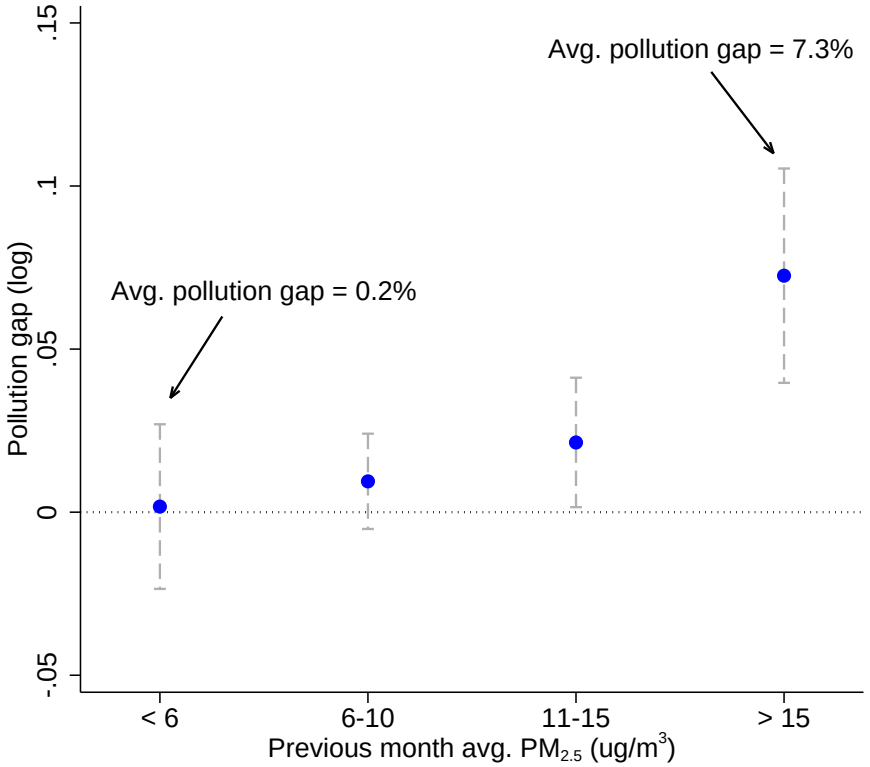
Figure 3. Placebo Test: Pollution Gap by Years Relative to Monitoring Site Retirement



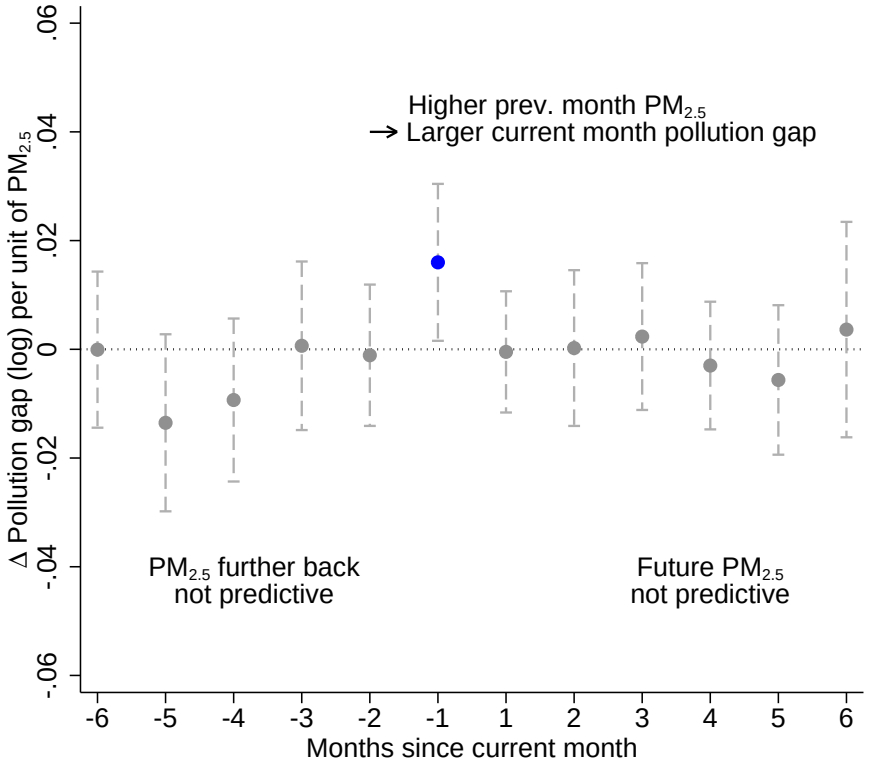
Notes: This figure plots the pollution gap between monitored (on-days) and unmonitored days (off-days) as a function of years relative to the retirement of a site that conducted monitoring every six days. Average estimates show off-days effect separately estimated before and after the retirement of the 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% confidence intervals constructed using standard errors clustered at the county level.

Figure 4. Heterogeneous Pollution Gap by County's Average $PM_{2.5}$ Level

Panel A. By Bins of Previous Month's $PM_{2.5}$ Level

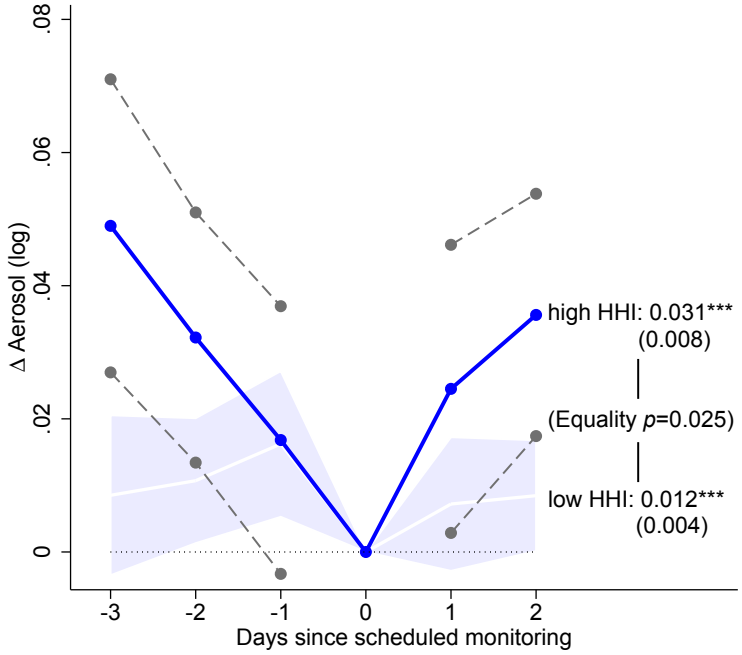


Panel B. By Leads and Lags of Monthly $PM_{2.5}$ Levels



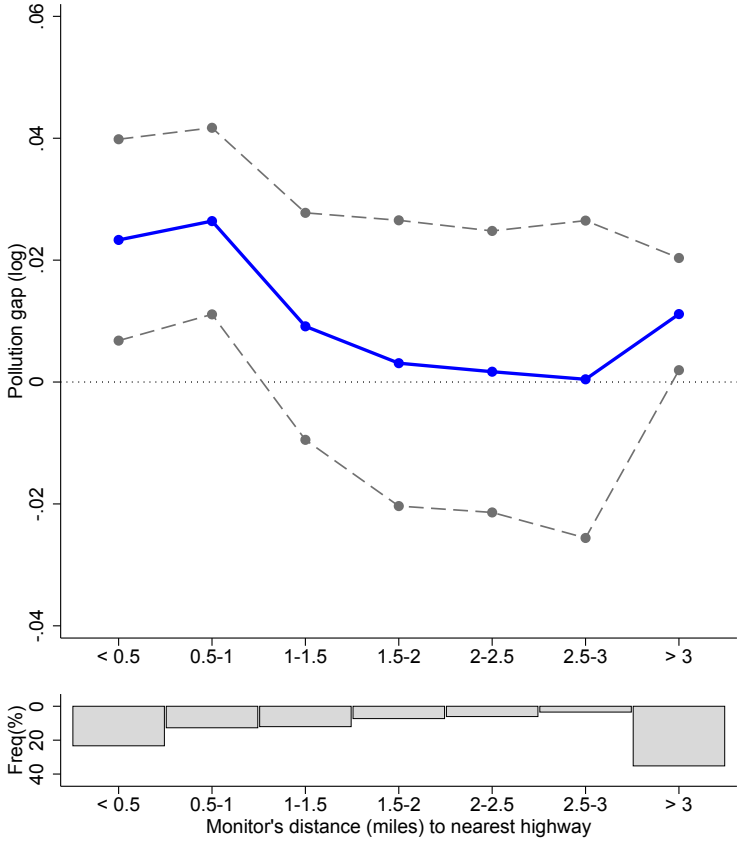
Notes: Panel A reports the interaction of the pollution gap with bins of realized $PM_{2.5}$ in the past month, controlling for interactions with other five lags and all six leads. Panel B reports the interaction of the 1-in-6-day pollution gap with six lags and six leads of the county's realized *mean* $PM_{2.5}$ concentration. The highlighted estimate at month -1 in panel B therefore corresponds to the slope across the estimates in panel A. Both regressions include fixed effects dummies (site, year, month-of-year, and day-of-week) and weather controls. In both panel, dashed range bars plot 95% confidence intervals constructed using standard errors clustered at the county level.

Figure 5. Heterogeneous Pollution Gap by County's Emission Herfindahl-Hirschman Index



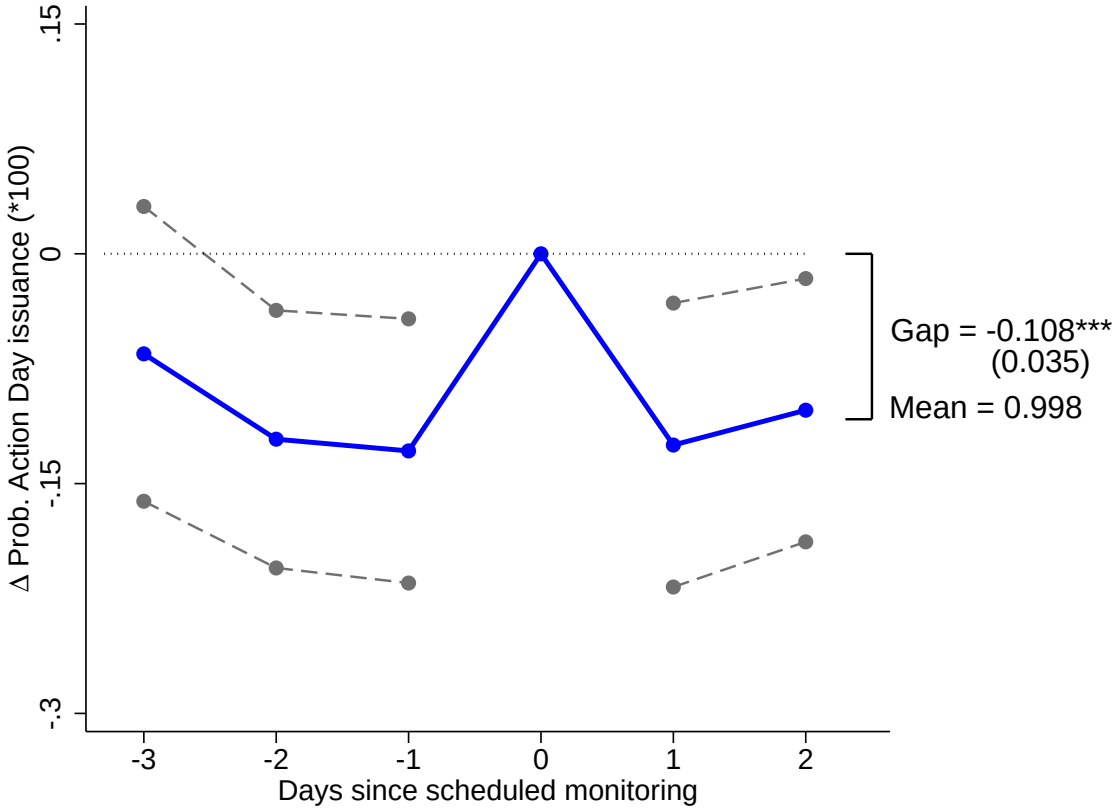
Notes: The figure displays 1-in-6-day pollution pattern separately for HHI≥ 0.9 (foreground graph objects) vs. HHI< 0.9 counties (background graph objects). Regressions include fixed effects dummies (site, year, month-of-year, and day-of-week) and weather controls. Dashed lines and the shades represent 95% confidence interval constructed from standard errors clustered at the county level. "Equality p" corresponds to the null hypothesis that there is no difference in the pollution gap between the two groups.

Figure 6. Heterogeneous Pollution Gap by Monitor's Distance to Highway



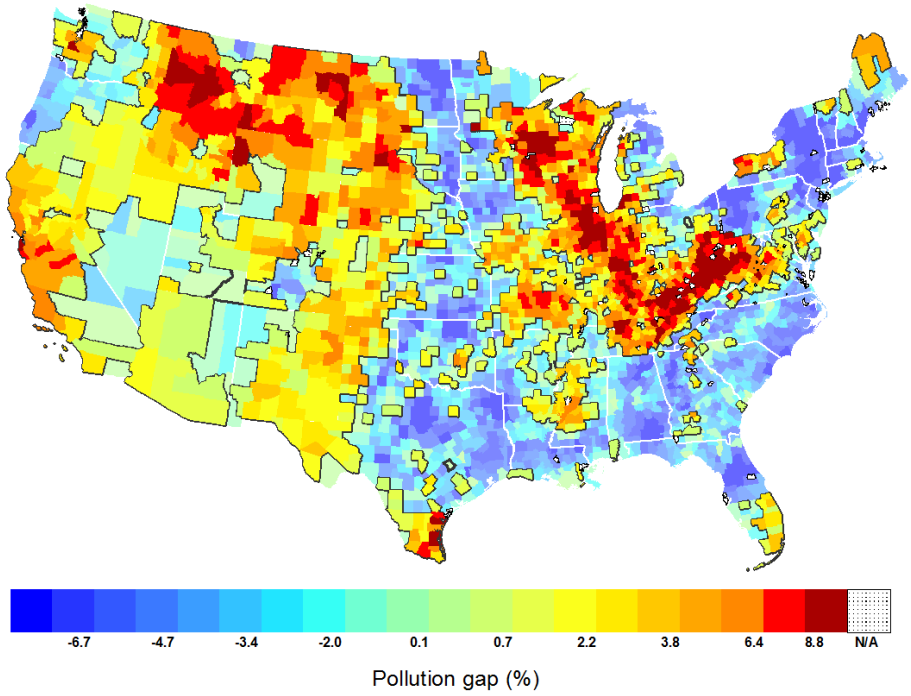
Notes: The figure plots interaction of pollution gap with the 1-in-6-day PM monitor's distance (bins) to the nearest highway. The group "$>$ 3" pools all monitors that fall more than 3 miles from the nearest highway. Regressions include fixed effects dummies (site, year, month-of-year, and day-of-week) and weather controls. Dashed lines represent 95% confidence intervals constructed using standard errors clustered at the county level.

Figure 7. Local Governments' Strategic "Pollution Action Day" Declarations



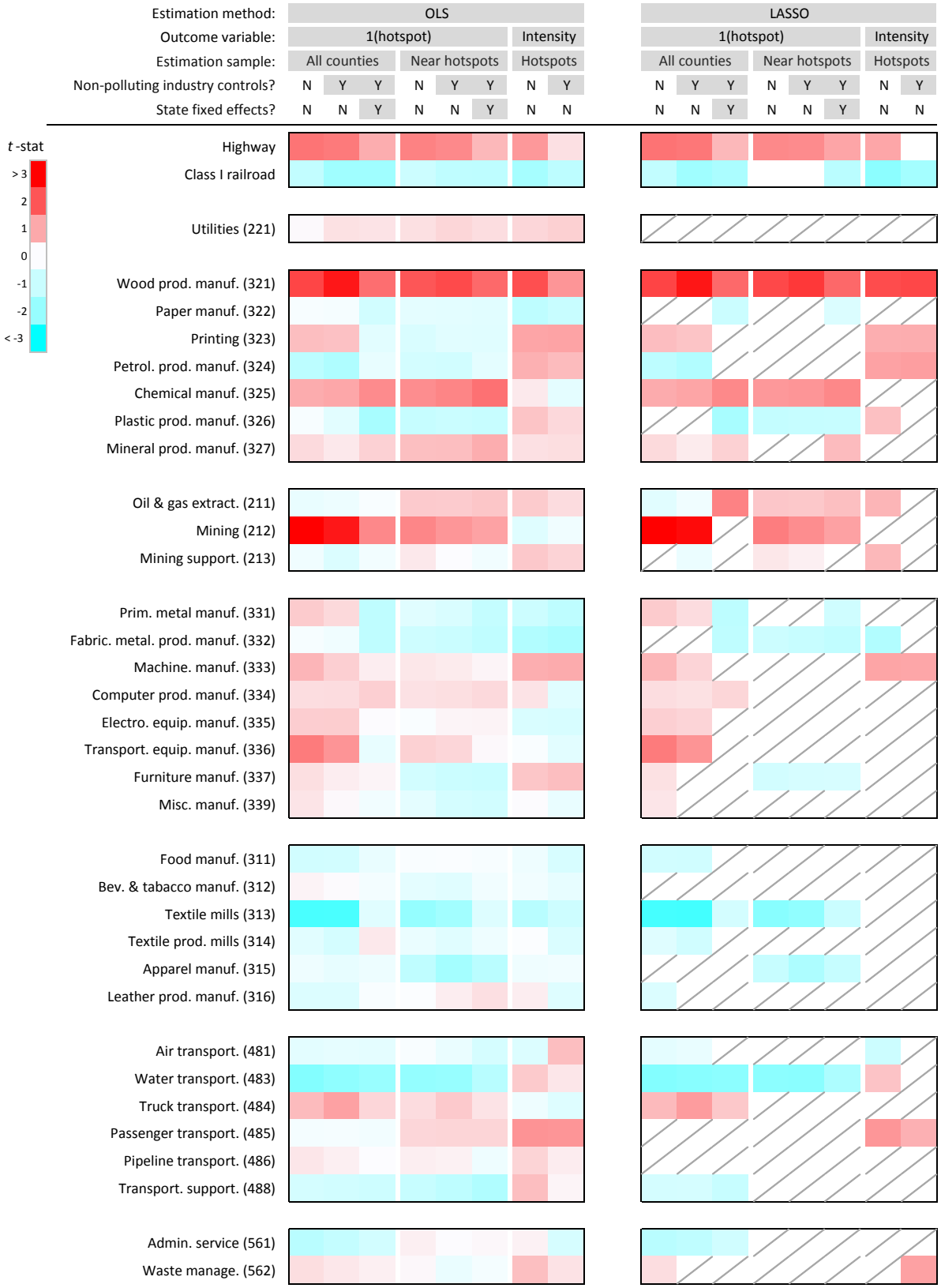
Notes: The outcome variable is core-based statistical area (CBSA) × daily dummy for whether any action day is issued. The sample spans 2004-2013 and includes 14,945 issuances across 171 CBSAs. In cases of issuances that span a consecutive number of days, only the first day of issuance is counted. See text for more detail. Dashed lines represent 95% confidence intervals constructed using standard errors clustered at the CBSA level.

Figure 8. County-Level 1-in-6-Day Pollution Gap



Notes: The map plots county-level 1-in-6-day pollution gap estimates. Red (blue) areas correspond to higher (lower) aerosol levels during off-days. For all counties, off-days and on-days are defined using EPA's monitoring schedule. The "N/A" category corresponds to counties with fewer than 6,000 available aerosol observations.

Figure 9. Industry Correlates of Pollution Gap Hotspots



Notes: Each column represents t-statistics from a single regression, with model specifications laid out in the header. Two-digit NAICS industry blocks are ordered by its share of national total PM_{10} emissions according to EPA's 2011 National Emissions Inventory. In the interest of space, the graph only reports results for polluting sectors, i.e. those contribute at least 1% of total PM_{10}. Redness (blueness) indicates that the county being an industrial hot spot positively (negatively) predicts the likelihood of the county being a 1-in-6-day pollution gap hot spot.

Table 1. Summary Statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	aerosol stats ($\times$ 100)				monitor stats			site stats		county stats			
	p25	mean	p75	Total observations	*N*	*N* 1/6day	*N* 1/6day > NAAQS	*N*	*N* 1/6day	*N*	pop (million)	*N* 1/6day	pop 1/6day (million)
Year													
2001	3.0	11.4	15.6	9,337,915	1,816	876	52	1,343	776	697	193.2	429	142.6
2002	3.3	14.1	18.1	9,301,308	1,933	889	57	1,413	774	725	206.3	428	143.3
2003	3.2	12.6	17.0	9,513,247	1,775	831	63	1,322	725	676	200.3	399	142.4
2004	2.6	10.7	14.7	9,322,004	1,865	850	25	1,377	737	694	204.9	403	149.6
2005	2.8	12.2	16.1	10,160,251	1,792	785	60	1,311	679	653	199.7	362	141.2
2006	3.2	12.4	16.4	10,174,374	1,817	828	86	1,296	690	657	202.8	368	141.2
2007	3.7	14.5	18.7	10,295,391	1,761	740	126	1,262	600	652	208.2	322	138.8
2008	3.7	12.3	16.7	10,228,175	1,628	663	73	1,158	538	606	204.6	287	133.9
2009	3.8	11.4	15.7	9,469,757	1,728	681	39	1,199	537	633	210.1	297	132.2
2010	3.3	10.5	14.2	10,078,748	1,702	659	38	1,159	522	619	209.3	285	132.0
2011	4.5	13.4	17.2	10,298,153	1,585	579	39	1,077	458	578	200.0	262	124.8
2012	4.0	13.3	17.8	10,806,884	1,629	540	20	1,088	426	582	195.4	253	114.5
2013	3.0	11.6	15.6	9,079,310	1,699	522	22	1,109	406	596	203.4	229	112.9

Notes: Each row represents statistics for a calendar year. Columns 1, 2, and 3 show 25th-percentile, mean, and 75th-percentile values of grid-daily level aerosol concentration. Column 4 shows total number of grid-daily observations. The monitor sample includes all monitors that collected enough samples during the year to be considered eligible for NAAQS comparison. Column 5 reports the total number of particulate matter (PM) monitors, including all $PM_{2.5}$ and PM_{10} monitors that are eligible for NAAQS comparison (see the text for more detail). Column 6 counts the number of 1-in-6-day monitors, defined as monitors that are required to sample either 60 or 61 days of PM data for each calendar year. Column 7 counts the number of 1-in-6-day monitors that exceeded any PM standard in that year (but not necessarily violating the NAAQS, as violation is based on 3-year average values). Column 8 aggregates monitor counts from column 1 to the monitoring site level, acknowledging the fact that there might be multiple PM monitors within the same monitoring sites. Column 9 counts the number of sites that contain at least one PM monitor that follows the 1-in-6-day schedule. Columns 10 and 12 aggregate site counts in column 8 and 9 to the county level, respectively. Columns 11 and 13 report the corresponding populations that lived in the monitored counties.

Table 2. Off-days vs. On-days Pollution Gap: 1-in-6-Day Monitoring Sites

Dep. var. = Aerosol concentration (log)				
	(1) Sample: sites w. any 1/6d monitor	(2) Sample: sites w. any 1/6d monitor	(3) Sample: sites w. only 1/6d monitor	(4) Sample: counties w. only 1/6d monitor
1(off-days)	0.016***	0.016***	0.018***	0.018***
	(0.004)	(0.004)	(0.004)	(0.006)
Ctrls		✓	✓	✓
N	685,060	685,060	427,846	176,225
N (site)	1,193	1,193	899	489

Notes: Each column reports a separate regression. The column names indicate the sample used. Columns 1 and 2 use all sites that have at least one 1-in-6-day PM monitor. Column 3 includes sites that have standalone 1-in-6-day monitor. Column 4 includes sites in counties with only 1-in-6-day monitors. Controls include fixed effects dummies (site, year, month-of-year, and day-of-week) and weather controls. Standard errors are clusterd at the county level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Table 3. Off-days vs. On-days Pollution Gap: "Placebo" Sites

Dep. var. = Aerosol concentration (log)			
	(1) Sample: retired 1/6d sites	(2) Sample: 1/1d sites	(3) Sample: Non-PM 1/6d sites (HAPs)
1(off-days)	-0.0020	0.0023	0.0023
	(0.0046)	(0.0080)	(0.0044)
Ctrls	✓	✓	✓
N	372,989	231,532	370,020
N (site)	490	556	792

Notes: Each column reports a separate regression. The column names indicates the sample used. Column 1 includes areas that had 1-in-6-day PM monitoring sites that retired. Column 2 includes 1-in-1-day sites. Column 3 includes 1-in-6-day HAPs sites. Controls include fixed effects dummies (site, year, month-of-year, and day-of-week) and weather controls. A simple power calculation suggests that the power of detecting a 1.5% mean difference between off-days vs. on-days at a 5% significance level is 94% (80%, 91%) for column 1 (2, 3). Standard errors are clusterd at the county level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Table 4. Pollution gap estimates by "Action Day" declaration

Dep. var. = Aerosol (log)	(1)	(2)
1(*off-days*) × 1(warning)	0.069***	0.051***
	(0.014)	(0.013)
1(*off-days*) × 1(no warning)	0.011**	0.013***
	(0.005)	(0.005)
1(*off-days*) × 1(no "Action Day" program)	0.011*	0.016***
	(0.011)	(0.006)
Ctrls		✓
N	685,060	685,060

Notes: Each column represents a separate regression. "1(warning)" and "1(no warning)" indicate whether the 1-in-6-day cycle includes an Action Day issuance or not. "1(no "Action Day" program)" indicate monitors that live in counties that had never issued any Action Day warnings from 2004 to 2013. Coefficient estimates on the group main effects are not reported in the interest of space. Controls include fixed effects (site, year, month-of-year, day-of-week), daily temperature bins, precipitation, and wind speed bins. Standard errors are clustered at the county level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Table 5. Summary Statistics: California High School Exit Exam (CAHSEE) Scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	English				Math			
	Overall	Feb-Mar	May	Other	Overall	Feb-Mar	May	Other
Scale score	367.2	375.2	338.2	341.8	370.3	378.3	342.2	344.1
(275-450)	[24.2]	[21.2]	[14.1]	[10.8]	[24.7]	[21.9]	[12.4]	[10.3]
Number of tests	52	15	10	27	59	19	7	33
Number of tests taken	8,370,191	6,411,603	403,451	1,555,137	8,301,056	6,383,758	388,841	1,528,457

Notes: Statistics are computed from school-subject-daily level data. Columns 1 to 4 (5 to 8) present statistics for English (math) tests. Columns 1 and 5 report overall statistics. The remaining columns report statistics by month of test administration. Standard deviations are reported in brackets.

Table 6. 1-in-6-Day Gap: Standardized Test Performance (CAHSEE, California)

	(1)	(2)	(3)	(4)	(5)
	Sample:		Sample:		
			Schools close to natt. sites		
	All schools		0-10 miles	10-50 miles	>50 miles
Panel A. Dep. var. = Scale score (std.)					
1(off-days)	-0.053**	-0.025	-0.063**	-0.009	0.011
	(0.024)	(0.026)	(0.032)	(0.030)	(0.024)
N	122,540	116,922	34,537	45,948	36,429
N (dates)	91	91	91	91	91
Panel B. Dep. var. = Number of tests taken (log)					
1(off-days)	-0.046	-0.068	-0.115*	-0.084	-0.011
	(0.052)	(0.046)	(0.060)	(0.052)	(0.042)
N	206,519	189,555	49,296	75,650	64,602
N (dates)	91	91	91	91	91
Test month-of-year ctrls.	✓				
Full ctrls.		✓	✓	✓	✓

Notes: Each cell corresponds to a separate regression. Regressions in columns 1 and 2 include all schools. Columns 3, 4 and 5 break down the sample by the distance between the school and the closest non-attainment monitor. In panel A, the outcome variables are standardized scale scores. In panel B, the outcome variable is the logged number of test takers. Standard errors are two-way clustered at the school and the exam date levels. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Table 7. Summary Statistics: Daily Crime Rates

	(1)	(2)
	Estimation sample	NIBRS population
Violent crime (per million)	15.88	11.93
	[14.99]	[23.54]
Property crime (per million)	110.31	97.92
	[55.26]	[59.87]
Other crime (per million)	118.33	92.36
	[69.72]	[116.11]
Number of counties	47	404

Notes: Column 1 presents statistics for counties included in the estimation, i.e. counties with 1-in-6-day sampling and with reported crime data for at least 10 years to NIBRS over the 2001-2013 period. Column 2 presents statistics for all counties that ever reported to NIBRS during the same period. Standard deviations are reported in brackets.

Table 8. 1-in-6-Day Gap: Crime

	(1)	(2)	(3)	(4)	(5)
	Sample:			Sample:	
	Counties w. 1/6d sampling	NIBRS counties w. 1/6d sampling			
Dep. var.	Aerosol (log)	Aerosol (log)	Violent crime (per million)	Property crime (per million)	Other crime (per million)
Panel A. No ctrls.					
1(off-days)	0.013***	0.027**	0.257**	1.004**	0.284
	(0.004)	(0.013)	(0.104)	(0.426)	(0.417)
Panel B. Full ctrls.					
1(off-days)	0.009**	0.019	0.255**	0.960**	0.255
	(0.004)	(0.011)	(0.110)	(0.416)	(0.417)
Mean dep. var.			15.88	110.31	118.33
N	224,847	25,981	68,666	68,666	68,666
N (county)	356	47	47	47	47

Notes: Each cell corresponds to a separate regression. Estimation samples are restricted to counties with no high frequency PM monitors. Column 1 replicates the main pollution regression at the county level. Column 2 is the pollution regression restricted to counties included in the crime regressions. Columns 3 to 5 present crime regression results. Panel A reports estimation with no covariates. Panel B reports estimations with full set of controls, including county fixed effects, year fixed effects, month-of-year fixed effects, day-of-week fixed effects, and day-of-month fixed effects. Standard errors are clustered at the county level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.