Unwatched Pollution: The Effect of Incomplete Monitoring on Air Quality[1]

Eric Zou

March 2017

PRELIMINARY; COMMENTS WELCOME!

Abstract

Modern polluters occasionally reveal their surprising ability to temporarily hide polluting activities during regulatory testing. Such capability is not envisioned by conventional wisdom, which considers polluters' regulation avoidance to be a gradual process of locational or technological adjustments. I examine whether this capability exists among a broader set of polluters by evaluating the impact of an announced ambient air pollution sampling schedule on air quality. I study the U.S. Environmental Protection Agency's cyclical monitoring schedule that samples particulate matter air pollution once every six days in hundreds of sites throughout the country. In the first part of the paper, I use a satellite measure to show that air quality near monitoring sites is significantly worse during days when pollution monitors are scheduled-off. Such "off-days" vs. "on-days" pollution gap can be traced back to industrial sources, especially for those that are close to out-of-compliance monitors. I further evaluate the possibility of local government coaching, relating geographic heterogeneity in the pollution gap to state government characteristics such as government efficiency and resources devoted to air compliance. In the second part of the paper, I exploit the pollution variation driven by the monitoring schedule and examine its impact on cognitive and behavioral responses. I show that the pollution gap coincides with poorer standardized test performance and higher crime rates during monitor off-days.

[1] Department of Economics, University of Illinois at Urbana-Champaign. Email: yzou9@illinois.edu. I am extremely grateful to Tatyana Deryugina, Don Fullerton, Nolan Miller, and David Molitor for invaluable guidance and support. This paper has also benefited from comments and suggestions by David Albouy, Max Auffhammer, Dan Bernhardt, Mark Borgschulte, Olivier Deschenes, Eliza Forsythe, Corbett Grainger, Joseph Shapiro, and seminar participants at the University of Illinois Applied Microeconomics Research Lunch, and the University of Illinois Program in Environmental and Resource Economics. All errors are mine.

1. Introduction

Within environmental regulations, polluters' regulation avoidance is usually considered as a gradual process of adjustments such as relocating of businesses and changes in production technologies. Recent experience defies this intuition. Consider, for example, the Volkswagen emission scandal where vehicles are programmed to dramatically reduce NOx emission during regulatory testing (Gates, Ewing, Russell, and Watkins, 2016). The example highlights the possibility that modern polluters are able to avoid regulations by temporarily hiding polluting activities without making any long term adjustments to technologies. It remains unclear, however, whether this capability extends to a more general set of polluters, in which case the regulators' failure to envision such capability may lead to ineffective environmental monitoring.

In this paper, I examine the impact of an announced ambient air pollution sampling schedule on air quality. I study the U.S. Environmental Protection Agency's (EPA) cyclical air pollution monitoring schedule that samples particulate matter pollution once every six days in hundreds of sites throughout the country. At least three features motivates the possibility of strategic polluting behavior ("gaming") against the monitoring schedule. First, the schedule is publicly available. Every year, the EPA produces a monitoring calendar which clearly highlights the dates when sampling is scheduled. The calendar is published online before the beginning of the calendar year. For example, Figure 1 shows the sampling calendar for 2001, published on the EPA's website in December 2000. Second, sampling results play a key role in deciding compliance toward the national air quality standards. Designation of violation status is usually based on very few observations of standard-exceedance, providing no "second chance" for compliance once a badly polluted day is realized. Finally, non-compliance comes with tremendous costs for both the state governments and the polluters.

The paper is divided into two parts. The first part presents evidence on the causal impact of the EPA's incomplete monitoring schedule on air quality. The empirical analysis uses a satellite air pollution measure to investigate differences in pollution concentration near monitors on days when the monitors are scheduled on ("on-days") vs. days scheduled off ("off-days"). As will be discussed in detail below, since the schedule follows strict once every six days ("1/6day") cycles, comparison of pollution concentrations across off-days and on-days within a typical 1/6day cycle reveals the causal impact of the schedule on air pollution.

The main finding is previewed in Figure 2, which plots the average pollution time path within a typical 6-day monitoring cycle near monitors that follow the 1/6day schedule using data from 2001 to 2013. I use day 0 to mark the on-day where air pollution level is normalized to zero. The graph therefore traces out logged changes in air quality during the off-days relative to the on-day. The graph features a stable pollution path except for a sharp drop during the on-day. On average, the pollution "gap" between the off-days and the on-day is 1.6% in magnitude. As will be discussed in further detail below, the pollution gap closes when monitors retire. Further internal validity checks also confirm that no such gap is observed near monitors where no gaming is expected, such as monitors that sample air quality every day.

I explore the sources of gaming in three steps. First, I show that the 1/6day pollution gap is larger near monitors with a higher potential to violate the EPA's air quality standards, providing confirmation to the hypothesis that gaming is driven by the incentive to avoid regulatory punishments. Next, I show that the observed pollution gap near pollution monitors is likely contributed by industrial sources, near which I find a similar 1/6day pollution cycle that echoes the monitoring schedule. This effect also exhibits a clear gradient with respect to polluting facilities' distance to the nearest non-compliance monitor. Finally, I evaluate whether gaming is entirely self-coordinated among individual polluters. From various perspectives I find the answer to be negative. Although gaming concentrates in areas with major polluters, areas where total emission is scattered among multiple polluters also exhibit modest gaming. This motivates an examination of the possibility of a coaching role played by local governments, who share the cost of non-compliance. I first present an illustrative example where government coaching is directly observed: using records on the universe of air quality public advisories issued by local governments, I show that these advisories, which call for citizens to reduce energy and automobile usage to prevent air quality deterioration, are issued more frequently during the monitor on-days. I then report an analysis which relates gaming to local political environment, finding strong correlations between gaming and state government characteristics such as government efficiency, environmental preferences, and resources devoted to air quality compliance.

In the second part of the paper, I exploit the air quality variation driven by the monitoring schedule and study its impact on cognitive and behavioral responses. I present two examples. First, I use school level data from the California High School Exit Exam (CAHSEE) and show that when an exam date steps on a monitor off-day, standardized exam performance is significantly worse than if an exam that takes place on an on-day. In my setting, the effect size represents about 4.4% of the Black-White test performance gap. Second, using FBI's crime data for a subset of counties across the U.S., I show that violent crime rate is 1.6% higher on off-days relative to on-days. I also report a modest effect on property crime (< 1%) and no effect on other crimes.

(IN PROGRESS) This study is related to the literature on the consequences of incomplete air regulation, which has been focusing on medium-run and long-run substitution of regulated activities toward unregulated sectors (Gray and Shimshack, 2011; Shimshack, 2014). Examples include increased siting of polluting plants in areas with less focused air regulation (Becker and Henderson, 2000), increased foreign investment in polluting industries in response to domestic air regulation (Hanna, 2010), and green house gas leakage due to regional emission cap-and-trade policies (Fowlie, Reguant, and Ryan, 2016). This paper extends the literature and shows modern polluters' ability to exploit short-run gaps in regulatory monitoring to achieve compliance. Perhaps most closely related to this paper is Reynaert and Sallee (2017), who documented performance gap between vehicles' on-road and laboratory fuel consumption. In a similar spirit, Vollard (2017) finds that illegal discharges of oil waste from shipping increase substantially after sunset when visual inspection becomes difficult. On the side of state air quality monitoring practice, my paper is in line with two recent studies of local governments' air monitoring practice and its implications to our understanding of air pollution data. Grainger, Schreiber, and Chang (2016) use satellite data to study states' strategic pollution monitor placement. Muller and Rudd (2016) document determinants of pollution monitor entries and exits. Finally, this paper adds to the current literature on the causal impact of air pollution exposure on test performances (e.g. Ham, Zweig, and Avol, 2014; Ebenstein, Lavy, and Roth, 2016) and criminal activities (e.g. Reyes, 2007; Herrnstadt, Heyes, Muehlegger, and Saberian, 2016).

The remainder of the paper is organized as follows. Section 2 provides brief background on particulate matter (PM) regulation and monitoring in the U.S. Section 3 presents identification of the main off-days vs. on-days pollution gap. Section 4 explores mechanisms underlying the main findings. Section 5 and 6 study the effect of regulation driven air pollution variation on standardized test performance and crime activities, respectively. Section 7 concludes.

2. Background on Particulate Matter (PM) Regulation and Monitoring

This section provides a brief summary of the U.S. EPA's PM regulation and monitoring practice, focusing on dimensions that are relevant to the empirical part. I begin with the general regulation framework in section 2.1. I then discuss aspects of monitoring in section 2.2. More institutional and administrative details can be found in the Appendix.

2.1. PM Regulation

Regulation of PM pollution in the U.S. is coordinated under the Clean Air Act (CAA). Rather than directly imposing emission restrictions on polluters, the CAA gives the EPA the authority to design and enforce a National Ambient Air Quality Standards (NAAQS) which sets maximum pollution concentration levels for various ambient pollutants such as PM. While the air quality standards are applied nation-wide, realization of regulatory punishments is local. The EPA requires each monitor within an area, usually a county, to meet the associated NAAQS, or the entire area will be designated as in violation of the NAAQS. In cases of violation, or "non-attainment", the parent state has to develop a State Implementation Plan (SIP) that details how plant-specific regulations for all major sources within the state's non-attainment areas will be implemented in order to achieve compliance. Such regulations usually require adoption of pollution abatement technologies and emission limits. The CAA provision also establishes stringent penalties that both the local government and individual sources will face in case of sustained non-compliance. These include (1) highway sanctions that prohibit the approval of almost any highway projects and grants, and (2) emission offset sanctions that require reduced emissions from existing pollution sources for any new or modified emission sources, with the reductions at least doubling the increases. Existing literature has shown that being designated with PM non-attainment leads to significant improvement in local air qualities (Auffhammer, Bento, and Lower, 2009), losses in employment and earnings (Walker, 2013) and reductions in manufacturing plants productivity (Greenstone, List, and Syverson, 2012).[2]

The EPA adopts stringent metrics in air quality non-attainment designations. During the study period from 2001 to 2013, there were two regulation metrics used for PM non-attainment designation, limiting

[2] A rich literature documents the significant effects of other provisions of the CAA 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).

not only the mean but also the maximum level of PM concentration. For example, the NAAQS for fine particulate matter (PM 2.5, or PM with a diameter less than 2.5 micrometers requires the 3-year average of annual mean PM 2.5 values to be lower than 15 ug/m^3 and at the same time requires the 3-year average of annual 98^{th} percentile values to be lower than 35 ug/m^3.[3] According to the EPA, while the mean standard helps reduce long run exposure, the maximum (i.e. 98^{th} percentile) standard protects the public from excessive short term pollution fluctuations. Implementation of the maximum pollution standard implies that once a single day of exceedance is realized, it would be almost impossible to stay compliant for that year with regard to the maximum standard. This is expected to raise the propensity of short term monitor gaming, and in fact, I will show in section 3 that gaming is most pronounced for monitors on the verge of violating the maximum standard.

2.2. PM Monitoring

Unlike monitoring of most gaseous air pollutants which uses automated fluorescence or chemiluminescence methods, PM monitoring is filter-based which involves manual sample collection, transportation, storage, and subsequent laboratory weighing analysis.[4] In practice, a cyclical sampling framework is employed to sample ambient PM at fixed time intervals. From 2001 to 2013, about 98% of PM monitors follow either 1/6day (42% of monitors), 1/3day (33%), and 1/1day (22%) schedules where more frequent sampling rate is applied to areas with higher baseline pollution.[5] Although this paper focuses on the 1/6day schedule where gaming is the most expected, I also report pollution responses to the 1/3day schedule. In addition, I take advantage of the existence of the 1/1day sites to conduct "placebo" checks which confirms that no gaming is observed at these sites where PM samples are taken every day.

The EPA's sampling schedule is publicly available. At the end of each calendar year, the federal EPA publishes a sampling schedule calendar on its official website, highlighting the dates when monitors with different sampling rates should be turned on for the next calendar year. Figure 1 presents the sampling calendar for 2001.

[3] NAAQS standards change over the years due to EPA's periodic review of new evidence on the health impacts of air pollution. For example, before 2006 the PM 2.5 maximum cap was 65 ug/m^3. I provide further detail about how changes in NAAQS caps are dealt with in this research.
[4] See the Appendix for description for recent development in continuous PM monitoring technology.
[5] The EPA has also used a once every 12 day monitoring schedule, but it is rarely used in the context of PM regulation. Assignment of frequency depends on factors such as historical PM concentration at the monitor. See the Appendix for more discussion.

Whereas the federal EPA designs air quality standards and determines states' compliance status, the authority of pollution monitoring is often granted to the states. The fact that states are charged with monitoring air quality and at the same time bear the regulation punishments if their monitoring data show non-compliance raises the concern that the quality of monitoring data may be compromised by the perverse incentive structure. The EPA implemented a number of measures to prevent states and polluters from gaming the monitoring system. For example, to prevent states from selective sampling, a monitor is considered eligible for NAAQS comparison only if it has sampled more than 75% of required sample in each quarter of the year. States can supply a makeup sample in cases where a scheduled sample is missed, but the makeup sample must be collected within seven days since the originally scheduled date in order to be considered valid. To further prevent states from over-sampling low pollution periods, the EPA only accepts an applicable number of samples with the highest pollution readings in cases where more samples than required are taken. These anti-gaming measures perhaps also reveals the degree to which the EPA recognize states' ability to game the monitoring system.

Part I. The Effect of Incomplete Monitoring on Air Quality

3. Does Air Quality Deteriorate When Pollution Monitors Are Scheduled Off?

3.1. Data and Summary Statistics

Monitor data. First, I obtain PM monitor characteristics from the EPA's Air Quality System (AQS) from 2001 to 2013. I use AQS' annual summary data files which contain monitor level information on scheduled number of monitoring days, actual monitoring days, latitude and longitude location, as well as annual PM concentration statistics such as the mean and the max. I identify 1/6day (1/3day, 1/1day) monitors by finding monitors that are required to sample 60 or 61 (121 or 122, 365 or 366) days a year. I cross check the validity of this definition by comparing assigned schedule to actual daily monitor operation status from AQS' daily monitor summary files, finding that these monitors follow the 1/6day schedule more than 94% of the time. 1/6day monitors may occasionally deviate from the monitoring schedule (e.g. power outages and sampler malfunctions), which explains the remaining 6% deviations. Finally, monitoring schedules are obtained from the EPA's annual publications of sampling schedule calendar that first became available in 2001.

Satellite data. My main pollution measure is a satellite measure of atmospheric aerosol concentration. The data is obtained from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol product, derived from the MODIS algorithm installed on NASA's satellite Terra which retrieves atmospheric aerosol level using a flexible set of spectral radiance instruments that are able to distinguish atmospheric aerosols from other climatological parameters such as water vapor (Remer et al., 2005; Donaldson and Storeygard, 2016). The resulting data are a series of daily aerosol maps for the contiguous U.S. with a spatial resolution of 10km×10km. From these maps I build a daily panel dataset linking each 10km×10km grid cell from 2001 to 2013.[6]

I choose MODIS aerosol to be the main air quality measure in this research for two main reasons. First, it provides a consistent measure of air quality in the absence of ground monitoring data. Previous research has used the same data as the proxy for air pollution in developing country contexts where monitoring data are not available (Foster, Gutierrez, and Kumar, 2009; Chen, Jin, Kumar, and Shi, 2013). Second, existing atmospheric science literature has shown that the aerosol measure is a strong predictor of ground level PM concentrations in various contexts (Liu, Franklin, Kahn, and Koutrakis, 2007; Lee, Coull, Bell, and Koutrakis, 2012; Zhang and Lee, 2015). This relationship is much expected, as atmospheric aerosol is composed mainly of sulfates, black carbon, mineral dust, and sea salts, all of which contribute to PM (Voiland, 2010). In the Appendix, I also re-discover this relationship, where I correlate monitor-daily level PM concentrations to the observed daily aerosol level within the 10km×10km grid where the monitor falls in. This is done for every monitors located in the lower 48 states. In general, I found the relationship to be positive, precise, and fairly linear across the distribution.

Summary statistics. Table 1 presents satellite air pollution and monitor summary statistics by calendar year. Column 1 to 4 shows 10km×10km grid-daily level satellite aerosol pollution statistics. The MODIS aerosol measure has a theoretical range from -5 to 500, and the mean level during my sample period is 12.4 (SD = 14.2), with over 95% percent of observations fall within the range between 0 and 100.[7] Pollution level stayed relatively stable, having declined an average of 0.5% per year from 2001 to 2013.

[6] The measure is called aerosol optical depth (AOD), a measure of the degree to which solar beam transmission is absorbed or scattered by atmospheric aerosols. More details about construction of the satellite pollution variable are included in the Appendix.

[7] The raw aerosol optical depth (AOD) measure ranges from -0.05 to 5. For presentation purpose I scale the raw measure by 100.

The decline rate is about 1.42% per year at PM 2.5 monitors and about 1.05% at PM 10 monitors. This is roughly on par with trends of monitor data, which shows an annual decline rate of 2.3% for PM 2.5 and 1.35% for PM 10.

Column 5 to 7 of Table 1 show monitor level statistics, including describe annual total number of monitors, number of 1/6day monitors, and number of 1/6day monitors exceeding PM NAAQS for that year. In both summary statistics and the following analyses, I restrict to NAAQS-eligible monitors, i.e. monitors that obtain at least 75% of required samples for each quarter of the year and therefore are eligible to be used by states to show NAAQS designation. Column 5 and 6 show that total number of monitors decrease over time. This trend began in 1997 with a NAAQS revision which initiated PM 2.5 monitoring and redirected sources from monitoring of coarser particulates pollution.[8] Column 7 shows that, every year, about 7% of 1/6day monitors exceed NAAQS. 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. As I discuss further below, both of these features are exploited in the empirical analysis. First, I exploit monitor retirement events to show that the off-days vs. on-days pollution gap disappears as monitors retire. Second, I explore variations in baseline pollution level to show that gaming is targeting at monitors with high potential of violating NAAQS.

Columns 8 and 9 present monitor level statistics aggregated to the monitoring site level. A monitoring site is a geographic unit in which multiple monitors may live. Since within a monitoring site locations of different monitors are not distinguished between each other, the main analysis is done at the monitoring site level. To be conservative in aggregating up individual monitor level schedules to the site level, I define a site to be a 1/6day site if any PM monitor in that site follows the 1/6day schedule. Defining the sample this way is expected to bias me against finding strategic responses to the 6-day monitoring schedule. For example, some monitoring site may have a 1/1day monitor *and* a 1/6day monitor, where the latter is used to provide quality assurance data for the former.[9] In analysis below I report confirmation that evidence of schedule gaming is stronger if I restrict to sites with a standalone 1/6day monitor.

Column 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 roughly 200 million population in the lower 48 states live in these counties, and column 13 shows that about 65% of them live in counties

---

[8] For the 1997 NAAQS revision see U.S. EPA (1997b).

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

that have at least one 1/6day monitor. Note that, although total population coverage of PM-monitoring counties has stayed almost constant over time, population that live in 1/6day counties has declined significantly during the study period, driven by the substantial decrease in the number of 1/6day monitors shown in column 6.

Table 2 reports that monitoring compliance is high. An average 1/6day monitors took 58.4 samples (SD = 2.2) in a year, while 60 or 61 samples are required. More than 96% of these monitors took at least 90% of required samples. In contrast, few monitors took full samples. Compliance is similar among 1/3day and 1/1day monitors.

Finally, the PM monitor network is geographically dispersed. For example, 1/6day monitors are observed in every one of the lower-48 states during the study period. A map of monitors and their schedules follows can be found in the Appendix.

3.2. Empirical Specification

The strict 1/6day design of the monitoring schedule motivates a straightforward identification strategy which estimates the causal effect of the schedule on pollution by simply comparing levels of air pollution across days of a 1/6day 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.[10] 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 omitted from 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 extensive length of the panel and the strict 1/6day

[10] 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% of grids contain a single site, and less than 1% of grids contain more than three sites. I confirm that dropping duplicative grid-day observations has negligible impacts on the results.

cyclicality of the off-days treatment implies that very few confounders may bias β's from identifying the causal effect of the monitoring schedule. I therefore report two types of specifications below. 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. 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

$$Aerosol_{st} = \beta \cdot Offdays_t + Time_t + \alpha_s + X_{st}\gamma + \varepsilon_{st} \qquad (2)$$

All components in this estimation equation are the same with equation (2), with the only difference that, rather than having five dummies separately indicating days of a 1/6day monitoring cycle, equation (2) includes the $Offdays_t$ 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 the β's 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 schedule. In other words, I assume that the only reason that ambient air quality would show a significant pattern once every six days is because polluters react to the incentive of monitoring avoidance generated by the 1/6day sampling schedule. I report two types of "placebo" tests to support the validity of this assumption. First, I exploit about 490 cases of monitoring site retirement from 2001-2013, contrasting 1/6day pollution gaps before and after site retirement. The underlying idea is that, if gaming is targeting at the sampling schedule, then no pollution gap should be observed in the same place where the 1/6day site was after it retired. Second, I repeat the 1/6day pollution examination to areas with "placebo" sites where gaming is not expected. These include sites that follow the 1/1day schedule, as well as sites that do follow the 1/6day schedule, but monitoring pollutants not linked to any regulation standards.

3.3. Baseline Results: 1/6day Pollution Gap

I first estimate equation (1) using my preferred sample which includes all monitoring sites containing at least one 1/6day 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 1/6day monitoring cycle, averaged across all cycles in the sample. As described in the introduction, within a typical monitoring cycle, air pollution exhibits a flat path, except for a sharp drop during the on-day. This pattern is a striking revelation of polluters' ability to manipulate ambient air quality at the monitoring sites on a daily basis. In the appendix, I report the same graph conditioning on the full set of covariates and the graphical pattern looks visually identical.

Table 3 reports the average 1/6day off-days vs. on-days pollution gap using equation (2). Results in column 1 corresponds to Figure 2 and shows that air pollution is on average 1.6% 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 1/6day PM monitor. In this case, the pollution gap rises to about 1.8%. The effect persists if I further restricts the sample to counties with only 1/6day monitors (column 4). It's good that gaming on average appear stronger in standalone sites, because it provides confirmation that gaming is against 1/6day monitors. [11]

3.4. "Placebo" Tests

To boost the confidence in the internal validity of the empirical design, I provide two types of "placebo" tests that establish null off-days effects in places where gaming is not expected.

The first test explores the retirement of 1/6day monitoring sites. If gaming is targeting at the 1/6day schedule, then one should expect the disappearance of the pollution gap after sites are removed. To operationalize this test, I first draw upon information in the EPA's monitor listing file and identify 490 cases of 1/6day monitoring sites retirement events. The analysis then uses the satellite measure to track air quality in the areas where these sites located, and compare 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

[11] Coefficient estimates obtained using the preferred sample and the restrictive sample is not statistically distinguishable. A joint test of equal effects across sites with multiple monitors and a standalone 1/6day monitor yields a *p*-value of 0.448.

removed because the monitoring calendar is universally applied, and hence I know what the sampling dates would have been even in the absence of a monitoring site. Figure 3 reports the results, where the 1/6day pollution gap is shown as a function of years relative to sites' retirement. The gap is about 2.1% while the site was still operating; for literally the same area, the gap closes immediately after monitor retires. This pattern provides support to the identification assumption that gaming is targeting at the monitoring schedule.

In the second type of "placebo" checks, I apply the same logic and estimate the 1/6day pollution gap near sites where gaming is either not feasible or not necessary. These include about 560 monitoring sites that follow the 1/1day schedule, and about 800 hazardous air pollutants (HAPs) monitoring sites that also follow the 1/6day schedule although the pollutants were not subject to any regulatory standards.[12] Table 4 reports that no significant 1/6day pollution pattern is detected near these sites. Table 4 also reports simple power calculation which shows that the "placebo" tests have enough statistical power to detect an effect that is similar in size to the effect found in the main analysis (Table 3) at a conventional significance level. These findings again support the identification assumption that no 1/6day pollution gap would have been observed in the absence of the 1/6day monitoring schedule.

3.5. Pollution Gap at the 1/3day Monitoring Sites

I now repeat the same analysis for sites that follow the 1/3day monitoring schedule. Despite the high sampling frequency, it is an empirical question whether polluters can engage in effective gaming against sampling schedule on a 3-day basis.

My analysis finds no evidence of gaming against the 1/3day sites. Figure 4 shows that, on average, pollution path within a typical 3-day monitoring cycle exhibits a "V" shape, but the off-days vs. on-days difference in pollution is not significant. Regression estimates in Table 5 confirm this finding. Column 1 and 2 show that the average 1/3day pollution gap is about 0.3% and is not statistically significant whether or not control variables are included. Column 3 reports that restricting to sites with standalone 1/3day monitors does not increase the estimate. When estimation sample is further restricted to counties with only 1/3day monitor (column 4), I find a pollution gap of 0.54, and it is marginally significant (p = 0.076). Again, simple calculation suggests sufficient statistical power to detect an effect size similar to the 1/6day

[12] 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.

pollution gap (power > 0.999 for detection of a 1.5% effect at 5% significance level). This simple analysis perhaps points to a straightforward policy implication that gaming can be avoided by increasing adoption of the 1/3day sampling schedule.

4. Sources of Monitor Gaming

I now turn to the exploration of the sources of monitor gaming. This section contains three parts. First, in section 4.1, I present an analysis where gaming, i.e. the 1/6day pollution gap, is compared between areas with high vs. low potential of violating the regulatory air quality standards. Next, in section 4.2, I take a more direct approach to trace out the source of gaming, using the satellite measure to detect abnormal pollution patterns near industrial point sources. Finally, section 4.3 probes the possibility of local government coaching by relating gaming to characteristics of the political environment.

4.1. Heterogeneity by the Potential to Violate the NAAQS

I begin by examining the interaction between the pollution gap observed at the monitoring sites and the sites' potential to violate the EPA's NAAQS standards. Because I hypothesize that the avoidance of NAAQS violation, or "non-attainment", is the underlying motivation for gaming in the first place, a larger pollution gap is expected for sites that have violated, or near the violation of, the EPA's standards. This analysis is also motivated by the findings of Auffhammer, Bento, and Lowe (2009) who found a more profound reduction in PM 10 pollution for non-attainment sites relative to attainment sites during a similar time frame, which suggests differential targeting of resources in NAAQS compliance.

During my study period, three different NAAQS standards of PM pollution are used by the EPA for non-attainment designation: (1) the PM 2.5 24-hour standard, requiring the 3-year average of annual 98th percentile values to be less than 35 ug/m^3; (2) the PM 2.5 annual standard, requiring the 3-year average of annual mean values to be less than 15 ug/m^3; (3) the PM 10 24-hour standard, requiring the 3-year average of annual 99th percentile values to be less than 150 ug/m^3.[13] This last standard on PM 10 was

[13] The PM 2.5 24-hour standard was revised from 65 ug/m^3 to 35 ug/m^3 in October 2006. According to the EPA's non-attainment designation history, the first annual NAAQS designation under this new standard is done in December 2009 based on the 3-year average pollution concentration using monitoring data from 2006, 2007, and

rarely violated during my study period. Using these regulatory metrics, for each year and each regulatory standard, I assign non-attainment status to individual monitoring sites. I then re-estimate equation (2), now allowing the 1/6day pollution gap coefficient to vary by sites' non-attainment status.

Table 6, panel A reports the 1/6day pollution gap coefficients estimated separately for attainment (i.e. NAAQS compliant) vs. non-attainment (i.e. NAAQS violating) sites for each NAAQS standard. Column 1 suggests that the pollution gap at sites violating the PM 2.5 24-hour standard is about 3.4%, more than twice the size than the gap observed at attainment sites. The difference, however, is not statistically significant due to the large standard error on the non-attainment sites coefficient (p = 0.273). Column 2 and 3 show little evidence of heterogeneity with regard to the PM 2.5 annual standard and the PM 10 24-hour standard.

The attainment vs. non-attainment comparison likely suffers from low power due to the scarcity of non-attainment sites. In addition, by focusing on sites are already in non-attainment, the comparison fails to capture preventive gaming which might take place for sites that are on the verge of non-attainment. I report a specification which interacts the off-days dummy with a second metric of NAAQS violation that better captures the dynamic incentive of monitor gaming and at the same time increases statistical power. The idea is to leverage on the fact that non-attainment designation decisions are based on 3-year average pollution statistics. Therefore, the incentive to engage in monitor gaming is the strongest if a site has already violated an air standard twice in the past two years, in which case there is pressing need to bring pollution down in the current year in order to avoid falling into non-attainment. Specifically, for each PM standard, I allow the 1/6day pollution gap to vary by the number of years in the previous two years that the site has exceeded the standard. Table 6, panel B summarizes the results. Once again, column 1 reveals that a larger gap (about 5.3% off-days vs. on-days difference) is observed near sites that are on the verge of violating the PM 2.5 24-hour standard, while no statistically significant heterogeneity is detected for the other two standards.

While Table 6 provides only suggestive evidence of a larger gap among non-attainment sites, the analysis shines some light on the potential incentives at play. For example, one possible explanation of why larger gaming appears to consistently emerge under the PM 2.5 24-hour standard is that the standard

---

2008. Consequently, there were no PM 2.5 24-hour designations for year 2007 and 2008. In my analysis regarding the PM 2.5 24-hour standard, these years are included and flagged as a separate group. Regression coefficients for this group is not reported in the main text, and is available upon request. Before October 2006, there was also a PM 10 annual standard requiring 3-year average of annual mean value to be less than 50 ug/m^3. This standard was rarely violated since the 2000s and was revoked in 2006.

is targeting at almost the maximum statistics, leaving no "second chance" of compliance once a particularly high pollution day is realized. Gaming is hence more rewarding in this context relative to the PM 2.5 annual mean standard, where the impact of a high pollution day can be averaged out over time. I now move on to higher power settings where the interaction between gaming and non-attainment can be better estimated with available data.

4.2. Gaming near Industrial Sources

I now begin to examine the source of the 1/6day pollution gap observed at the monitoring sites. I take the most direct approach, looking for abnormal pollution patterns near industrial sources that echo the 1/6day monitoring schedule. Specifically, I use the satellite pollution measure to test whether air pollution is higher on off-days relative to on-days near major point sources of pollution, as would be the case if polluters respond to the monitoring schedule by exhibiting differential polluting behavior across on- and off-days. To further attribute this behavior to gaming, I estimate a flexible specification which allows the off-days effect to depend on the facility's distance to the nearest non-attainment 1/6day monitoring site. Because polluters are not expected to have control over ambient air pollution over long distances, gaming is expected to decrease over the distance between the polluter and the closest monitoring site. The specific estimation equation is as follows

$$Aerosol_{pt} = \left(Offdays_t \times 1\left[Dist_{pt} = d\right]\right)' \cdot \beta_d + Time_t + \alpha_p + X_{pt}\gamma + \varepsilon_t \qquad (3)$$

The outcome variable $Aerosol_{pt}$ is satellite pollution measure near plant *p* on date *t*. In implementation, the indicator function $1\left[Dist_{pt} = d\right]$ is a categorical variable that indicates decile distance from plant *p* to the nearest non-attainment site. Interacting this categorical variable with the off-days dummy $Offdays_t$ allows the key coefficient β to vary by distance bin d. The rest of the notation is analogous to that in equation (2), except that, in the interest of space, X_{pt} is understood to include the main distance bin variable $1\left[Dist_{pt} = d\right]$.

My analysis focuses on polluters observed in the EPA's Toxic Release Inventory (TRI), which contains annual location and self-reported emission information for all facilities that are required to report under the 1986 Emergency Planning and Community Right-to-Know Act (EPCRA). These facilities span a wide range of industries including mining, utility, manufacturing, wholesale distribution, and hazardous waste treatment. By the EPCRA, facilities in these industries are required to be included in the TRI if they have at least 10 full-time employees and process or use any one of EPCRA-listed toxic pollutants by more than the threshold amounts. As a limitation, reporting to TRI is based on toxic pollutants emission, and thus the TRI does not include PM emitters that do not produce or use any one of the EPCRA-listed pollutants. As a robustness check, I repeat the analysis in this section using an alternative sample of polluters from the EPA's National Emissions Inventory (NEI) which provides a near census of polluters, although NEI only provides snapshots of polluter profiles in selected years. The robustness checks along with more discussions about data sources are included in the Appendix.

Figure 5 plots the β_d estimates from equation (3). Results suggest a clear distance gradient: for facilities that are the closest to non-attainment monitoring sites (average = 0.5 miles), satellite detects that air quality on an off-day is on average 2.2% worse than an on-day; the gap shrinks as the distance increases, and no gap is observed for polluters that are more than 15 miles away from the closest site. The visual evidence provides support that industrial sources contribute to the 1/6day pollution gap, and that avoidance of NAAQS non-attainment appears to be driving the gaming behavior. In the Appendix, I report further examinations that explore heterogeneity in the distance gradient along two dimensions: (1) I show an absence of a clear distance gradient for out-of-state polluter-monitor pairs, which is consistent with the fact that consequences of NAAQS violation are only felt within the state border, and (2) I report substantial heterogeneity in gaming across industries and I explore determinants of such heterogeneity.

The discovery of gaming near industrial sources prompts the question about the underlying coordination among polluters in gaming the monitoring schedule. Is the observed 1/6day pollution gap contributed solely in areas with major polluters? If not, how does the incentive and coordination play out in areas with multiple polluters? I provide answers to these questions by first relating gaming to a Herfindahl-style measure (HHI) of emission concentration at the local level. Specifically, for each county *c* and year *y*, I define its emission HHI to be

$$\mathrm{HHI}_{cy} = \begin{cases} \dfrac{1}{1-1/N_{cy}}\left(\sum_{i=1}^{N_{cy}} \left(\dfrac{Emission_{icy}}{Emission_{cy}}\right)^2 - 1/N_{cy}\right) & \text{if } N_{cy} > 1 \\ 1 & \text{if } N_{cy} = 1 \end{cases} \quad (4)$$

where $Emission_{icy}/Emission_{cy}$ is the share of air pollutants emission by polluter *i* in county *c* and year *y*, and N_{cy} is the total number of polluting facilities in the county-year. The value of the HHI therefore ranges from 0 to 1, with higher value representing the highest emission concentration, i.e. areas where emissions are concentrated in the hands of few polluters. The heterogeneity analysis therefore compares the magnitude of pollution gap in areas with high HHI vs. areas with low HHI. To address the concern that emission information from the TRI is self-reported which has been shown to bear significant biases (de Marchi and Hamilton, 2006; Koehler and Spengler, 2007), I confirm that results are similar if I simply compare pollution gap in areas with a single polluter with areas with multiple polluters. These checks are reported in the Appendix.

Figure 6 reports heterogeneous 1/6day pollution pattern by high (≥ 0.9) vs. low (< 0.9) HHI. Results show significantly stronger gaming in areas with high levels of emission concentration. In high HHI areas, the off-days vs. on-days pollution gap averages 3.1%, and the pollution pattern within a typical 6-day monitoring cycle exhibits a "V" shape, as oppose to the "T" shape observed at an average monitoring site. The difference may occur for various reasons. For example, it may take multiple days for ambient pollution to clear up as major polluters ramp down emission. Alternatively, this "V" pattern can arise when major polluters engage in coordination where relatively smaller polluters emit around on-days and the off-days are reserved for the big polluters.

One important observation from Figure 6 is that gaming is not driven entirely by high emission concentration areas: the magnitude of the 1/6day pollution gap in low concentration areas is roughly 1.2%, and is individually statistically significant. This result suggests that the pollution gap does not arise solely from polluters' self-coordinated monitoring avoidance: in the absence of coaching, it is implausible that multiple polluters may successfully coordinate themselves in gaming ambient air quality. In the following subsection, I probe the possibility of the existence of extra coordination and coaching mechanisms which support gaming. I focus on the role of local government, who would share the penalty of NAAQS in cases of non-attainment.

In the appendix, I report additional analysis which extends the analysis to "mobile" sources of pollution. I examine whether the timing of prescribed burning, a tool used in forest management and farming that intentionally starts fires to reduce excessive fuel accumulation, is strategically adjusted to avoid the EPA's monitoring schedule. I find no evidence that it is so.

4.3. Gaming and the Local Political Environment

This subsection presents evidence on the observable political environment correlates of monitor schedule gaming. The goal of the analysis is to shed light on the possibility that local governments, who share the penalties of NAAQS enforcement in cases of air quality non-attainment, play a coaching role in gaming the EPA's 1/6day sampling schedule.

I begin with an illustrative example in a context where a specific format of government coaching is directly observed. Throughout the country, many local air pollution control agencies adopt air pollution "Action Day" programs which issue public warnings when ambient pollution concentration is expected to reach unhealthy level. When an action day is issued, citizens are advised to "take actions" to prevent deteriorations of air quality by reducing energy and automobile usage. From 2004 to 2013, 346 reporting areas have voluntarily adopted these programs. These areas are a mix of cities, counties, metro areas, and states covering a total of 51% of the U.S. population in the lower 48 states. I obtain all Action Day records from the EPA's Airnow program. To avoid double counting issuances in cases of overlapping and nested jurisdictions, I aggregate the data to the core-based metro area level.[14] This gives me a total of 14,945 issuances from 2004 to 2013. A map showing the geographic distribution of the Action Day issuances can be found in the Appendix.

I examine whether Action Day advisories, which call for public energy conservation to reduce air pollution, are more likely to be issued on days when PM monitoring is scheduled. Figure 7, panel A plots the raw histogram of Action Day issuances by days relative to the on-day. Whereas Action Days are expected to distribute evenly across the 6-day cycle, I find that 17.4% of Action Days occurred on on-days, relative to the off-days average of 16.5%. Although the raw difference is small (about 5.5% more on-day warnings in percentage term), it is masked by the fact that Action Day advisories usually last for days, and so an on-day advisory is associated with increased chance of the following days being action days, which dilutes the off-days vs. on-days difference. Figure 7, panel B therefore adjusts for consecutive action days by plotting the histogram of the onset of Action Days "episodes". Using this adjustment, for a period of consecutive Action Days, only the first day is included in computing the histogram. Post adjustment, I find

[14] 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).

18.2% of Action Days are issued on on-days, relative to an off-days mean of 16.4%. In relative terms, this implies that 11.1% more Action Days are issued on on-days than off-days.

This example illustrates one mechanism through which local governments appears to coach citizens in gaming the EPA's monitoring schedule. I now turn to a more general evaluation, relating gaming to characteristics of local governments. I focus on two groups of measures. The first group of measures capture the general political environment, including

(1) Government size. Size is defined as the annual government employment as a share of total employment. I build government size measures both at the state and the county level. The data are obtained from the Bureau of Economic Analysis;

(2) Corruption. Corruption is defined as the 13-year (2001-2013) average of per capita number of federal convictions among state and local public officials. This measure is available at the state level. The measure is sourced from the Report to Congress on the Activities and Operations of the Public Integrity Section (PIN), which has been previously used in economic research of corruption in the U.S. (Glaeser and Saks, 2006; Leeson and Sobel, 2008; Grooms, 2015);

The second group of measures capture local government's environmental attentiveness, and in particular, the resources devoted to NAAQS compliance. I define

(3) "Pro-environment" score. I use the 13-year (2001-2013) average of the League of Conservation Voter's (LCV) score which is based on state representatives' voting records on environmental issues, with a higher score corresponding to a stronger environmental preference (Dietz et al., 2015; Grooms, 2015);

(4) State's history of challenging the EPA's PM non-attainment designations. Under the Clean Air Act, when the EPA revises the NAAQS, each state has an opportunity to recommend designations. This usually takes the form of a list of the state's counties that the state believes should be designated non-attainment. The recommendation is reviewed by the EPA, who then notifies the state whether or what parts of its recommendation is accepted, along with a preliminary designation. The state will then have two months to challenge the designation, usually by presenting new data analyses and arguments that support the original recommendation, before the final designation decision is made by the EPA. These communications are published on the EPA's website. Using this information, I create two measures that reveal states' NAAQS-attentiveness and the political resources available for NAAQS compliance. In the first measure, I observe whether the state has ever challenged the EPA's designations. In my study sample of the lower 48 states, thirty one of them have done so. Not surprisingly, many states that did not

challenge the EPA are those with no areas violating the NAAQS. To better tease out the role of state's attentiveness and resources, I create a second measure of challenge in which I count the total number of pages each states have put together in challenging the EPA's designation. This measure has a wide spread among the thirty one states that ever filed a challenge (mean = 62.4, SD = 69.8);

(5) Presence of air quality Action Day programs, as described earlier in this subsection.

Figure 8 summarizes the results. On the left panel, each row reports the coefficient on the interaction term between the off-days dummy and an indicator for above-median government characteristics. These coefficients are obtained from separate regressions, interacting the off-days dummy with one characteristics at a time. I find that gaming concentrates in states with below-median government size, while county government size does not correlate with gaming. Corruption, as proxied by below-median per capita federal convictions of public officials, is *negatively* correlated with gaming, although the correlation is not statistically significant. More specifically, I find gaming is more pronounced under administrations with stronger environmental focus. These are above median “pro-environment” states, states with a history of challenging the EPA’s preliminary PM NAAQS designation, states that spent above median efforts in challenging, as well as states with Action Day advisory infrastructures. As state characteristics likely correlate with one another, on the right panel I report joint estimation results where all interactions between gaming and state characteristics are included simultaneously in the model. This exercise pinponits three strong predictors of gaming: below median state government size, above median state “pro-environment” score, and a history of challenging the EPA’s designation.

Although my estimates do not represent the causal influence of local government on gaming, the pattern is consistent with the view that effective strategic responses to the federal monitoring schedule require coordination and management at the high level, which is more likely to occur under the administration of governments with more capacities and resources for environmental compliance.

Part II. Cognitive and Behavioral Responses to Schedule-Driven Air Pollution

In the first part of the paper, I use a satellite-based measure to show that ambient air quality significantly deteriorates when the EPA's regulatory particulate matter pollution monitors are scheduled off. The analysis points to several contributing incentives underlying the effect, such as the avoidance of regulatory punishments associated with air quality non-attainment designation. In this second part of the paper, I examine whether the air quality variation driven by the monitoring schedule has adverse consequences for the nearby population.

I focus on standardized test scores and criminal activities, two cognitive and behavioral outcomes suggested by the recent literature to be causally influenced by short term air pollution fluctuations (Ebenstein, Lavy, and Roth, 2016; Herrnstadt, Heyes, Muehlegger, and Saberian, 2016). The analysis is a simple extension of the primary examination on pollution. Whereas part I of the paper establishes that the 1/6day monitoring schedule generates a gap in air pollution between monitor off-days and on-days, here I examine whether the gap coincides with poorer test performance and higher crime rates during the off-days.

5. Monitor Off-days vs. On-days Test Scores Gap

5.1. 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 comprises 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. At most one session is administered in a given month. 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 missed the census may take makeup sessions held in May. In cases of failure, students are allow to retake the exams in future sessions (including the census 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 interception with the EPA’s 1/6day 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, the data has a cell size limit of 10 student, so that test scores are masked if they are averaged over 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 58 counties in California, aggregated over 91 different tests dates and more than 14 million individual tests taken from 2004 to 2013. Table 7 reports test performance statistics. The average math (English) scale score is 367.2 (370.3) within the 275-450 scale. As expected, February and March sessions have much higher average score due to the fact that tests on other months are mostly taken by students who did not pass CAHSEE on their first attempt in the February and March census. In the analysis below, I control for this compositional difference by including month-of-year fixed effects.

5.2. Estimation Framework 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 per 6 tests, step on the EPA’s monitor 1/6day off-days. However, the scarcity of test dates implies that the interaction between test dates and monitoring dates alone does not ensure strong balancedness 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 which compares test scores on off-days vs. on-days, including only month-of-year dummies to control for different student composition for test administered in different month of the year. I then report a much richer specification to ensure balancedness 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.

Since schools are within a relatively narrow geographic extent (i.e. the state of California), I begin the analysis with all schools in the data regardless of their location within the state. In other words, 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 off-days vs. on-days pollution gap and examine whether test scores responses are stronger for schools that locate 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 level.

Apart from test scores, I also examine whether number of test taken reduces on monitor off-days, as would be the case if pollution causes sickness among students. The econometric specifications are analogous to those used in test scores examination, except that (1) the outcome variable is now log number of tests taken in a school-test cell, and (2) regressions are no longer weighted by the number of tests taken.

Table 8 reports the main results. I first focus on the upper panel, 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 that control for differences in test takers composition. This specification shows that taking an exam on an off-day reduces test score significantly by 5.3% of a standard deviation. In column 2, I include the full set of controls as described earlier to ensure balancedness 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% of a standard deviation and it is not statistically significant. In column 3 to 5, I repeat the estimation in column 2, but separately for schools that are close (< 10 miles) and far away (10-50 miles and > 50 miles) from the nearest non-attainment PM monitor. Results suggest that the effect of off-days is driven by schools close to monitors. Column 3 shows that for schools that are within 10 miles from the nearest non-attainment monitor, taking a test on an off-day reduces scores by a statistically significantly 6.3% of a standard deviation. In contrast, column 4 and 5 show that no significant responses are detected for schools that are beyond 10 miles, although the standard errors do not allow me to rule out small effects. The impact of monitor off-days on test scores are moderate in size. Take the estimate in column 3 for example. On an average off-day, test score is about 6.3% of a standard deviation lower, which is roughly 1.54 points in scale score. The average white-black CAHSEE test score gap from 2004-2013 is 32.15 points. Thus, the effect size of a monitor off-day exam is about 4.8% of the white-black test score gap.

In the lower panel of Table 8, I repeat the same estimation but now testing 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 column 1 and 2 suggest that 4.6 to 6.8% fewer tests are taken on off-days, but the effects are not precisely estimated. Column 3 through 5 again present estimation separately for schools close to vs. far away from non-attainment monitors. The analysis provides suggestive evidence of a distance gradient. Column 3 shows that about 11.5% fewer tests are taken on an off-day for schools less than 10 miles away from the nearest non-attainment monitor, although the effect is marginally significant. In column 4 and 5, schools that are farther away show smaller and insignificant responses. Due to the lack of statistical precision, I cannot make strong conclusion that fewer students attend tests administered on monitor off-days.

6. Monitor Off-days vs. On-days Crime Gap

6.1. Background and Data

I use crime data from the FBI's National Incident-Based Crime Reporting System (NIBRS) from 2001 to 2013. This data contains 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 the city (or county) law enforcement agencies. The number of participating jurisdictions in the NIBRS has grown over time. By 2013, NIBRS covers about 92 million population in 33 states, and accounts for more than 28% of all crime reported to the FBI Uniform Crime Reporting Program.

Despite the growing coverage, crime prevalence in NIBRS-participating jurisdictions is understood to be not representative of overall crime rates. For example, jurisdictions with fewer population are known to be disproportionately more likely to report data (James and Council, 2008). Differences in levels of crime is not a threat to the identification strategy which uses day-to-day variation in pollution monitoring status within a same area. To alleviate concerns about compositional changes in the NIBRS-covered population over time, my analysis also restricts to jurisdictions that have participated in NIBRS for at least 10 years during the study period.[15] However, I recognize that sensitivity to air pollution may differ among population living in reporting and non-reporting areas. This is an important caveat regarding external validity throughout the interpretation of results obtained in this section.

I begin by first constructing daily crime rates at the county level using information on jurisdictions' county location and the population covered.[16] In other words, crime rate is defined as the reported number of crime incidents in the county divided by the population covered by NIBRS within the county. This data is then merged with counties' PM monitoring frequency. From 2001 to 2013, 356 counties follow the 1/6day sampling schedule (defined as counties where all PM monitors follow the 1/6day schedule) and crime data are available in 47 counties. These counties span 19 different states, covers a total of 1.33 million population, and represents 67% 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 rates variables for violent crime (aggravated assault, robbery, forcible rage, murder, and nonnegligent manslaughter), property crime (burglary, larceny-theft, motor vehicle theft, and arson), and other crime (non-violent and non-property crime). Table 9 presents summary statistics for counties in the estimation sample, i.e. counties that follow the 1/6day schedule and with crime data available, as well as the NIBRS population.

6.2. Estimation Framework 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. Again, I report results from two types of econometric 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.

---

[15] In unreported results, I confirm that the findings in this section are similar if I use a less balanced panel or if I use a strictly balanced panel, i.e. counties that consistently report to NIBRS for the entire 2001-2013 period.

[16] Every year, about 10 percent of crime incidents occur in jurisdictions that cross county borders. In these cases where a jurisdiction spans multiple counties, 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 crime incident to each of the counties that the jurisdiction spans, with the assigning probability proportional to the county's population share within the jurisdiction.

Table 10, column 1 begins the analysis by repeating the pollution analysis. The estimation conceptually corresponds to column 4 of Table 3, whereas now pollution is measured at the county level, i.e. average of all 10km × 10km pixels within the county border. For example, column 1, panel A shows that the off-days vs. on-days pollution gap is estimated to be 1.3% which is smaller than the effect size observed at the monitoring sites (a 1.8% pollution gap), likely due to the fact that gaming is targeting at the monitoring sites and therefore the average effect observed at the county level is lower. In column 2, the pollution regression is repeated again using the estimation sample, which yields a pollution gap of 2.7% without controls and 1.9% with controls. The 95% confidence intervals of these estimates overlap with the mean coefficients in column 1.

Column 3 to 5 reveal 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% increase. Whereas previous literature documents little evidence on the effect of pollution on property crime, in column 4 I report that property crime, including burglary, larceny-theft, motor vehicle theft, and arson, 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% out of the daily mean of 110.31 per million. Column 5 continues the analysis with crime in other categories where 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

(IN PROGRESS) In this paper, I use a satellite-based measure to show that gaps in the U.S. Environmental Protection Agency's air particulate matter pollution sampling schedule lead to worse air quality during days when regulatory sampling is not scheduled. My findings echo recent experience with the automobile industry which reveals polluters' surprising ability to avoid regulation by temporarily hiding polluting activities during regulatory testing, and suggests that such capability may extend to a much broader set of air polluters. My findings therefore provide updates to the conventional wisdom which usually considers polluters' regulation avoidance as a gradual process of locational and technological adjustments. A policy implication follows that pollution regulation should be designed in ways that envision polluters' ability to take advantage of short-term gaps in the monitoring system.

My results also point to the consequences of conflicts in incentives when pollution monitoring is decentralized to the state governments who, at the same time, bear the regulatory punishments in cases of non-compliance. While this study makes no causal conclusion on the influence of state governments in gaming the EPA's monitoring schedule, the broadly consistent pattern in the correlations between gaming and observable government characteristics such as government efficiency and resources available for air compliance provides suggestive evidence on a potential role of local governments in coordinating strategic responses against the federal monitoring schedule.

Finally, this study proposes a new source of air quality variation driven by the design of the regulatory monitoring system. I provide two examples in which I exploit such variation to study the consequences of air pollution on standardized test performance and criminal activities, two cognitive and behavioral outcomes suggested by the recent literature to be causally influenced by short term air pollution fluctuations.

References

[1] **Auffhammer, Maximilian, Antonio Bento, and Scott Lowe.** 2009. Measuring the effects of the Clean Air Act Amendments on ambient PM_{10} concentrations: The critical importance of a spatially disaggregated analysis. *Journal of Environmental Economics and Management*, 58: 16-26.

[2] **Becker, Randy, and Vernon Henderson.** 2000. Effects of air quality regulations on polluting industries. *Journal of Political Economy*, 180: 379-421.

[3] **Berkman, Mark, Charles Gibbons, and Stephen Lagos.** 2015. A review of the National Fire Incident Reporting System and the National Fire Protection Association's Upholstered Furniture Fire Statistics. <http://firepreventionalliance.org/nfirs.pdf>.

[4] **Chay, Kenneth, and Michael Greenstone.** 2005. Does air quality matter? Evidence from the housing market. *Journal of Political Economy*, 113: 376-424.

[5] **Chen, Yuyu, Ginger Zhe Jin, Naresh Kumar, and Guang Shi.** 2013. The promise of Beijing: Evaluating the impact of the 2008 Olympic Games on air quality. *Journal of Environmental Economics and Management*, 66: 424-443.

[6] **de Marchi, Scott, and James Hamilton.** 2006. Assessing the accuracy of self-reported data: An evaluation of the Toxic Release Inventory. *Journal of Risk and Uncertainty*, 32: 57-76.

[7] **Dietz, Thomas, Kenneth Frank, Cameron Whitley, Jennifer Kelly, and Rachel Kelly.** 2015. Political influences on greenhouse gas emissions from US states. *PNAS*, 12: 8254-8259.

[8] **Donaldson, Dave, and Adam Storeygard.** 2016. The view from above: Applications of satellite data in economics. *Journal of Economic Perspective*, 30: 171-198.

[9] **Foster, Andrew, Emilio Gutierrez, and Naresh Kumar.** 2009. Voluntary compliance, pollution levels, and infant mortality in Mexico. *American Economic Review: Papers & Proceedings*, 99: 191-197.

[10] **Fowlie, Meredith, Mar Reguant, and Stephen Ryan.** *Forthcoming*. Market-based emissions regulation and industry dynamics. *Journal of Political Economy*

[11] **Gates, Guilbert, Jack Ewing, Karl Russell, and Derek Watkins.** 2016. Explaining Volkswagen's emissions scandal. *New York Times*, July 19.

[12] **Glaeser, Edward, and Raven Saks.** 2006. Corruption in America. *Journal of Public Economics*, 90: 1052-1072.

[13] **Grainger, Corbett, Andrew Schreiber, and Wonjun Chang.** 2016. How states comply with federal regulations: Strategic Ambient Pollution Monitoring. mimeo.

[14] **Gray, Wayne, and Jay Shimshack.** 2011. The effectiveness of environmental monitoring and enforcement: A review of the empirical evidence. *Review of Environmental Economics and Policy*, 5: 3-24.

[15] **Greenstone, Michael.** 2002. 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: 1175-1219.

[16] **Greenstone, Michael, John List, and Chad Syverson.** 2012. The effects of environmental regulation on the competitiveness of U.S. manufacturing. mimeo.

[17] **Grooms, Katherine.** 2015. Enforcing the Clean Water Act: The effect of state-level corruption on compliance. *Journal of Environmental Economics and Management*, 73: 50-78.

[18] **Ham, John, Jacqueline Zweig, and Edward Avol.** 2014. Pollution, test scores and distribution of academic achievement: Evidence from California schools 2002-2008. mimeo.

[19] **Henderson, J. Vernon.** 1996. Effects of air quality regulation. *American Economic Review*, 86: 789-813.

[20] **Herrnstadt, Evan, Anthony Heyes, Erich Muehlegger, and Soodeh Saberian.** 2016. Air pollution as a cause of violent crime: Evidence from Los Angeles and Chicago. mimeo.

[21] **Koehler, Dinah, and John Spengler.** 2007. The toxic release inventory: Fact of fiction? A case study of the primary aluminum industry. *Journal of Environmental Management*, 85: 296-307.

[22] **Lee, Hyung Joo, Brent Coull, Michelle Bell, and Petros Koutrakis.** 2012. Use of satellite-based aerosol optical depth and spatial clustering to predict ambient $PM_{2.5}$ concentrations. *Environmental Research*, 118: 8-15.

[23] **Leeson, Peter, and Russell Sobel.** 2008. Weathering corruption. *Journal of Law and Economics,* 51: 667-681.

[24] **Liu, Yang, Meredith Franklin, Ralph Kahn, and Petros Koutrakis.** 2007. Using aerosol optical thickness to predict ground-level $PM_{2.5}$ concentrations in the St. Louis area: A comparison between MISR and MODIS. *Remote Sensing of Environment*, 107: 33-44.

[25] **Lorraine Remer, Y.J. Kaufman, D. Tanre, S. Mattoo, D.A. Chu, J.V. Martins, R.-R.Li, C. Ichoku, R.C. Levy, R.G. Kleidman, T.F. Eck, E.Vermote, and B.N. Holben.** 2005. The MODIS aerosol algorithm, products, and validation. *Journal of the Atmospheric Science*, 62: 947-973.

[26] **Muller, Nicholas, and Paul Rudd.** 2016. What forces dictate the design of pollution monitoring networks? NBER Working Paper 21966.

[27] **Reyes, Jessica Wolpaw.** 2007. Environmental policy as a social policy? The impact of childhood lead exposure on crime. *The B.E. Journal of Economic Analysis & Policy*, 7: Article 51.

[28] **Reynaert, Mathias, and James Sallee.** 2016. Corrective policy and Goodhart's Law: The case of carbon emissions from automobiles. mimeo.

[29] **Shimshack, Jay.** 2014. The economics of environmental monitoring and enforcement. *Annual Review of Resource Economics*, 6: 339-360.

[30] **United States Environmental Protection Agency.** 1997a. Memorandum: Guidelines for granting schedule exemptions for PM_{10} monitoring. <https://www3.epa.gov/ttnamti1/files/ambient/pm25/pm10_wav.pdf>

[31] **United States Environmental Protection Agency.** 1997b. Guidance for network design and optimum site exposure for $PM_{2.5}$ and PM_{10}. EPA-454/R-99-022.

[32] **United States Environmental Protection Agency.** 1999. Guideline on data handling conventions for the PM NAAQS. EPA-454/R-99-008.

[33] **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.

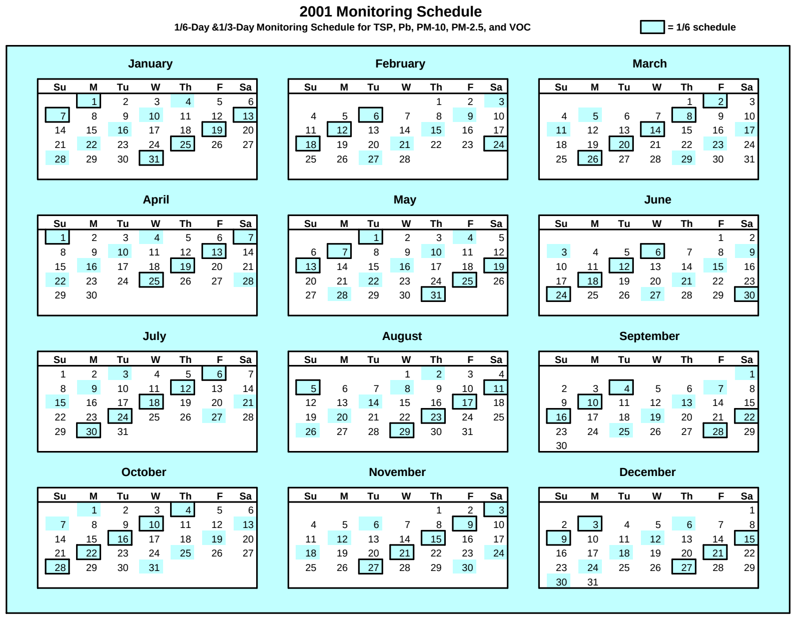
[34] **Voiland, Adam.** 2010. Aerosols: Tiny particles, big impact. <http://earthobservatory.nasa.gov/Features/Aerosols/>

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

[36] **Walker, W. Reed.** 2012. The transitional costs of sectoral reallocation: Evidence from the Clean Air Act and the workforce. *Quarterly Journal of Economics*, 1787-1835.

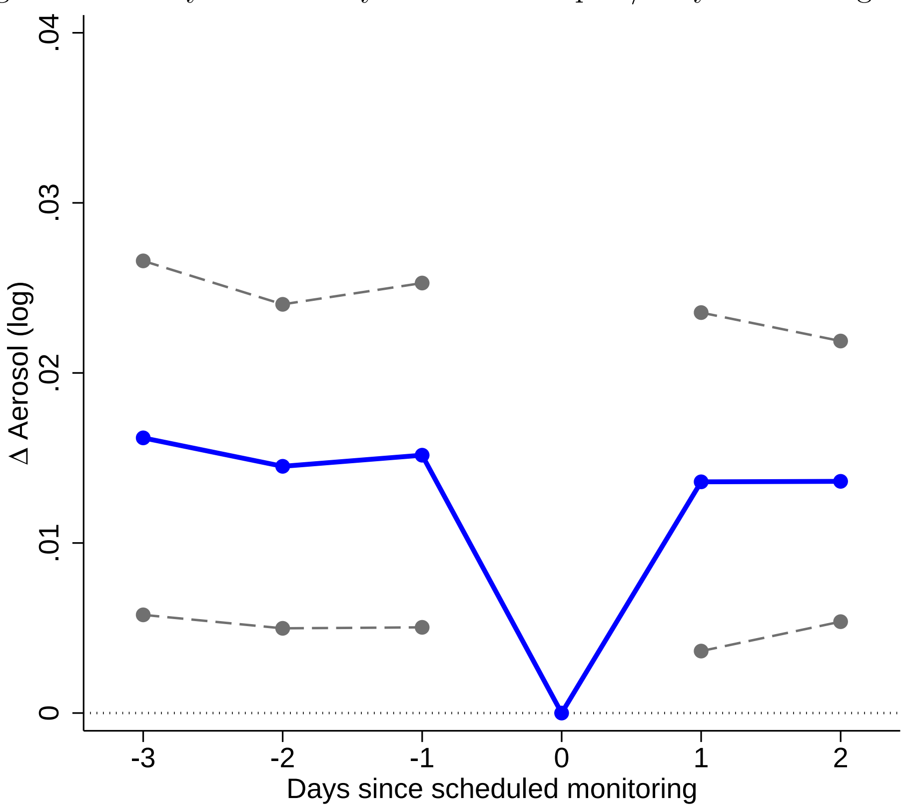
[37] **Zhang, Ying, and Zhengqiang Li.** 2015. Remote sensing of atmospheric fine particulate matter ($PM_{2.5}$) mass concentration near the ground from satellite observation. *Remote Sensing of Environment*, 160: 252-262.

Figure 1: EPA Monitoring Schedule, 2001



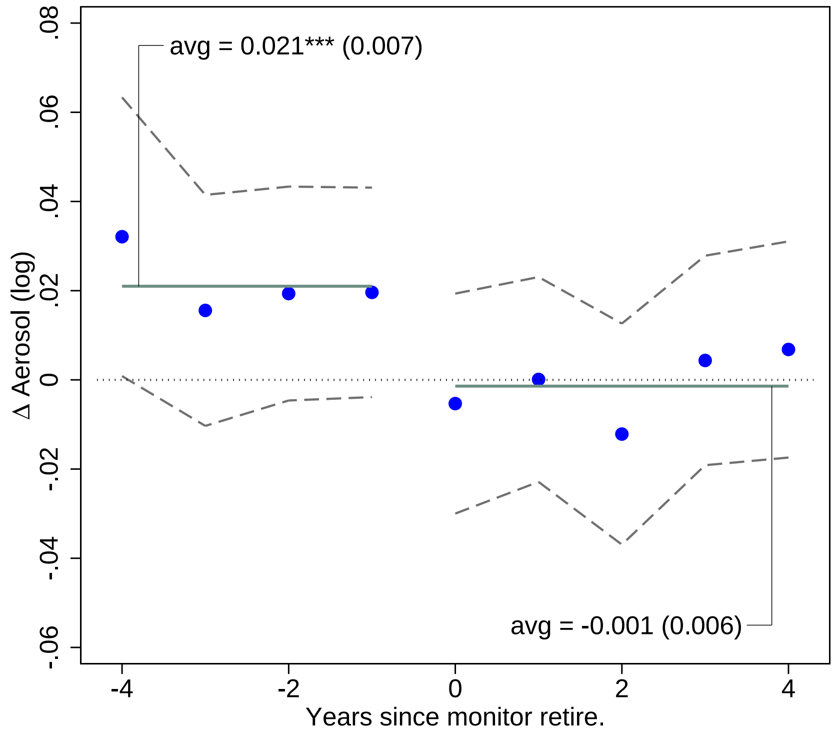
Notes: Figure shows the EPA's 2001 monitoring schedule calendar. Full archives of all calendars can be found here: <https://www3.epa.gov/ttn/amtic/calendar.html>.

Figure 2: Off-days vs. On-days Pollution Gap: 1/6day Monitoring Sites



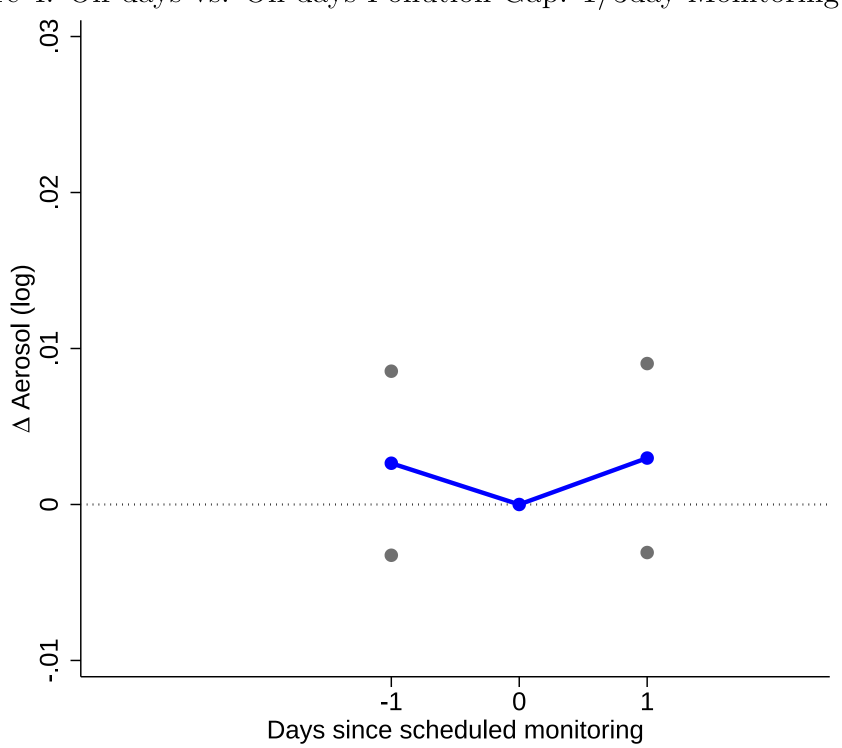
Notes: Figure plots the path of pollution concentration by days of 1/6day cycle. Day 0 corresponds to the scheduled sampling day which is normalized to 0. Pollution is measured by the satellite aerosol concentration within an 10km×10km area that contains a monitoring site. Sample includes all sites that contain at least one 1/6day PM monitor. The regression is conditional on no covariates. Dahsed lines represent 95% confidence intervals constructed using standard errors clustered at the county level.

Figure 3: 1/6day Pollution Gap by Years Relative to Monitoring Site Retirement



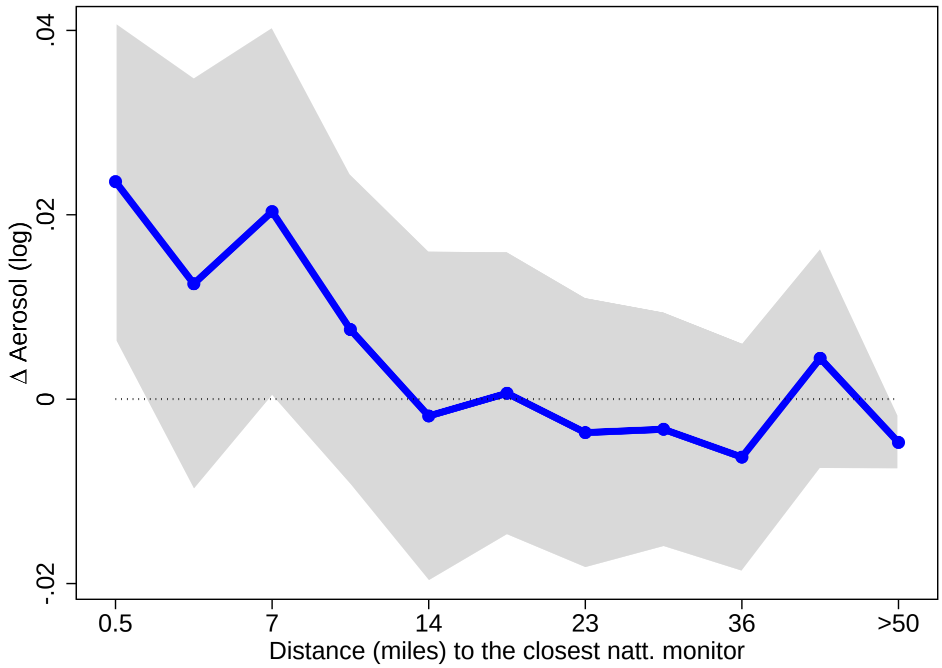
Notes: Figure plots off-days vs. on-days pollution gap as a function of years relative to site retirement. Average estimates show off-days effect separately estimated before and after monitoring site retirement. Sample includes retirement of 490 sites from 2001 to 2013. The regression is conditional on no covariates. Dahsed lines represent 95% confidence intervals constructed using standard errors clustered at the county level.

Figure 4: Off-days vs. On-days Pollution Gap: 1/3day Monitoring Sites



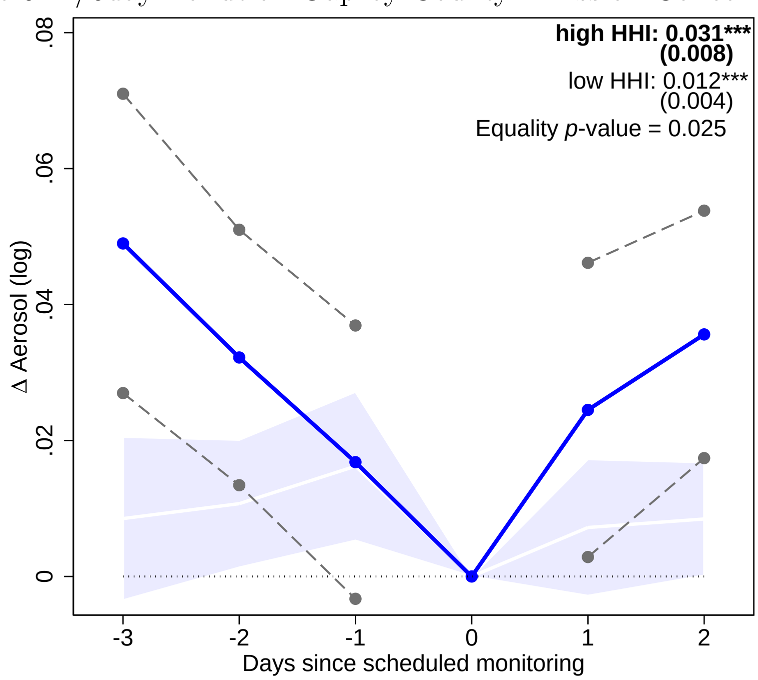
Notes: Figure plots the path of pollution concentration by days of 1/3day cycle. Day 0 corresponds to the scheduled sampling day which is normalized to 0. Pollution is measured by the satellite aerosol concentration within an 10km×10km area that contains a monitoring site. Sample includes all sites that contain at least one 1/3day PM monitor. The regression is conditional on no covariates. Dahsed lines represent 95% confidence intervals constructed using standard errors clustered at the county level.

Figure 5: 1/6day Pollution Gap: Industrial Sources



Notes: Graph shows off-days vs. on-days pollution gap near industrial sources, estimated by facilities' distance to the closest 1/6day non-attainment monitor. The farthest bin groups all facility-monitor pairs that are at least 50 miles apart. All regressions include fixed effects dummies (distance bin, site, year, month-of-year, and day-of-week) and weather controls. Gray shades present 95% confidence intervals constructed from standard errors clustered at the county level.

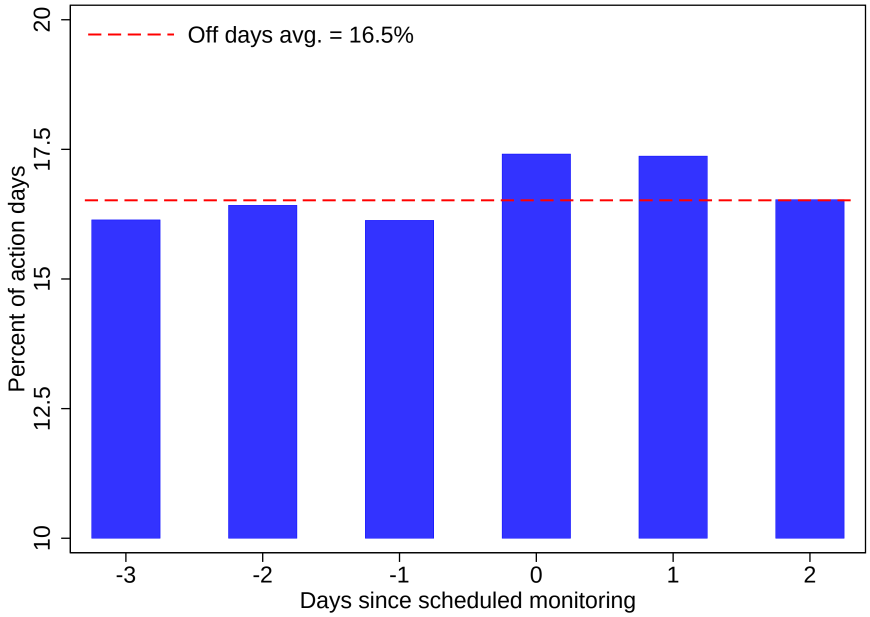
Figure 6: 1/6day Pollution Gap by County Emission Concentration



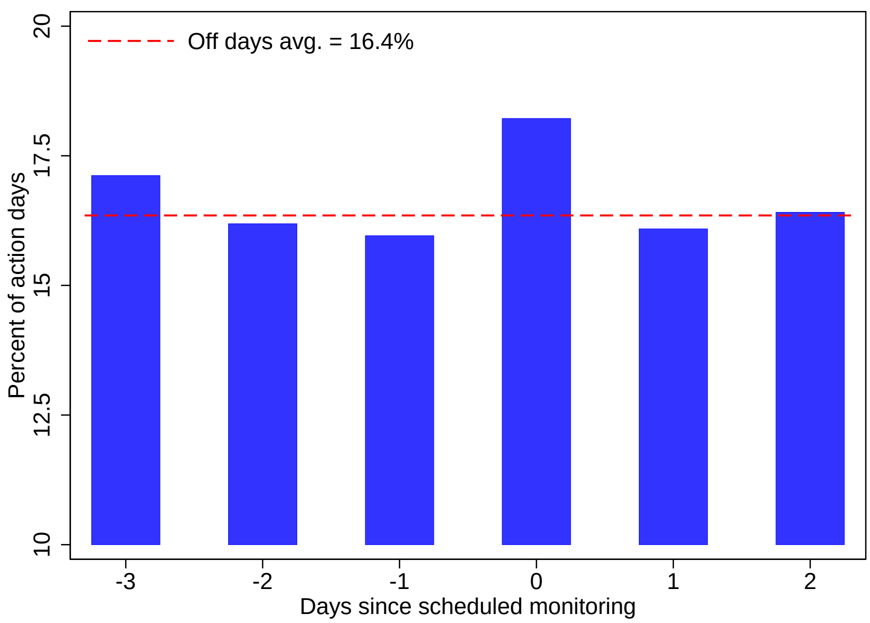
Notes: Figure displays 1/6day pollution pattern separately for high Herfindahl index (≥ 0.9) vs. low Herfindahl index (< 0.9) counties. Estimates are obtained from a single regression. Foreground graph objects represent estimates for the high Herfindahl index counties while the background graph objects show estimates for the rest of the samples. Dashed lines and the shades represent 95% confidence interval constructed from standard errors clustered at the county level. Point estimates shown on the upper-right corner shows average pollution gap. Equality p-value corresponds to the null hypothesis that there is no difference in the off-days effect for the two groups. All regressions include fixed effects dummies (site, year, month-of-year, and day-of-week) and weather controls.

Figure 7: Issuance of Action Day Advisories by Days in 1/6day Monitoring Cycle

Panel A: Raw distribution

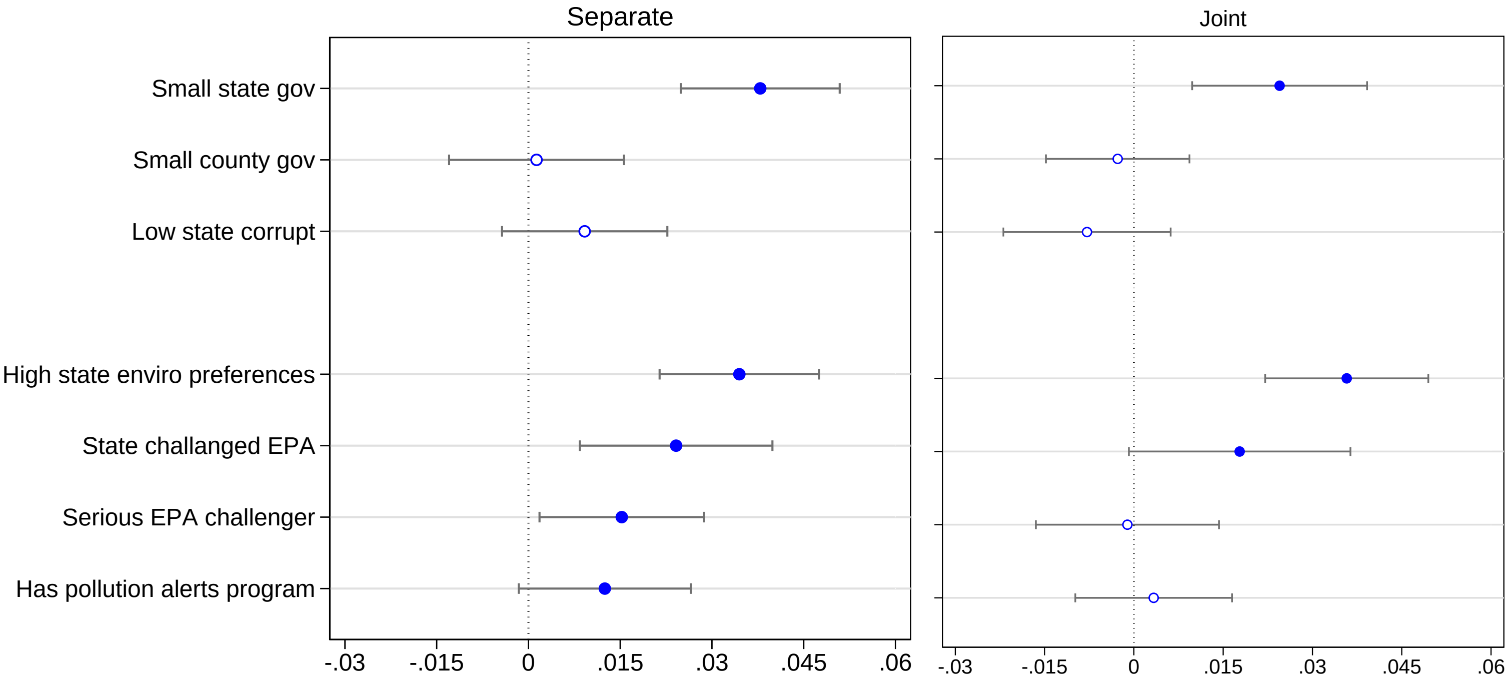


Panel B: Consecutive Action Days adjustment



Notes: Panel A shows the raw distribution of air pollution action day issuance by days in the 1/6day cycle. Day 0 marks the scheduled sampling day. Panel B shows adjusted distribution by including only the first issuance for consecutive action days episode. The sample includes all action day issuances reported to the EPA's airnow program from 2004 to 2013, aggregated to the core based statistical area (CBSA) by day level. See the text for more details.

Figure 8: 1/6day Pollution Gap by Local Government Characteristics



Notes: Graph reports interaction coefficient between off-days dummy and government characteristics including dummies for (from north to south): above median state government size, above median county government size, below median state corruption, above median state environmental preferences, whether state challenged EPA's preliminary PM designation from 1997-2012, above median effort in challenging letters, whether state has any air pollution Action Day programs. Left panel shows separate regressions, each including interaction for one characteristic at a time. Right panel shows a joint regression where all interaction terms are included simultaneously. Gray bars show 95% confidence interval constructed from standard errors clustered at the county level.

Table 1: Summary Statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	pollution stats				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. Column 1, 2, and 3 shows 25^{th}-percentile, mean, and 75^{th}-percentile values of grid-daily level aerosol concentration. Column 4 shows total number of grid-daily observations. Monitor sample includes all monitors that collected enough samples during the year to be considered eligible for NAAQS comparison. Column 5 reports 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 details). Column 6 counts number of 1/6day monitors, defined by monitors that are required to sample either 60 or 61 days of PM data for each calendar year. Column 7 counts number of 1/6day 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 number of sites that contain at least one PM monitor that follows the 1/6day schedule. Column 10 and 12 aggregate site counts in column 8 and 9 to the county level, respectively. Column 11 and 13 report corresponding population that lived in the monitored counties.

Table 2: Monitor Sampling Compliance

	(1) Samples required	(2) Samples taken	(3) Fraction taking $\geq$90% required samples	(4) Fraction taking 100% required samples
1/6day monitors	60 or 61	58.4 [2.2]	96.74%	19.21%
1/3day monitors	121 or 122	115.6 [4.4]	94.72%	5.42%
1/1day monitors	365 or 366	349.1 [13.0]	92.54%	6.33%

Notes: Statistics are computed from monitor-year observations. Sample includes all monitors eligible for NAAQS comparison. Standard deviation in brackets.

Table 3: Off-days vs. On-days Pollution Gap: 1/6day 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
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. Column name indicates the sample used. Column 1 & 2 use all sites that have at least one 1/6day PM monitor. Column 3 includes sites that have standalone 1/6day monitor. Column 4 includes sites in counties with only 1/6day 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 4: Off-days vs. On-days Pollution Gap: "Placebo" Sites

Dep. var. = Aerosol concentration (log)			
	(1) Sample: retired 1/6d sites	(2) Sample: continuously monitoring sites	(3) Sample: Non-regulatory 1/6d sites (HAPs)
off-days	-0.0020	0.0023	0.0023
	(0.0046)	(0.0080)	(0.0044)
Power$_{(1.5\%\text{ effect, }5\%\text{ sig.})}$	0.940	0.803	0.910
N	372,989	231,532	370,020
N (site)	490	556	792

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

Table 5: Off-days vs. On-days Pollution Gap: 1/3day Monitoring Sites

Dep. var. = Aerosol concentration (log)				
	(1) Sample: sites w. any 1/3d monitor	(2) Sample: sites w. any 1/3d monitor	(3) Sample: sites w. only 1/3d monitor	(4) Sample: counties w. only 1/3d monitor
off-days	0.0028	0.0029	0.0024	0.0054*
	(0.0026)	(0.0020)	(0.0025)	(0.0030)
Ctrls		✓	✓	✓
N	598,859	598,859	386,854	244,071
N (site)	1,064	1,064	849	562

Notes: Each column reports a separate regression. Column name indicates the sample used. Column 1 & 2 use all sites that have at least one 1/3day PM monitor. Column 3 includes sites that have standalone 1/3day monitor. Column 4 includes sites in counties with only 1/3day 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 6: 1/6day Pollution Gap: Non-attainment Sites

Dependent variable:	Aerosol concentration (log)		
		Standard:	
	$PM_{2.5}$ 24 hr (1)	$PM_{2.5}$ annual (2)	PM_{10} 24 hr (3)
Panel A. By attainment & non-attainment			
off-days × $\mathbb{I}$(attainment)	0.016***	0.012**	0.018***
	(0.006)	(0.006)	(0.004)
off-days × $\mathbb{I}$(non-attainment)	0.034**	0.018	-0.0010
	(0.016)	(0.014)	(0.0162)
Equality p - value	0.273	0.661	0.266
Panel B. By number of exceedance in the past two years			
off-days × $\mathbb{I}$(#violation last two yrs = 0)	0.014***	0.015***	0.015***
	(0.004)	(0.004)	(0.004)
off-days × $\mathbb{I}$(#violation last two yrs = 1)	0.015	0.015	0.023**
	(0.012)	(0.013)	(0.011)
off-days × $\mathbb{I}$(#violation last two yrs = 2)	0.053***	0.025*	0.013
	(0.018)	(0.015)	(0.019)
Equality p - value	0.083	0.803	0.725
N	685,060	685,060	685,060
N (site)	1,193	1,193	1,193

Notes: Each column-panel reports a separate regression. Column names indicate regulation standard used to determine site attainment status. In panel A, the off-days dummy is interacted with indicators of the monitoring site's violation of the NAAQS standards. In panel B, the off-days dummy is interacted with the monitoring site's number of NAAQS exceedance in the past two years. 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 7: 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. Column 1 to 4 (5 to 8) present statistics for English (math) tests. Column 1 and 5 report overall statistics. The remaining columns report statistics by month of test administration. Standard deviations are reported in brackets.

Table 8: 1/6day Gap: Standardized Test Performance (CAHSEE, California)

Independent variable: Indicator for taking an exam on an off-day					
	(1)	(2)	(3)	(4)	(5)
	Sample:		Sample:		
			Schools close to natt. sites		
	All schools		0-10 miles	10-50 miles	>50 miles
Scale score (std.)	-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
Test taken (log)	-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
Month-of-year ctrls.	✓				
Full ctrls.		✓	✓	✓	✓

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

Table 9: Summary Statistics: Daily Crime Rates

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

Notes: Column 1 presents statistics for counties included in the estimation, i.e. counties with 1/6day sampling and 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 10: 1/6day Gap: Crime

	(1)	(2)	(3)	(4)	(5)
	Sample: Counties w. 1/6d sampling	Sample: 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.					
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.					
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 restrict 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 restricting to counties included in the crime regressions. Column 3 to 5 present crime regression results. Panel A reports estimation with no covariates. Panel B reports estimation 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. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Online Appendix

Unwatched Pollution: The Effect of Incomplete Monitoring on Air Quality

Eric Zou

March 2017

PRELIMINARY; COMMENTS WELCOME!

Contents

A. Additional Background on Particulate Matter Regulation

This section provides more institutional and administrative details about the EPA's particulate matter (PM) monitoring practice. I first describe steps involved in obtaining PM samples. Next, I discuss determinants of schedule assignments and monitor placements. Finally, I introduce newly available continuous PM monitoring and their relationship with traditional periodic PM samplers.

A.1. Particulate Matter Sampling Procedures

The federal EPA outlines the practice standard for PM sample handlings in the *Quality Assurance Handbook for Air Pollution Measurement Systems* (US EPA, 2013). Manual sampling of PM is a delicate procedure that demands great care. Local monitoring agencies are advised to give "particular attention" to the handling of filters for PM as the process of filter handling is understood to be a major source of measurement error.

Filters are first pre-weighed before they are taken to the monitoring site to collect samples. They are then transported to the monitoring site where sampling takes place. After samples are collected, filters must be carefully removed from the monitoring device, placed in labeled, nonreactive containers, and sealed. Samples are then delivered to the laboratory for analysis, usually on the same day that the samples are taken. The integrity of PM samples are sensitive to a variety of factors such as temperature extremes, air pressure, and the physical handling such as packing and jostling. As a consequence, local monitoring agencies are required to develop standard operating procedures that take these considerations into account on a site-by-site basis. Also, the monitoring agency's personnel who has "custody" of the samples on each sampling day needs to make sure the security of the sample and that no tampering occurred. Because PM samples may be transferred among multiple parties through various stages of storage, processing, and analysis at the laboratory, a written "Chain of Custody" (COC) record form must exist that accompany the samples at all time from the field to the laboratory, listing the locations of the samples and the corresponding custodians.

A.2. Assignment of Sampling Schedules and Monitor Placement

Because manual PM sampling is costly, many monitoring sites employ periodic sampling framework. Other than the once every six days (1/6day) schedule studied in this paper, two other frequently used schedules are the 1/3day and the 1/1day (i.e. daily) schedules. As discussed in the main text, whereas states are granted the authority to carry out pollution monitoring, assignment and revisions of sampling frequency are determined by the regional EPA office which administers several states. For the current regional EPA delineation, see <https://www.epa.gov/aboutepa/visiting-regional-office>. In general, more frequent schedule is assigned to sites with higher chances of violating the NAAQS. To illustrate this, Figure D.8 reports results from a multinomial logistic model of selection into different PM sampling schedules. Probability of adopting the 1/6day schedule decreases with increasing PM concentrations both in terms of the annual metric and the 24-hour metric. Below I provide more details about the administration of sampling frequency assignment and revisions. I provide separate discussion for PM 2.5 sites and PM 10 sites as the rules for frequency assignment and revisions are slightly different.

PM 2.5 sampling frequency. In principle, all PM 2.5 samplers are required to sample at least once every three days (40 CFR Part 58). Individual sites can also request EPA Regional Administrator for reduction to once every six day schedule on a case-by-case basis. The EPA Regional Administrator may grant sampling frequency reductions after consideration of factors (including but not limited to the historical PM 2.5 data quality assessments, the location of current PM 2.5 design value sites, and their regulatory data needs) if the Regional Administrator determines that the reduction in sampling frequency will not compromise data needed for implementation of the NAAQS.

A PM 2.5 sampler may also follow the 1/6day schedule if it is a collocating sampler to a 1/3day or a 1/1day sampler. By the EPA's regulation, for each reporting organization (usually a state), 25% of its PM samplers are required to be collocated with an identical samplers to estimate data precision, and these collocating samplers sample at the 1/6day rate (40 CFR Part 58). This rate dropped to 15% in March 2003, when EPA decided that reduced collocation rate would not significantly deteriorate precision estimation. In principle, PM data collected by collocating samplers should not be used toward NAAQS comparison, unless the corresponding main sampler malfunctioned or did not collect a valid sample on a sampling day. Also, states should be clear about which samplers are collocators when reporting data to the AQS. Specifically, collocating samplers should all have a Parameter Occurrence Code (POC) of “2” in the AQS data whereas the main sampler has a POC of “1”. In practice, however, states had substantial misconceptions about how data from collocating samplers should be treated, e.g. in some cases states reported collocators' PM data for NAAQS comparison even when the main sampler has already collected valid samples; wrong POCs were also assigned to samplers. See EPA's memorandum *Use of Collocated PM2.5 Data and Parameter Occurrence Codes (POCs)* which can be found here: <https://www.epa.gov/sites/production/files/2015-09/documents/25colo_0.pdf>. For this reason, in the main analysis I do not attempt to identify and exclude collocating PM samplers from the estimation sample. I do confirm that near sites with standalone 1/6day samplers (i.e. sites where the 1/6day sampler must not be a collocator) the gaming effect is stronger.

PM 10 sampling frequency. The 1997 revision of PM NAAQS sets sampling frequency to a minimum of once in three days for all PM 2.5 and PM 10 sites. But for PM 10 monitoring, an exemption can be granted to a site that reduces the sampling frequency to once in six days if it can be shown that there is "little chance that the daily PM 10 standard will be exceeded" (U.S. EPA, 1997a). Specifically, a site is eligible for the exemption if a one-tail *t*-test of the difference between 3-year 99^{th} percentile value and the 24-hour standard of 150 ug/m^3 plus five is significant at the 10% level. In cases where this criteria cannot be satisfied, a site can still be considered eligible for exemption if the ratio of 3-year mean to the mean standard of 50 ug/m^3 is smaller than the ratio of 3-year 99^{th} percentile to the max standard of 150 ug/m^3 so that the mean standard is the "controlling standard".

A.3. Continuous Monitoring Technology

The past decade has seen enormous development of continuous PM monitoring technologies. In this subsection I briefly introduce some of these technologies and review the main barriers that prevent them from replacing the traditional manual PM sampling. I will focus on PM 2.5 monitoring, the focus of most of the innovations.

Manual sampling of PM2.5 acquires deposits over a 24-hour period on a Teflon-membrane filter from air drawn at a controlled flow rate through the Well Impactor Ninety Six PM2.5 inlet. If done appropriately, manual sampling obtains the most accurate measure of ambient PM 2.5 concentrations. In the EPA's language, this method provides the "reference" measure of PM 2.5 and is named the Federal Reference Method (FRM). Performance of any continuous monitoring technology is judged by its ability to replicate monitoring results from the FRM method. Below I cite descriptions of two most commonly used continuous technologies and their monitoring method from the EPA's 1998 *Guidance for Using Continuous Monitors in PM 2.5 Networks* which can be found here: <https://www3.epa.gov/ttnamti1/files/ambient/pm25/r-98-012.pdf>

Tapered Element Oscillating Microbalance (TEOM). "*Particles are continuously collected on a filter mounted on the tip of a glass element which oscillates in an applied electric field. The glass element is hollow, with the wider end fixed; air is drawn through the filter and through the element. The oscillation* *of the glass element is maintained based on the feedback signal from an optical sensor. The resonant frequency of the element decreases as mass accumulates on the filter, directly measuring inertial mass. The typical signal averaging period is 10 minutes. Temperatures are maintained at a constant value, typically 30°C or 50°C, to minimize thermal expansion of the tapered element.*"

Beta Attenuation Method (BAM). "*Beta rays (electrons with energies in the 0.01 to 0.1 MeV range) are attenuated according to an approximate exponential (Beer's Law) function of particulate mass, when they pass through deposits on a filter tape. Automated samplers utilize a continuous filter tape, first measuring the attenuation through the unexposed segment of tape to correct for blank attenuation. The tape is then exposed to ambient sample flow, accumulating a deposit. The beta attenuation measurement is repeated. The blank- corrected attenuation readings are converted to mass concentrations, with averaging times as short as 30 minutes.*"

Why does manual sampling of PM 2.5 sampling still dominate when these continuous monitoring technologies are available? I list some major barriers below. First, although some continuous technologies can provide reasonable proxies of PM 2.5 concentrations, their performance varies significantly across space and time. For example, the ability of both TEOM and BAM to provide FRM-comparable data compromises when the sampled aerosol is not stable. It is known that when the sampled PM 2.5 deposits contain a high fraction of volatile components, both TEOM and BAM sensors measure reduced amount of mass relative to the FRM method. Employment of continuous technologies therefore requires substantial validation efforts before the data can be used toward NAAQS comparison. Second, current regulation (40 CFR 58, Appendix D, Section 2.8.1.3.8) requires continuous PM2.5 monitors to be operated in large US metropolitan areas. However, data obtained from these monitors are only intended to be used for public reporting and forecasts of PM2.5 concentrations, not for NAAQS comparison. In other words, although states are required to implement continuous monitoring for public advisory purposes, they can choose to keep using data from manual sampling to show NAAQS compliance.

B. Additional Data Descriptions

This section provides more details on main variable construction as well as descriptions of secondary data sources.

B.1. Satellite Data

The satellite pollution measure from MODIS is called aerosol optical depth (AOD), which is a measure of the degree solar beam transmission is absorbed or scattered by atmospheric aerosols. The original measure has a range of -0.05 to 5, with smaller values corresponding to lower level of aerosols and therefore less air pollution. In all analyses, I multiply the original measure by a factor of 100 to reduce redundant decimals in presentation.

The key outcome variable of this study is a panel dataset of daily aerosol level with a 10km×10km grid spatial resolution. This variable is constructed from daily aerosol raster files which are produced at the spatial resolution of 10km×10km pixel array. In order to create the grid level panel dataset, original satellite pixels must be mapped onto a series of 10km×10km grids which correspond to the same ground areas over time. To execute the mapping, I first re-grid daily aerosol rasters into 0.1km×0.1km pixel sizes, and then map them onto a 10km×10km gridded map of the contiguous US provided by the US National Grid Information Center where the grid boundaries are fixed over time. In other words, the aerosol level for each 10km×10km grid-day is computed as the average aerosol level of all 0.1km×0.1km pixels that fall within the grid on that day. This procedure ensures that the grid dataset preserves the original resolution of the satellite rasters, and that each grid tracks aerosol levels for the same area over time. Figure D.1 provides a map of 2001-2013 average grid aerosol level for the lower 48 states.

Existing literature has documented a strong correspondence between the MODIS aerosol measure and ground level PM, which is a primary motivation for using it to detect gaming against the PM monitoring schedule (Liu, Franklin, Kahn, and Koutrakis, 2007; Lee, Coull, Bell, and Koutrakis, 2012; Zhang and Lee, 2015). As a replication of the aerosol-PM relationship in my study context, I correlate monitor-daily level PM concentrations to the daily aerosol level within the 10km×10km grid where the monitor falls in. Figure D.2 plots the distributions of standardized PM 2.5 and PM 10 concentrations within 9 aerosol bins. Across the distribution I find the raw correlation between PM and aerosol to be positive and fairly linear. Moreover, the correlation is stronger for PM 2.5 than for PM 10. I report regression estimates in Table D.1, Panel A, column 1 shows that the raw correlation is 0.028 standard deviation increase in PM 2.5 per unit increase in aerosol level. Column 2 and 3 adds increasingly flexible time and geographic controls, and the coefficient estimates stay stable. The magnitude is in line with previous atmospheric science studies which use various modeling strategies to estimate the AOD-PM relationship. Table D.1, panel B reports that the correspondence with coarse particulate matter is much weaker, where a unit increase in aerosol level is associated with 0.0067 standard deviation increase in PM 10 concentration.

B.2. Polluters Data

I draw polluter information from the following sources.

First, I obtain annual observations of polluters' location and reported total emission from the EPA's Toxic Release Inventory (TRI). By the 1986 Emergency Planning and Community Right-to-Know Act (EPCRA), a facility is required to report to the TRI if it satisfies all three of the following requirements: (1) it is included in a EPCRA-listed North American Industry Classification System (NAICS) code, which includes mining (NAICS 212), utilities (NAICS 221), Manufacturing (NAICS 31-33), Hazardous Waste (NAICS 562) among others. All federal facilities are included regardless of industry (in the data, more than 50% of the federal facilities are in the industry of national security (NAICS 928)); (2) it has at least 10 full time employees; and (3) it processes more than 25,000 pounds or uses in production more than 10,000 pounds of EPCRA-listed toxic pollutants during the year. At the time of this writing, the list contains about 690 individual pollutants. Key variables contained in the TRI are facility latitude and longitude, self-reported annual stack and fugitive emissions, and NAICS code.

As an alternative to the TRI, I use data from the EPA's 2011 National Emission Inventory (NEI). The data is created to support the EPA's National Ambient Air Quality Standards program under the Clean Air Act. Maintained by the EPA's Office of Air Quality Planning and Standards, the NEI combines polluter information from a variety of data sources, including the TRI, and provides the most comprehensive list of polluters in the US. The advantage of this data over the TRI is that it allows me to directly observe PM emitters. The disadvantage of the NEI 2011 data is that it only provides a snapshot of polluters in 2011 and so I'm forced to assume that polluter profiles stay unchanged over the study period of 2001 to 2013. Alternatively, I can restrict estimation to the year of 2011. In the analysis below I report both results.

Like in the monitor level analysis, air pollution near the polluters are inferred by the aerosol levels observed within the associated 10km×10km grids. Due to numerosity of polluters observed in the data, many grids naturally contain multiple polluters. To avoid repetitive observations, I first aggregate up the TRI (NEI) data from the polluter-year level to the grid-year level. For grids that contain multiple polluters, the “representative” polluter is coded to locate at the centroid of the grid. Each grid-year is then linked to PM monitors nearby using facility latitude and longitude (in cases of single-facility grids) and grid centroid latitude and longitude (in cases of multiple facility grids).

B.3. Controlled Burning and Forest Fire Data

As will be discussed further below in section C, a part of the analysis attempts to use fire incidents data to test whether prescribed burning schedules are strategically adjusted to avoid the EPA's monitoring. Here I describe the main data sources used in this analysis.

First, I observe controlled burning events from the Federal Emergency Management Agency's (FEMA) National Fire Incident Reporting System (NFIRS) Version 5.0. For each fire incident, NFIRS contains a type code which I use to identify controlled burning (incident code 632). Other key variables included in the NFIRS are the ZIP Code location and date of the incident. In computing distance from fires to PM monitors, I assume that all fires occur at the centroid of the ZIP Code location. Using NFIRS I identify 42,449 controlled burning incidents at the ZIP Code-daily level from 2001 to 2013.

There are two potential limitations with the analysis on prescribed burning. The first is the low statistical power to detect a meaningful effect due to the rareness of burning events, which I will illustrate in the analysis. Another inherent limitation is that the NFIRS database is neither a random sample nor a census of fires. In particular, fire departments in urban areas are overrepresented relative to those in rural areas (Berkman, Gibbons, and Lagos, 2015). In supplementary analysis described below I supplement the NFIRS data with forest fire records obtained from the National Fire and Aviation Management. This data combines administrative records from seven major fire and wildland management agencies including Bureau of Indian Affairs, Bureau of Land Management, Bureau of Reclamation, California Department of Forestry and Fire Protection, National Park Service Fire and Aviation Management, US Fish & Wildlife Service, and Forest Service. The advantage of this data is that it may provide a better coverage of fires occurred on wildland than the NFIRS does. However, it does not contain any identifiers for fires caused by control burning. I use this data to identify 443,393 wildland fire events at the ZIP Code-daily level from 2001 to 2013.

C. Additional Analysis

This section reports results from additional analysis mentioned in the main text. Except for robustness checks, much of the analysis reported here is conceptually important and interesting in its own right, but has limitations, such as limited statistical power, in actual implementation. This section keep records of the methods used and the results obtained, along with my interpretation.

C.1. Industry-Specific Effects

Section 4.1 of the main text shows that air pollution pattern near industrial polluters show a 6-day pattern that echoes the PM monitoring cycle, an effect that is particularly strong near nonattainment PM monitors. In the analysis I pool all polluters together, and an immediate extension of the exercise would be to explore industry-specific effects by estimating the same model separately for polluters in different industries. Such analysis may bring insights as of which industries are driving the observed gaming effect and therefore provides guidance for future pollution monitoring policy enforcement.

Table D.2 reports gaming effect estimates by industries included in the TRI. Each row reports a separate regression that uses all facilities in a 3-digit NAICS industry. For each industry regression, the off-days dummy is interacted with indicators for facility-monitor distance bins. That is, each facility is linked to the closest non-attainment 1/6day PM monitor, and off-days effect is estimated when the facility-monitor pair is < 15 miles (column 1), 15-50 miles (column 2), and > 50 miles (column 3) apart. Column 4 reports how many counties that the facilities span, and column 5 reports total number of observations included in the regressions. Results show that, first, the off-days effect is in general strongest when the facility is less than 15 miles away from the closest non-attainment monitor. Most coefficient estimates for facility-monitor pairs > 15 miles apart are small (44 out of 58 coefficients show less than a 1% effect) and statistically insignificant (53 out of 58 coefficients insignificant at the 1% level). Second, there is substantial heterogeneity across industries in gaming. Largest gaming is observed near gas extraction plants with a 19.6% off-days vs. on-days pollution gap, along with many manufacturing plants such as beverage & tobacco, paper, food, printing, leather, and transportation equipment production facilities. I also find gaming near non-military federal facilities, where pollution is 5.3% higher on off-days than on on-days.

Regarding the interpretation of the industry-specific results, two caveats are in order here. First, whether gaming is expected for facilities in a given industry depends not only on observed factors such as the PM intensity of emissions but also on unobserved factors such as the technological capacity to engage in short run gaming. While certain industries such as electric utilities and fossil fuel refineries certainly emit high level of PM, many facilities in these industries operate 24/7, which reduces the discretion to frequently ramp up and down production. Therefore, it is not entirely clear *a priori* which industries should be associated with the strongest gaming. Second, the satellite resolution determines that no differential impacts can be distinguished among facilities that locate within a same 10km×10km grid area. Therefore, the observed industry heterogeneity blends in industrial agglomeration patterns that are not explored here.

To further illustrate the point, I provide one potential correlate of industry gaming, focusing the subset of facilities in the manufacturing industry. For each 3-digit NAICS industry, I relate gaming to average capacity utilization from the Quarterly Survey of Plant Capacity Utilization (QPC) provided by the US Census Bureau since 2008. Each quarter, QPC surveys over 7,000 establishments with five or more production workers selected from the Census Business Register. Probability of selection is based on establishments' value of shipments within each industry. For each 3-digit NAICS industry, QPC provides statistics (mean and standard errors estimates) on the rate of capacity utilization defined as the ratio of a manufacturer's production to their full production. From 2008 to 2013, average capacity utilization rates range from 57.25% (NAICS=321, wood product manufacturing) to 86.08% (NAICS=322, paper manufacturing). Figure D.3 presents a simple scatterplot of industry-specific gaming coefficient against the industry capacity utilization rate. Graph shows a generally negative correlation, except for an outlier which is the paper manufacturing industry.

C.2. Robustness: Gaming Near Industrial Sources

In the main text, I show that a similar 1/6day pollution gap is find near polluting plants. The effect exhibits a distance gradient, where the gap closes as the distance between the polluter and the nearest non-attainment monitor increases. This subsection reports three sets of results that either extend or support the main analysis.

First, I report a more saturated model where the pollution gap is not only allowed to vary by the polluter-monitor distance, but also by whether the polluter and the monitor reside on the same side of the state border. Because consequences of NAAQS violation are only felt within the state border, polluters need not game monitors in neighbor states. Figure D.4 reports the result. The right hand side of the graph reports coefficients for same-state polluter-monitor pairs. Like in the main estimates, a clear distance gradient is identified, where the pollution gap is mostly identified when the polluter is less than 15 miles away from the closest non-attainment monitor. The left hand side of the graph shows that the same estimation on different-states polluter-monitor pairs yield no significant pattern. As a precautious note, the different-states analysis is based on a much smaller sample, as shown by the bar charts imbedded in the same figure. This is largely due to the fact that states tend not to site monitoring stations along state borders. Due to the power limitation, the same-state and different-states comparison of the pollution gap estimates is not entirely conclusive.

Second, because polluters in my analysis are drawn from the EPA's Toxic Release Inventory (TRI) which only contains polluters who emitted above-threshold amount of toxicants, one might concern about external validity with regard to effects near other PM emitters. I report a robustness check where I draw polluters from the 2011 National Emissions Inventory (NEI) which is a near census of all polluters for the 2011 snapshot. In Figure D.5 I show two specifications, one where I assume polluter profile stay constant from 2001 to 2013 (panel A), and the other where I restrict estimation to the year of 2011 (panel B). Results are both similar to the main analysis.

Third, because the TRI's plant-emission data used to create the HHI are found to contain measurement error due to biased reporting, I report a robustness check where the HHI is replaced with an indicator variable that simply indicates whether a single polluter is observed in the county. Figure D.6 reports the robustness results. Analysis using simple count of polluters as a measure of emission concentration replicates the HHI-based heterogeneity analysis.

C.3. Effects of Incomplete Monitoring on Controlled Burning

In this section, I expand the analysis on sources of gaming by looking at mobile pollution sources. Since it's not possible to track mobile sources over time using the satellite measure, in this subsection I use a different approach that detects gaming by searching for abnormal patterns in occurrences of mobile pollution events. I focus on control burning (a.k.a. prescribed fires), a tool used in forest management and farming that intentionally starts fires to reduce excessive fuel accumulations and decreases the hazard of large fires. Two features make control burning a good context to examine monitor gaming. First, burning directly produces significant particulates pollution. Second, dates for burning are chosen. A burning plan must be approved prior to ignition. I rely on detailed fire records data to test the hypothesis that timing of control burning has been strategically distributed across on-days and off-days to game the monitoring schedule.

As described in section B, I identify control burning events from the Federal Emergency Management Agency's (FEMA) National Fire Incident Reporting System (NFIRS Version 5.0). I also employ an alternative dataset from the National Fire and Aviation Management which combines forest fire records from seven major fire and wildland management agencies. These include Bureau of Indian Affairs, Bureau of Land Management, Bureau of Reclamation, California Department of Forestry and Fire Protection, National Park Service Fire and Aviation Management, US Fish & Wildlife Service, and Forest Service.

The key identification here compares likelihood of control burning events between off-days and on-days near monitoring sites that follow the 1/6day schedule. To operationalize this, I first create a relational database of sites and burning. I link each site-day to every burning events that occurred in the US on that day. In other words, the database is a daily panel of all 1/6day sites with each site-day expanded to all burning events on that day. Each observation therefore corresponds to a site-day-burning. For each observation I compute the distance between the burning and the site, and I also observe whether the burning occurred within the county (or state) of the monitoring site. All following analyses are based on reduced (i.e. collapsed) version of this relational database.

I test whether there are more burning events on off-days than on-days in the administrative area (county or state) where the monitoring site lives. I take all within county (or state) site-burning pairs and collapse the data to a balanced panel at the site-daily level. For each site-day, I create a dummy variable which indicates whether any burning occurred within the site's county (or state). This dummy variable is then used as the outcome variable. Results are reported in Panel A of Table D.3. The coefficient on off-days represents the gap in likelihood of burning between an average off-day and an average on-day. Column 1 and 2 shows no evidence that burnings occur differentially more frequently on off-days. The coefficients are not very precisely estimated, with the 95% confidence interval able to reject a 5% effect out of the mean. Column 3 and 4 replicate the analysis using a subsample that contains all non-attainment sites. Again, I find no statistically significant evidence of gaming.

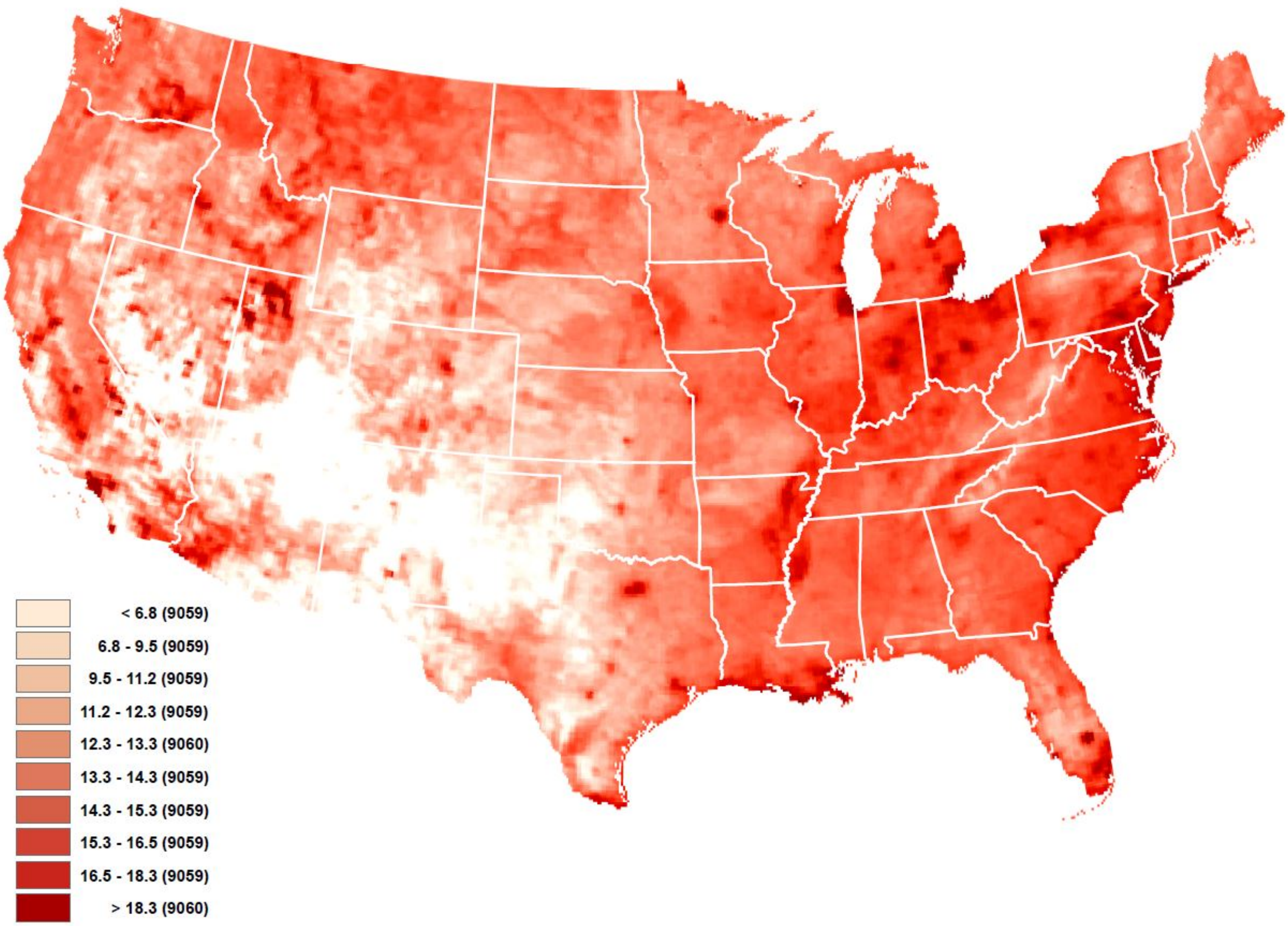
Analysis of prescribed burning suffers from insufficient statistical power due to the rareness of prescribed burning events. Whereas the analysis is based on more than 2 million observations, simple power calculation shows that the power of detecting a 1.5% effect on within-county burning at a 5% significance level, i.e. an effect size similar to the main pollution regression, is only 11.7% (panel A, column 1). Statistical power improves significantly to 97.9% for within-state burning analysis (panel A, column 2), but the test has potentially discarded geographic information, such as proximity to monitors. Not surprisingly, the tests using the non-attainment sites sample in column 3 and 4 are also underpowered, with less than 10% odds of detecting a 1.5% effect at a 5% significance level.

Table D.3, panel B reports a replication of the analysis using occurrence of forest fire as the outcome variable. As forest fires occur at incidents rate that are almost an order of magnitude larger than prescribed burning, tests of more fires on the off-days improves substantially on the statistical power side. Results again show no significant change in fire incidents during off-days, with precisely estimated standard errors. The downside of the analysis is that the outcome variable mixes in non-intentional fires which may mask the responses from prescribed burning.

To sum up, results in this subsection show no significant evidence of gaming of the EPA's PM monitoring schedule through strategically adjusting timing of prescribed burnings. This analysis, however, cannot rule out the possibility that more powerful tests could detect modest but significant responses. One such example would be that selective approval of control burning plan takes into account wind direction. In that case, an average effect estimate as used in this subsection will be diluted.

D. Additional Figures and Tables

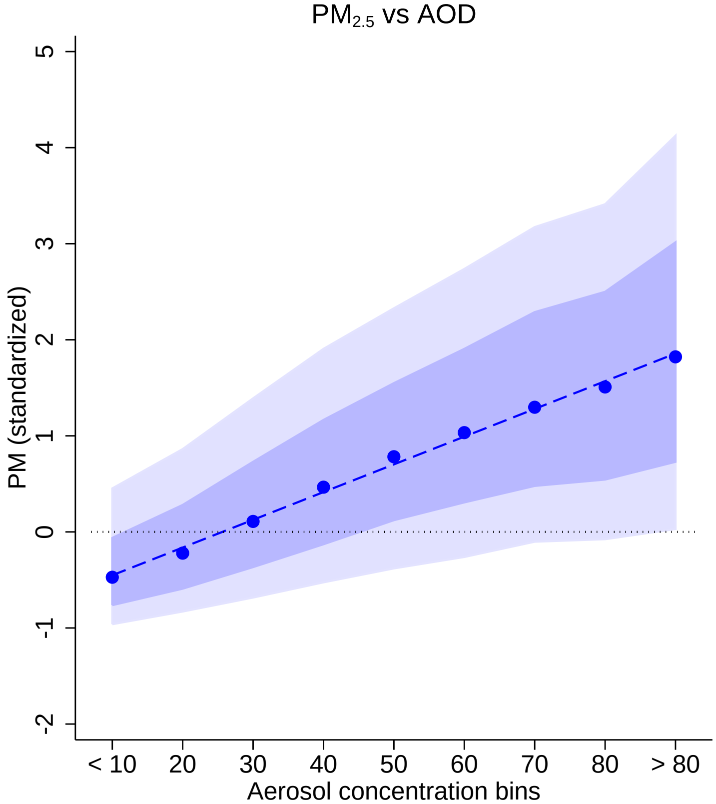
Figure D.1: 10km×10km Level Aerosol Concentration, 2001-2013 Average



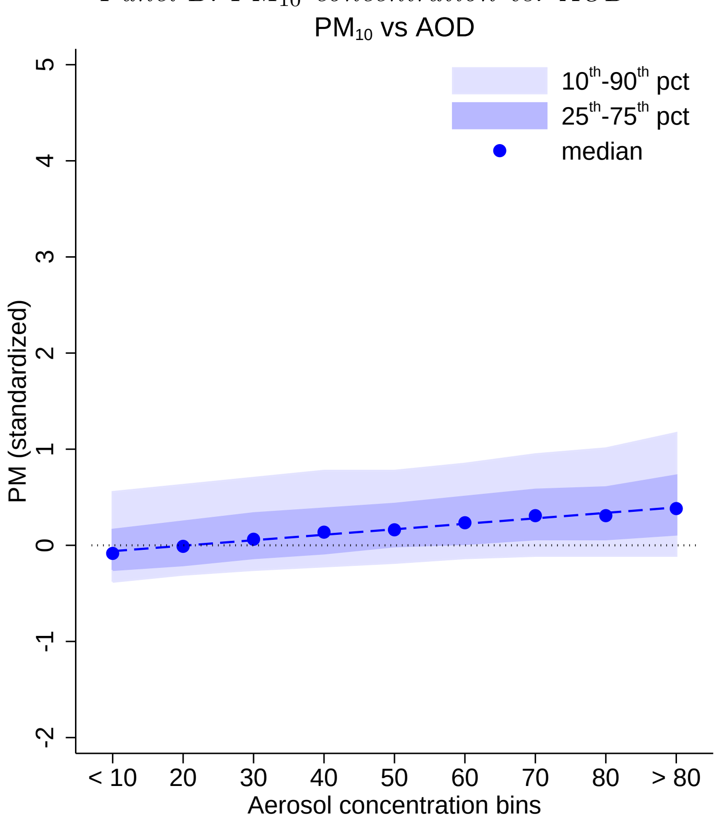
Notes: Map shows 13 year (2001-2013) average 10km×10km pixel level aerosol concentration. Legend presents ranges of deciles aerosol concentration. Number of pixels in each decile in parentheses.

Figure D.2: PM and Aerosol Correlation, 2001-2013

Panel A: $PM_{2.5}$ concentration vs. AOD

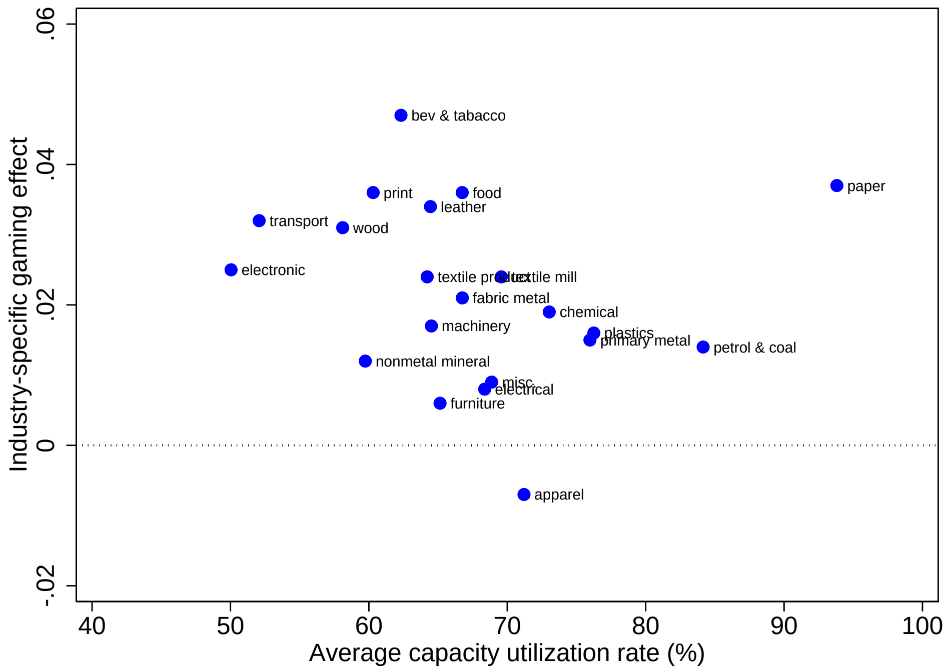


Panel B: PM_{10} concentration vs. AOD



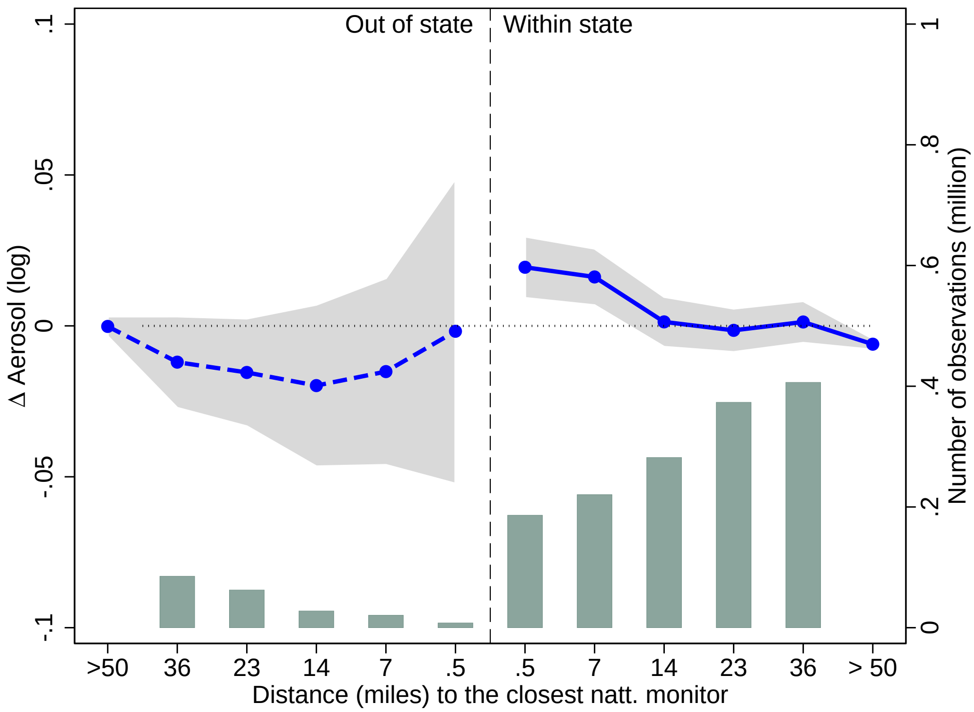
Notes: Graph presents the correlation between monitor PM readings and the satellite pollution measure, defined as the aerosol level within the 10km×10km area where the monitor lives in. The distribution of standardized $PM_{2.5}$ (panel A) and PM_{10} (panel B) concentrations within each aerosol level bin is displayed. The "<10" bin polls all observations where aerosol is less than 10, whereas the ">80" bin polls all observations with aerosol is greater than 80. The aerosol measure has a range of -5 to 500, whereas more than 95% of observations fall in the 0-100 range. Within each bin, the 10^{th}, 25^{th}, 50^{th}, 70^{th} and 90^{th} percentile PM values are shown.

Figure D.3: Correlation: 1/6day Pollution Gap vs. Industry Level Capacity Utilization Rate



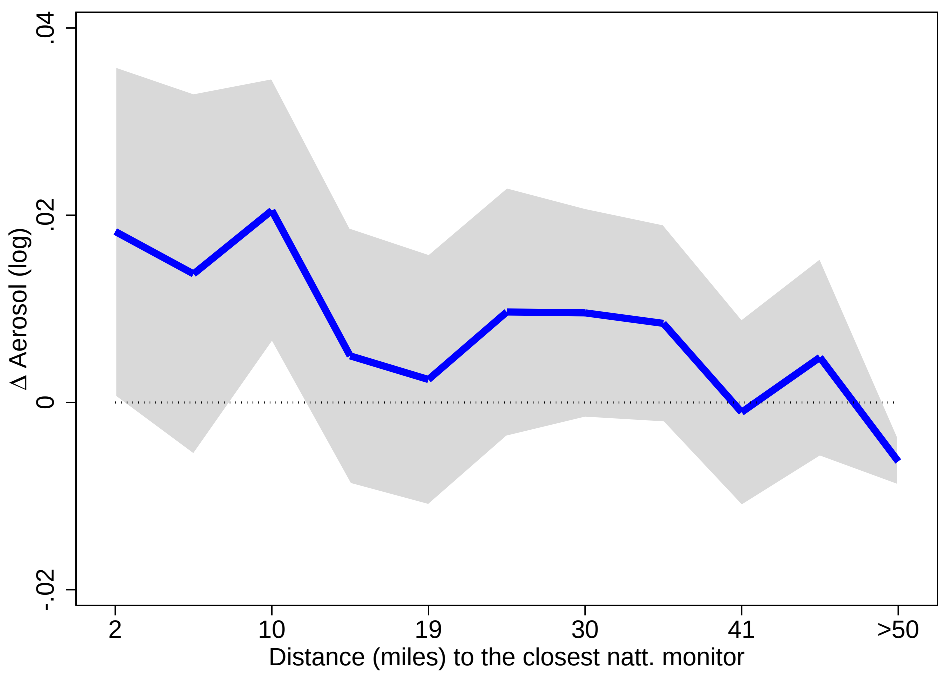
Notes: Each dot represents a 3-digit NAICS industry in the manufacturing sector. y-axis shows industry-specific 1/6day pollution gap from Table D.2. x-axis shows average 2008-2013 3-digit NAICS industry level capacity utilization rate. See the appendix text for more details.

Figure D.4: 1/6day Pollution Gap Near Industrial Sources



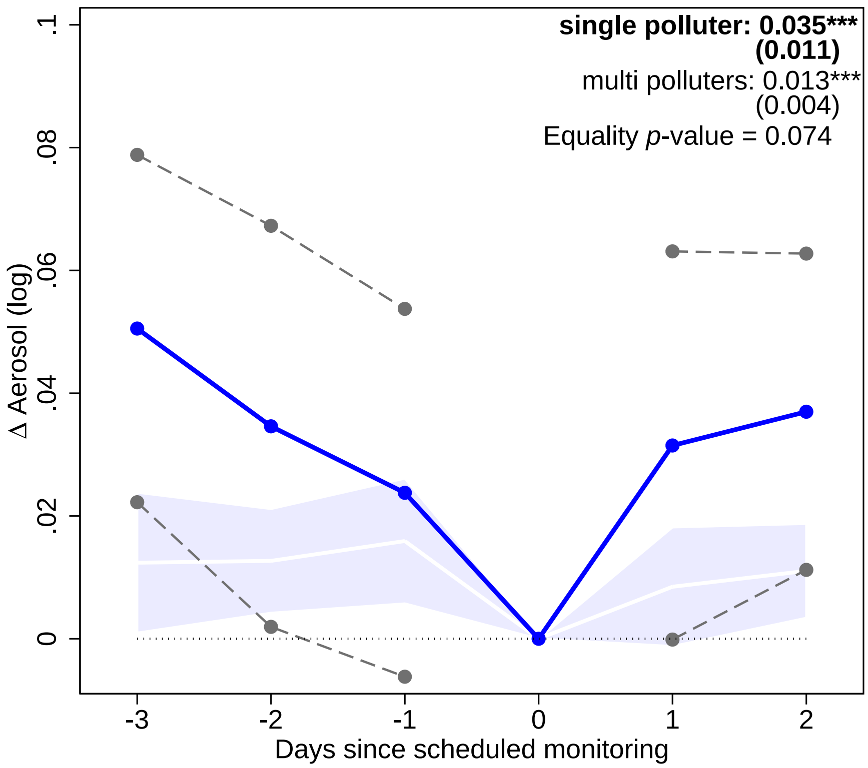
Notes: Graph shows the 1/6day pollution gap near industrial sources, stratified by facilities' distance to the closest 1/6day non-attainment monitor and whether the facility and the monitor locate on the same side of the state border. The farthest bin groups all facility-monitor pairs that are at least 50 miles apart. The regression includes detailed weather controls, distance bin fixed effects, site fixed effects, year fixed effects, month-of-year fixed effects, and day-of-week fixed effects. Gray shades present 95% confidence intervals constructed from standard errors clustered at the county level.

Figure D.5: Robustness: 1/6day Pollution Gap Near Industrial Sources



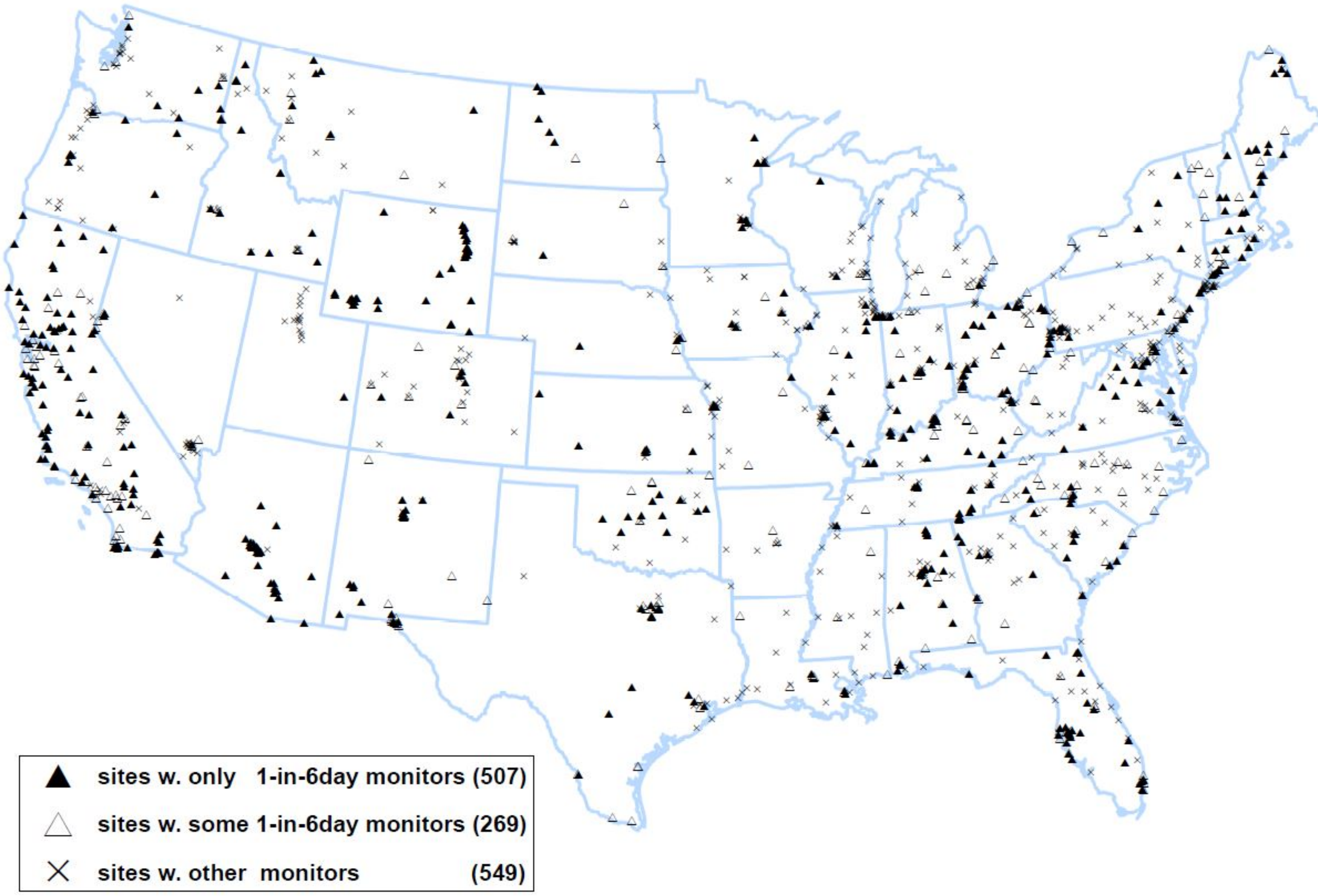
Notes: Graph shows replication of 1/6day pollution gap near industrial sources using the National Emissions Inventory (NEI) 2011 as the source of polluting facility information. The sample is restricted to polluters in 3-digit industry with total emission greater than 0.01 ton. This includes industries with the following 3-digit NAICS: 211, 212, 221, 311, 321, 322, 324, 325, 327, 331, 488, and 562. The underlying regression covers facilities that locate in 18,487 satellite 10km×10km grids, spanning 2,964 counties, and 22,633,006 total observations.

Figure D.6: 1/6day Pollution Gap by Local Emission Concentration



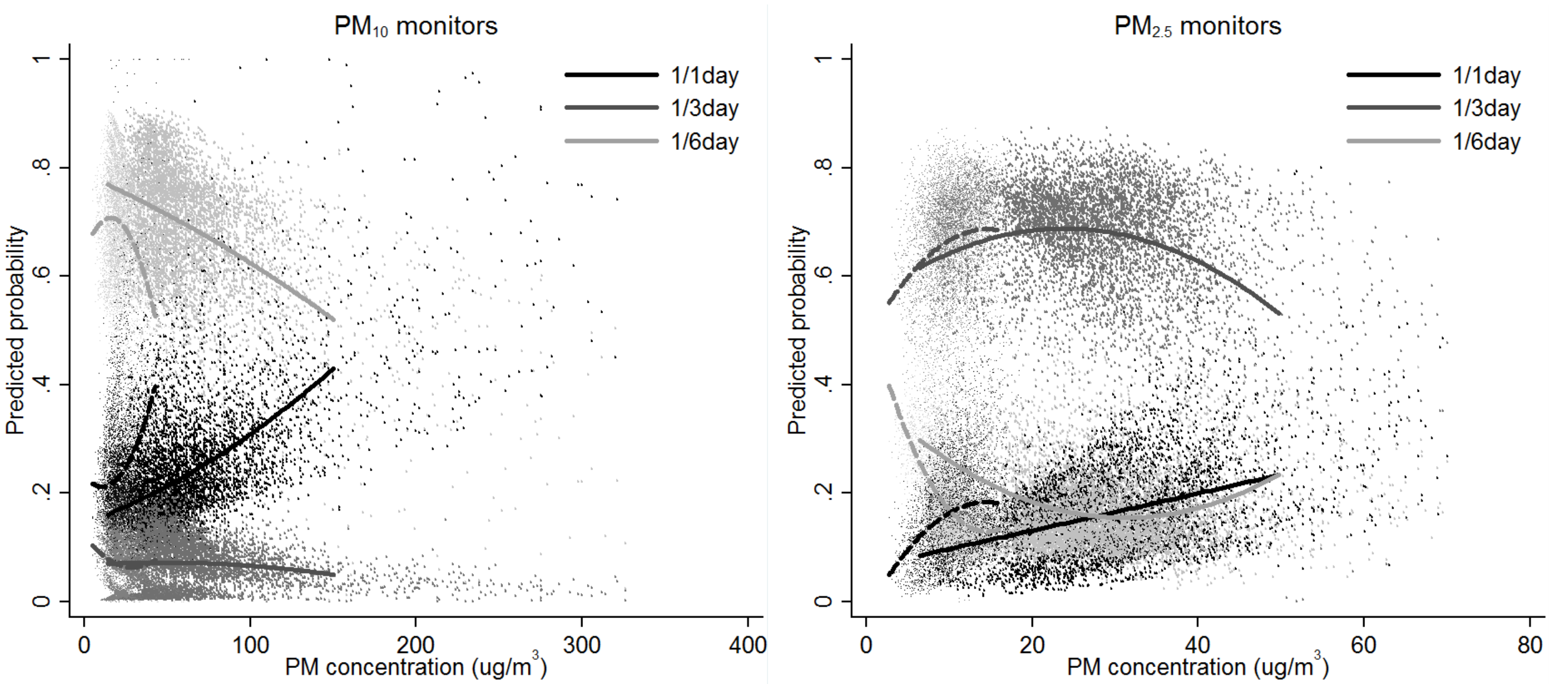
Notes: Graph displays 1/6day pollution cycles separately for counties with a single polluters vs. counties with multiple polluters. Results are from a single regression. Foreground graph objects represent estimates single polluter counties, while the background graph objects show estimates for the rest of the samples. Dashed lines and the shades represent 95% confidence interval constructed from standard errors clustered at the county level. Point estimates shown on the upper-right corner shows average 1/6day pollution gap (i.e. the difference between average day -3, -2, -1, 1, 2 pollution and day 0 pollution). Equality *p*-value corresponds to the null hypothesis that there is no difference in the 1/6day pollution gap for the two groups. Regression includes detailed weather controls, site fixed effects, year fixed effects, month-of-year fixed effects, and day-of-week fixed effects.

Figure D.7: Distribution of Monitors, 2001



Notes: Graph plots the 2001 snapshot of the spatial distribution of all PM monitoring sites in the lower 48 states. Solid triangles show sites with only 1/6day PM monitors. Hollow triangles show sites with some 1/6day monitors mixed with monitors following other schedules. Crosses show PM sites with no 1/6day monitors.

Figure D.8: Predicted Probability of Monitoring Schedule Assignment by Annual PM Concentration



Notes: Graph reports predicted probability of monitoring schedule assignment for PM_{10} (left panel) and $PM_{2.5}$ (right panel) by annual PM concentration. Predictions are obtained from a multinomial logistic model that predicts selection into monitoring schedule by annual average and 99^{th} percentile PM value fully interacted with Census region dummies, 5 year lags in annual average as well as 99^{th} percentile value, and calendar year dummies. Each dot on the graph represent a monitor-pollutant metric. Lines show quadratic fits of predicted probability over annual average concentration (dashed) and annual 99^{th} percentile concentration (solid).

Table D.1: Linear Regressions: Ground Particulate Matters Concentration and Satellite Aerosol

Independent variable:	Aerosol (lvl)		
	(1)	(2)	(3)
Panel A: Dependent variable = $PM_{2.5}$ (std.)			
Aerosol	0.028***	0.028***	0.027***
	(0.001)	(0.001)	(0.001)
N	502,410	502,402	351,284
N (site)	1,676	1,668	1,306
Panel B: Dependent variable = PM_{10} (std.)			
Aerosol	0.0067***	0.0065***	0.0060***
	(0.0006)	(0.0003)	(0.0004)
N	534,446	534,438	402,758
N (site)	1,745	1,737	1,236
FEs: site		✓	
FEs: day-of-year		✓	
FEs: year		✓	
FEs: site×day-of-year			✓
FEs: state×year			✓

Notes: Table shows linear regression coefficients where standardized daily PM concentration is regressed on the MODIS aerosol measure. Each column in a panel is a separate regression. The unit of anlaysis is a monitoring site-day. PM is defined as average concentration across all monitors within a site on a given day. Aerosol level is defined as the aerosol concentration of the 10km×10km area which contains the monitoring site. Aerosol level ranges from -5 to 500, whereas >95% observations fall within the 0-100 range. Panel A reports results for $PM_{2.5}$ and Panel B reports results for PM_{10}. Column 1 reports regression with no controls. Column 2 controls for site, day-of-year, and year fixed effects. Column 3 controls for site×day-of-year and state×year fixed effects. Standarad errors are clustered at the site level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Table D.2: 1/6day Pollution Gap by 3-Digit Industry

Dependent variable:	Aerosol (log)				
	(1)	(2)	(3)	(4) N	(5) N
	0-15 miles	15-50 miles	> 50 miles	county	
Mining					
Gas extraction (211)	0.196	0.179	-0.016	14	4,906
Coal mining (212)	0.050	0.037	0.042	200	171,102
Utility					
Utility (221)	0.022	0.004	-0.003	765	1,204,593
Manufacturing					
Beverage & tabacco (312)	0.047	0.015	0.005	106	106,456
Paper (322)	0.037	-0.003	-0.009	409	419,352
Food (311)	0.036	-0.005	-0.003	898	1,140,043
Printing (323)	0.036	0.003	0.002	232	202,538
Leather (316)	0.034	0.072	0.008	53	34,565
Transport. equip. (336)	0.032	0.01	-0.004	943	1,240,341
Wood (321)	0.031	-0.010	-0.008	574	610,995
Electronic product (334)	0.025	0.004	-0.002	544	812,197
Textile mills (313)	0.024	-0.041	-0.012	186	190,670
Textile (314)	0.024	-0.025	-0.011	68	64,973
Fabricated metal (332)	0.021	0.000	-0.003	1,116	1,977,566
Chemical (325)	0.019	-0.006	-0.006	1,187	2,189,099
Machinery (333)	0.017	0.001	-0.008	822	988,760
Plastics (326)	0.016	-0.006	-0.007	973	1,343,845
Primary metal (331)	0.015	-0.009	-0.001	834	1,318,243
Petroleum & coal products (324)	0.014	0.015	-0.005	472	517,930
Nonmetallic mineral (327)	0.012	0.002	-0.007	1,014	1,369,429
Miscellaneous (339)	0.009	-0.002	-0.001	426	387,245
Electrical equip. (335)	0.008	0.001	-0.006	561	636,763
Furniture (337)	0.006	-0.007	0.004	327	274,422
Apparel (315)	-0.007	-0.045	-0.003	18	8,722
Distribution					
Electronic markets wholesale (425)	0.120	0.042	0.079	10	4,280
Chemical & petroleum wholesale (424)	0.025	-0.001	-0.001	523	606,200
Hazardous waste treatment					
Hazardous Waste (562)	0.030	0.001	-0.001	239	232,983
Federal facilities					
Non-military	0.053	0.009	0.007	245	155,193
Military (928)	-0.020	0.008	-0.008	251	234,470

Notes: Table reports 1/6day pollution gap estimated separately by 3-digit NAICS industry. Sample includes all polluting facility-years included in the EPA's Toxic Release Inventory database from 2001 to 2013. Each row reports a separate regression. Row names report industry names with the 3-digit NAICS code shown in parentheses. Each facility-year is linked to the closest non-compliance monitor, and 1/6day pollution gap is estimated when the facility-monitor pair is $<$ 15 miles (column 1), 15-50 miles (column 2), and $>$ 50 miles (column) apart. Column 4 shows the number of counties spanned, and column 5 reports total number of observations in each regression. Shaded cells highlight regression coefficients that are individually significant at the 1% level. Standard errors are clustered at the county level.

Table D.3: The Effect of 1/6day Monitoring Schedule on Control Burning and Forest Fire

	Sample: All sites		Sample: Natt. sites	
	(1)	(2)	(3)	(4)
	In-county fire	In-state fire	In-county fire	In-state fire
Dependent variable = Indicator for controlled burning events (coeff.×100)				
off days	0.002	-0.179	0.008	0.161
	(0.017)	(0.164)	(0.073)	(0.192)
Power$_{\text{(1.5\% effect, 5\% sig.)}}$	0.117	0.325	0.052	0.108
Mean dep. var. (×100)	0.750	17.3	0.583	16.8
N	2,404,615	2,404,615	79,127	208,478
N (site)	1,193	1,193	94	182
Dependent variable = Indicator for forest fire events (coeff.×100)				
off days	0.054	0.052	0.001	-0.191
	(0.067)	(0.081)	(0.142)	(0.119)
Power$_{\text{(1.5\% effect, 5\% sig.)}}$	0.623	0.999	0.150	0.700
Mean dep. var. (×100)	6.76	40.6	11.7	49.4
N	2,304,303	2,304,303	201,882	201,882
N (site)	1,175	1,175	182	182

Notes: Each column in a panel represents a separate regression. Panel A uses control burnings as outcome measures, and panel B uses all wildfire events as outcome measures. Outcome variables are dummy for within county fire (column 1 and 3) and within state fire (column 2 and 4). See the appendix text for more details. All regressions include detailed weather controls, site fixed effects, year fixed effects, month-of-year fixed effects, and day-of-week fixed effects. Power calculation estimates power of tests detecting a 1.5% mean difference between off-days vs. on-days at a 5% significance level. Standard errors are clustered at the county level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.