For Online Publication

Appendix to:

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

Eric Yongchen Zou

University of Oregon

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. History of Particulates Pollution Sampling, 1950s – 1970s

The practice of intermittent pollution monitoring dates back to the 1950s when the U.S. first started collecting data for particulates pollution. The National Air Sampling Network was established in 1955 under the Federal Air Pollution Research and Technical Assistance Act. The Act establishes a purely informational role of the government to conduct research on air pollution matter. No air quality standards are set and regulatory provisions are included to punish polluters. Initially, air sampling stations are set in 17 communities. The network expands to 83 communities by the end of 1956. During this period, sampling is done on a weekly basis, where the sampling day is chosen by the state operator at their convenience.

Two problems with the Network were (1) the geographic coverage was poor, and (2) since monitoring days are chosen by convenience, very few samples were obtained on weekends, biasing the resulting average concentration. A major revision is put into effect in 1957. First, the Network expands to a "national" scale that contains more than 150 stations distributed in each state. Second, it starts to implement a "modified random" sampling scheme. In this scheme, pollution samples are taken biweekly, and within each 2-week window the sampling day is randomly chosen at the beginning of the year. In a 1957 study, the U.S. Public Health Service concluded that "sufficiently reliable measures of air pollution in a specific area may be obtained by sampling on a limited basis and the additional accuracy to be gained by daily sampling is not sufficient to justify the increased operation costs." (U.S. Public Health Service, 1957[1])

At the time, the low monitoring frequency was warranted by the fact that the purpose of TSP sampling was purely informational. This changed in the 1970s when the Clean Air Act established *regulatory standards* for TSP. Because compliance status was based on annual

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

statistics of TSP, data from biweekly sampling was no longer sufficient. Therefore, a new monitoring practice must be developed which has a higher data capture rate while brings minimum burden to operation agencies as possible. Such a method was proposed by Akland (1972)[2], who studied a cyclical sampling schedule which samples TSP once every k^{th} day. In the paper, Akland used six years of daily TSP data from a monitoring site in Buffalo, and showed that mean and precision statistics computed from subsamples using a 1-in-3 day schedule or a 1-in-13 day schedule are not significantly different from those computed from the entire sample. The Clean Air Act subsequently adopted a 1-in-6 day monitoring schedule (for the vast majority of TSP sites) in the TSP regulation and the practice continued in PM regulation starting from 1980s.

A.2. Modern 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* (U.S. 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.

Atmospheric PM is measured by the amount of particle deposition when a PM monitor forces air through a size-discrimination filter which is typically made of glass fibers (for PM10 measurement) and Teflon (for PM2.5 measurement). PM concentration is then computed by the ratio of amount of particle deposition (in ug) to the volume of air that carried the particles (in m^3). Due to the need for measuring air flow, modern PM monitors are also under microprocessor control that uses real time temperature and barometric pressure readings to determine flow rate. Regular maintenance effort is needed to ensure that monitor measurements function properly. Sampler dust, especially build-up in the air inlet, must be cleaned roughly after 15 days of monitoring. Calibration needs to be verified every 90 days to ensure the accuracy of air flow measurement.

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,

---

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

nonreactive containers, and sealed. Samples are then delivered to the laboratory, usually on the same day that the samples are taken. In the lab, filters must be “equilibrated” in a controlled environmental for 24 hours (20 degree C and 40% humidity) before weighing analysis which must be done within 10 days since the sample collection.

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.3. 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-in-6-day) schedule studied in this paper, two other frequently used schedules are the 1/3day and the 1/1day (i.e. daily) schedules. See Appendix Figure D.3 for a snapshot of monitor frequency distribution in year 2001 and 2013. By the EPA’s rule, 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. No reward is given to over-sampling: the EPA only accepts an applicable number of samples with the highest pollution readings in cases where more samples than required are taken. Appendix Table D.6 shows that the assigned monitoring frequency is closely followed by state governments. 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.

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>. Below I provide more details about the administration of sampling frequency assignment and revisions. I provide separate discussion for PM2.5 sites and PM10 sites as the rules for frequency assignment and revisions are slightly different.

PM2.5 sampling frequency. In principle, all PM2.5 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 PM2.5 data quality assessments, the location of current PM2.5 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 PM2.5 sampler may also follow the 1-in-6 day 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-in-6 day 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* $PM_{2.5}$ *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-in-6 day samplers (i.e. sites where the 1-in-6 day sampler must not be a collocator) the gaming effect is stronger.

PM10 sampling frequency. The 1997 revision of PM NAAQS sets sampling frequency to a minimum of once in three days for all PM2.5 and PM10 sites. But for PM10 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 PM10 standard will be exceeded" (U.S. EPA, 1997b). 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".

Characteristics of sites with different sampling frequencies. In general, more frequent schedule is assigned to sites with higher chances of violating the NAAQS. I provide some industrial and government characteristics by monitoring schedule in Appendix Table D.2. Because temporal data availability varies, for each monitoring group I report cross-sectional statistics. Thus, each cell in the table shows the county-level characteristics for a monitor that follows a specific monitoring schedule, averaged across the entire study period. Starting the first row, I find that the odds of a monitor ever violating a PM NAAQS standard from 2001-2013 is about 50 percent at a 1-in-1-day site, much higher than at sites that follow intermittent monitoring (about 30 percent). Daily traffic volume seems to be higher at 1-in-1-day locations, suggesting every-day monitoring is more likely to be employed in urban centers. Continuing the rows, I find little evidence that county-level composition of polluting industries (including manufacturing, utility, mining, and construction) differ significantly across schedules. For instance, the county-level fraction of employment in the manufacturing sector is roughly about 10 percent at 1-in-1-day, 1-in-3-day, and 1-in-6-day sites. There is also little evidence that government characteristics differ. For example, counties with different monitoring schedules are similar in government sector size, level of environmental friendliness (measured by state-level index of the League of Conservation Scorecard), as well as blue/red party affiliation shares (at the state-level). Overall, I find that observed pollution levels as well as past NAAQS violation histories appear to be the driving forces underlying monitoring frequency assignment.

A.4. 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 PM2.5 monitoring, the focus of most of the innovations.

Manual sampling of PM2.5 acquires deposits over a 24-hour period on a size-discrimination filter from air drawn at a controlled flow rate through the PM2.5 inlet. If done appropriately, manual sampling obtains the most accurate measure of ambient PM2.5 concentrations. In the EPA's language, this method provides the “reference” measure of PM2.5 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.*"

Although some continuous technologies can provide reasonable proxies of PM2.5 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 PM2.5 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. Separately, 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.

Appendix Figure D.4 plots the number of PM2.5 monitors by sampling frequency and method since 2001. There is a big rise in the information use of continuous PM2.5 monitors starting 2004, the year AirNow launched. Continuous monitors were not used for NAAQS regulatory purposes until 2009. Over the entire time period, 23% of continuous monitors are used for such regulation. While regulation use of continuous monitors has risen in recent years, these continuous monitors are only slowly replacing manual monitors. Extrapolating the roughly linear trends in the figure above, the entire PM2.5 monitoring system will not become continuous until 2035. Trends also suggest that continuous monitors are mostly replacing 1-in-3-day and 1-in-1-day manual monitors; the replacement of 1-in-6-day monitors is occurring more slowly. The differential resistance of upgrading 1-in-6-day monitors to continuous might be an interesting fact in and of itself. But overall, the trends suggest that intermittent pollution monitoring is not coming to an end in the near term.

B. Additional Data Descriptions

B.1. Satellite Data

The satellite data is sourced from the NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) algorithm installed on *Terra*, a sun-synchronous satellite that crosses the equator on descending passes at approximately 10:30 AM local time. MODIS contains a total of 36 spectral bands (ranging from 0.4 to 14.2 microns) that passively measures reflected solar and thermal radiation emission, and produces data for a wealth of atmospheric elements such as surface temperature, moisture, water vapor, etc. This paper uses MODIS' aerosol variable which measures aerosol optical depth (AOD) at 0.55 microns wavelength. Importantly, while the AOD retrieval method hasn't fundamentally changed since its invention in late 1990s, the algorithm gets refined overtime. These updates are reflected in new "collections" of data, i.e. releases of re-processing of raw imageries that incorporate algorithm refinements. This paper uses MODIS' collection 6 AOD data, which is the most updated version at the time of writing.[3] Notably, MODIS collection 6 features the availability of an enhanced retrieval algorithm called "Deep Blue" that substantially improves the ability to measure AOD over bright surface (e.g., snow and deserts). I use the "AOD_550_Dark_Target_Deep_Blue_Combined" AOD measure that incorporates this new algorithm through a merge between the "Deep Blue" and the original "Dark Target" measure which only retrieves AOD over dark surface.

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 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 raster files into 0.1km×0.1km pixel sizes, and then map them onto a 10km×10km gridded map of the contiguous U.S. 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

3 Sayer, A. M., L. A. Munchak, N. C. Hsu, R. C. Levy, C. Bettenhausen, and M-J. Jeong. "MODIS Collection 6 aerosol products: Comparison between Aqua's e-Deep Blue, Dark Target, and "merged" data sets, and usage recommendations." *Journal of Geophysical Research: Atmospheres* 119, no. 24 (2014).

that each grid tracks aerosol levels for the same area over time. Appendix Figure D.5 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 (Liu, Franklin, Kahn, and Koutrakis, 2007; Lee, Coull, Bell, and Koutrakis, 2012; Zhang and Lee, 2015). As a replication of this relationship in my study context, I correlate monitor-daily level PM2.5 concentrations to the daily aerosol level within the 10km×10km grid where the monitor falls in. Appendix Figure D.6 plots a simple bin-scatter of PM2.5 concentrations by the corresponding aerosol level.

B.2. Industry Data

County Business Pattern. Employment data are drawn from the Census Bureau's annual County Business Pattern (CBP) data which contains information on industry × county level employment counts. For each 3-digit NAICS industry available in the CBP, I compute its employment concentration for every county, defined as the share of employment in the 3-digit industry relative to the county's total employment. Importantly, in the CBP data about 50 percent of employment counts at the 3-digit industry × county level are masked to avoid disclosure. In cells where employment counts are masked, I use the CBPs' establishment count by employee size class information to impute employment count. Specifically, a county's employment count for a masked industry j is imputed as

$$Employment_j = \sum_s Establishment_{js} \times \left(\frac{MaxEmployeeSize_s + MinEmployeeSize_s}{2}\right)$$

where $Establishment_{js}$ is number of industry j 's establishments in employee size class s, and $Max(Min)EmployeeSize_s$ is the upper (lower) end of the employee count range of employee size class s.

Toxic Release Inventory. 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.

In Section C.4 below, I used the the TRI's self-reported emission amount to construct a county-level Herfindahl-style index (HHI) of emission concentration. Specifically, for each county c and year y, I define its emission HHI to be

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

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.

B.3. Weather Data

Temperature and Precipitation. Temperature and precipitation measures are derived from the National Climatic Data Center's Global Historical Climatology Network (GHCN) which contains daily weather information from about 9,000 weather stations across the U.S. I compute county × daily level temperature and precipitation using 20 mile inverse distance weighting: a county × day's temperature (precipitation) is computed as the weighted average of readings from all monitors within 20 miles to the county's centroid, where the reading from a given monitor is assigned a weight equal to the inverse of the distance between the monitor and the county's centroid.

Wind. Wind direction and speed data are drawn from the North American Regional Reanalysis (NARR, 2001-2013) produced by the National Centers for Environmental Prediction. NARR contains wind conditions information at a spatial resolution of 32km × 32 km grid cell. For each grid cell × day, NARR reports the horizontal (u-wind) and vertical (v-wind) components of the wind vector. From 2001 to 2013, I compute daily wind vector at each county by geographically interpolating the u- and the v- components between nearby grid cells within 20 miles to the county's centroid. I then convert the components to average wind speed and direction using trigonometry.

C. Additional Analysis

C.1. 1-in-6-Day Monitor Retirement

Section IIIA of the paper exploits monitor retirement events to examine changes in the level of pollution on on-days and off-days. Here I provide additional details of the analysis.

Synthetic weights. To improve the comparability of non-retiring sites to retiring sites in terms of pre-treatment pollution levels, I use the synthetic control method to assign weights to each control site in such a way that the weighted average control site will do a better job tracking trends in retiring sites' pre-treatment pollution levels in an MSE-minimizing sense. I compute synthetic weights based on data in periods t=-4 and t=-3 only, allowing me to use t=-2 and t=-1 as a validation sample of the synthetic control method's performance in my context. That is, if synthetic weighting does a good job capturing overall trends, we expect to a mechanically-zero difference between treatment and control pollution levels in period t=-4 and t=-3, but also a zero difference at t= -2 and -1. Any post-treatment difference in pollution for the treated and control units is then attributable to the effect of the treatment.

Difference-in-Differences Results. The graphical evidence in Figure 4 motivates a difference-in-differences design that compares retiring sites and (synthetically weighted) non-retiring sites, before and after retirement. Appendix Figure D.7 summarizes the findings graphically. The chart plots difference in pollution levels in treated site and (synthetic) control sites as a function of years relative to monitoring site retirement. To test for differential response across on- and off-days, the regressions are run separately for monitoring days (panel A) and nonmonitoring days (panel B). Two major patterns emerge. First, in both panels, the difference between treated and control units is close to zero before treatment. Note that, as explained earlier, this finding is mechanical for t=-3 (because data in that event year are used to generate synthetic weights), but is not mechanical for t=-2 or t=-1. This pattern suggests that synthetic weights perform well in capturing factors that cause treatment and controls to be on a differential pre-trend in the first place. Second, Appendix Figure D.7 features a rise in pollution on monitored days after treatment. The individual event-year coefficients are not statistically significant, but the overall increase is judged as visually obvious. For a non-retiring site, there is no evidence of changing pollution levels before and after retirement of other sites. This pattern echoes the observation from raw data: monitoring retirement appears to only increase on-day pollution levels, without changing pollution levels on off-days.

C.2. Strategic Action Day Warnings and Political Characteristics

Section IIIC documents evidence on strategic air quality Action Day warnings. To explore potential mechanisms that could explain why some areas are issuing strategic warnings, below I report an additional analysis that sheds some light on political characteristics associated with areas that issue strategic Action Day advisories.

I first document substantial heterogeneity in strategic advisories (henceforth, the "advisory gap") across states. Appendix Figure D.8 reports a specification in which I allow the off-days/on-day advisory gap to vary by each state in my sample. Several patterns emerge. First, not every state shows evidence of strategic advisories. I detect an individually significant strategic advisory effect in six states (NJ, OH, TN, NY, ME, PA). Second, there is meaningful variation in effect sizes across states, ranging from about 1.5 percentage points (out of a mean daily advisory rate of 1 percent) in some states to effectively zero in others. Third, almost no states exhibit "positive" advisory gaps (i.e., more advisories on unmonitored days). This is reassuring because local agencies should have no incentive to issue more advisories on non-monitored days.

The presence of substantial heterogeneity across states provides an opportunity to investigate political characteristics underlying strategic advisories. For example, do strategic advisories often arise in states perceived less environmentally friendly? To shed light on some of the potential mechanisms, I obtain additional data to build five state-level characteristics to measure potential political influences:

(a) Government size: government-sector (2-digit NAICS=92) employment as a share of total employment. Sector-specific employment data are sourced from the Bureau of Economic Analysis.

(b) Party affiliation: share of Democratic Party affiliation as of year 2006 (the middle point of my study period) using data from Gallup.

(c) "Pro-environment" score: League of Conservation Voters (LCV) score, which is based on state representatives' voting records on environmental issues. A higher score indicates to a stronger environmental preference (e.g., Dietz et al., 2015[4]).

(d) Corruption: per capita number of federal convictions among state and local public officials. These data are 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 United States (Glaeser and Saks, 2006[5]; Leeson and Sobel, 2008[6]; Grooms, 2015[7]).

(e) History of challenging federal NAAQS nonattainment-status designation: This is a novel measure I propose to capture a state's administrative capacity to respond to the EPA's NAAQS compliance regulations. Under the Clean Air Act, when NAAQS compliance is evaluated, each state has an opportunity to recommend designations before the EPA makes a preliminary designation decision, and an opportunity to challenge the preliminary designation before the final designation is made. Upon receiving the preliminary designation, the state is allowed two months to challenge, usually by presenting new data analyses and arguments that support the original recommendation. These communications are publicized on the EPA's website. In the lower 48 states, 31 have submitted challenges in the past. Using this information, I create a measure that

---

[4] Dietz, Thomas, Kenneth A. Frank, Cameron T. Whitley, Jennifer Kelly, and Rachel Kelly. "Political influences on greenhouse gas emissions from US states." Proceedings of the National Academy of Sciences 112, no. 27 (2015): 8254-8259.
[5] Glaeser, Edward L., and Raven E. Saks. "Corruption in america." Journal of Public Economics 90, no. 6-7 (2006): 1053-1072.
[6] Leeson, Peter T., and Russell S. Sobel. "Weathering corruption." The Journal of Law and Economics 51, no. 4 (2008): 667-681.
[7] Grooms, Katherine K. "Enforcing the Clean Water Act: The effect of state-level corruption on compliance." Journal of Environmental Economics and Management 73 (2015): 50-78.

potentially reveals states' attentiveness and the political resources available for NAAQS compliance. I count the total number of pages each state has put together in challenging the EPA's designation (zero if a state has not ever challenged). This measure has a wide spread; among the 31 states that ever filed a challenge (mean = 62.4, SD = 69.8), ranging from Colorado (a single page submission) to Georgia (294 pages submitted).

Appendix Table D.7 reports results from a series of regressions in which the magnitude of the advisory gap is allowed to vary by whether the state is above or below the median value of the government characteristics (indicated by the column names). The interactive coefficient "1(off-days) x 1(> median states)" therefore shows the additional degree of the advisory gap in above-median states relative to below-median states. That is, a negative coefficient means the above-median state issues more strategic advisories than below-median states. Results in Appendix Table D.7 suggest that places that issue strategic advisories not only have the bureaucratic incentive to game, but they also have the "administrative capacity" to do so. Column 4 shows that strategic advisories are more common in states that exhibit a history of challenging the EPA's decisions with high effort. Somewhat in line with this finding, column 5 shows that the effects concentrate in states with an above-median corruption index, suggesting that increased degree of "political flexibility" may also be needed for strategic advisories to be actually carried out. On the other hand, I find that the advisory gap does not depend on the size of the government (column 1) or on the degree of general environmental friendliness (columns 2-3).

C.3. Empirical Bayes Estimates of County-Level Pollution Gaps

Section IIID of the paper presents county-level estimates of the pollution gap. Here I provide additional details of the estimation.

Construction of Empirical Bayes estimates. To overcome noisy estimates at the county level, I construct Empirical Bayes (EB) estimates of county-level 1-in-6-day pollution gap. I use an approach adapted from previous research that uses big data to provide geographically localized causal estimates (e.g., Chetty and Hendren, 2018; Finkelstein, Gentzkow, and Williams, 2019).[8,9] The logic of the Empirical Bayes approach is to strike an optimal balance between the direct estimate of the local effect (which is unbiased but often contains substantial noise) and the estimate of the population effect (which often contains bias but little noise due to large number of observations). In my setting, the EB estimates are a combination of the raw pollution gap estimates, which are unbiased estimates of the pollution gap at the county level, and an estimate of the "regionwide" pollution gap, which is a biased estimate of the pollution gap in the interested county, but has low variance. To estimate the regionwide pollution gap for a given county, I use data from all counties within 150 miles of the interested county's geographic centroid. The 150-mile radius is used to represent the size of an average state in the United States. I do not use actual state boundaries to avoid the substantial variation in the size of states.

Following Chetty and Hendren (2018), the EB estimate mimics an MSE-minimizing linear predictor that would come out of a "hypothetical" OLS regression in which the true county-level effect is regressed on the raw county-level effect estimate and the regionwide estimate. In practice, the EB estimator is given by the following equation:

$$\beta_c^{EB} = \frac{\chi^2}{\chi^2 + s_c^2}\hat{\beta}_c + \frac{s_c^2}{\chi^2 + s_c^2} \cdot \hat{\gamma} \cdot (\tau_c - \bar{\tau})$$

where $\hat{\beta}_c$ (τ_c) is the raw pollution gap estimate for county c. $\hat{\gamma}$ is the coefficient estimate obtained from a univariate OLS regression of $\hat{\beta}_c$ on τ_c. s_c^2 are the sampling variance of the $\hat{\beta}_c$ estimates, and, in practice, they are estimated by the squared standard error of $\hat{\beta}_c$. χ^2 is the component of the variability of the true pollution gap across counties that is not explained by general variability across regions. In practice, χ^2 is estimated by the residual variance of a regression of $\hat{\beta}_c$ on τ_c, minus the average sampling variance of $\hat{\beta}_c$:

$$\chi^2 = \mathrm{var}\left(\hat{\beta}_c - \hat{\gamma} \cdot (\tau_c - \bar{\tau})\right) - E(s_c^2)$$

EB is therefore given by a linear combination of the (unbiased but noisy) raw pollution gap estimate $\hat{\beta}_c$ and the (biased but precise) regionwide pollution gap τ_c with relative weights

[8] Chetty, Raj, and Nathaniel Hendren. "The impacts of neighborhoods on intergenerational mobility II: County-level estimates." *The Quarterly Journal of Economics* 133, no. 3 (2018): 1163-1228.
[9] Finkelstein, Amy, Matthew Gentzkow, and Heidi L. Williams. Place-based drivers of mortality: Evidence from migration. No. w25975. National Bureau of Economic Research, 2019.

proportional to the signal-to-noise ratio in the $\hat{\beta}_c$ estimates. To put things in perspective, the average county-level regression underlying the estimation of $\hat{\beta}_c$ contains 35,236 observations, with a mean standard error of 0.0187 (in log scale); by contrast, the average region-level regression underlying the estimation of τ_c contains 2,224,094 observations with a mean standard error of 0.0022 (in log scale).

Figure 7 of the paper plots a map of county-level pollution gap estimates. In Appendix Figure D.9, I overlay the histogram of the raw county-level pollution gap estimates with the EB estimates. Relative to the raw estimates, the EB estimates show visually apparent shrinkage towards zero. It is worth noting that shrinkages mostly take effect on very large (negative or positive) observations – in part because many raw county-level regressions have quite sufficient numbers of observations to begin with.

Positive vs. negative pollution gaps. The analysis above shows that some large positive and large negative pollution gaps arise from sampling variation, and that the EB approach partially addresses the concern. Here I illustrate that a negative pollution gap can also be explained by wind transport (in addition to sampling variation). I show this in two ways.

First, in Appendix Figure D.9, I use color shades to indicate pollution gaps by distance to the nearest hot-spot county. Hot-spot counties (those with top-decile pollution gaps) are colored back; colors are lighter for counties farther away from the nearest hot-spot county. The color pattern of the histogram reveals that counties with a smaller pollution gap – that is, those to the left-hand side of the histogram – tend to be far away from hot-spot counties. In fact, the histogram features a gradual change in the gray scale, indicating a systematic relationship between a county's pollution gap and a county's distance to hot-spot counties. This pattern is consistent with wind transport of pollution gaps originating from hot-spot counties.

Second, to further explore the pattern observed in the previous exercise, I estimate the 1-in-6-day pollution profiles by distance to hot-spot counties, and by relative wind direction to the hot-spot counties. Appendix Figure D.10 summarizes the results. Panel A shows the "shape" of the 1-in-6-day pollution gap, estimated separately for counties in each decile of distance to hot-spot counties indicated by the y-axis. Thus, the "distance = 0 mile" group represents the hot-spot counties themselves; the shape represents average aerosol levels by six day of monitoring cycle, indicated by the x-axis. The chart suggests that: (1) the drop in pollution on the monitoring day becomes less salient as one moves away from the hot-spot county; (2) there appears to be a systematic shift of the timing of the pollution drop. For example, at about 170 miles from the hot-spot counties, the day with the lowest pollution level occurs on day 1 rather than on day 0. At about 300 miles, the day with the lowest pollution level occurs on day 3 (or equivalently, day -3 of the next six-day cycle). Perhaps as anecdotal evidence, the speed of such transport is consistent with the average wind speed of about 3 miles per hour in the United States; and (3) the negative pollution gap arises partly because the regression estimates always use the monitoring day (day 0, represented by the vertical dashed lines in the chart) as the reference day. Finally, Panel B of Appendix Figure D.10 repeats the same analysis (as in Panel A) separately for downwind counties (those with an absolute wind-bearing of less than 30 arc-degrees relative to hot-spot counties) and upwind counties (those with an absolute wind bearing of over 150 degrees). Results show that the shift is stronger for counties in downwind directions.

C.4. Pollution Gap and County's Emission Herfindahl Index

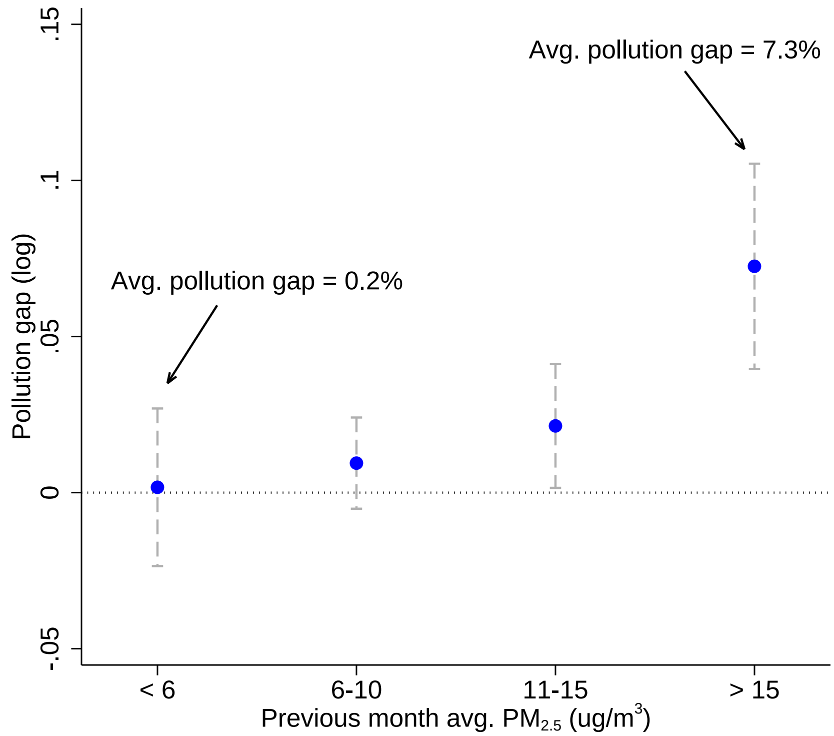
While cooperative polluting activities are unlikely to drive the pollution gap, they might be observed in counties with single/major polluters. I test this economic prediction by examining heterogeneous pollution gaps by county's emission Herfindahl-Hirschman index (HHI) constructed from the EPA's Toxic Release Inventory (TRI) data which contain annual observations of plants' reported total air emissions. The HHI ranges from 0 to 1 and takes larger values in counties where fewer polluters contribute to total emissions. Appendix Figure D.11 reports heterogeneous 1-in-6-day pollution patterns by high ($\geqslant$ 0.9) vs. low (< 0.9) HHI. In unreported results, I show this finding is robust to various cutoffs of HHI. Results show significantly stronger gaming in areas with high levels of emission concentration, where the pollution gap averages 3.1 percent. However, I cannot reject the existence of a significant pollution of about 1.2 percent in low HHI regions as well.

C.5. Pollution Gap and Monitor's Distance to Highway

As complimentary evidence to the warning analysis, I augment equation (2) of the paper by allowing the pollution gap estimate to vary flexibly by the monitor's distance to nearest highway segment. I focus on monitors located within 3 miles of the highway (< 0.5 miles, 0.5-1 miles, … , 2.5-3 miles) and all other monitors are pooled into a single group (> 3 miles). Appendix Figure D.12 suggests a distance gradient. I find that monitors within 1 mile of highways exhibit a strong response (roughly 2.5 percent pollution gap), while the pollution gap is not detectable for monitors that are between 1 mile and 3 miles away from highways. The results also suggest that the pollution gap observed near highways is likely to only partially explain the main finding on the average pollution gap, as a significant pollution gap is also precisely estimated for all monitors that fall more than 3 miles away from highways.

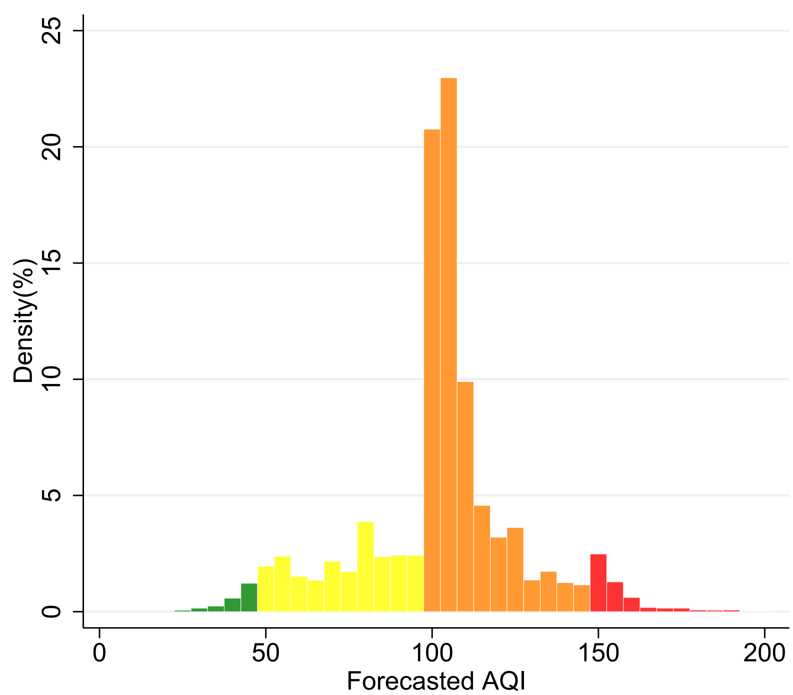
D. Appendix Figures and Tables

Figure D.1: Heterogeneous Pollution Gap by County's Previous Month PM2.5 Level



Notes: This figures reports an extension of Figure 5's regression, where the previous month's PM2.5 enters the estimation model as separate level bins, rather than a single linear term. The underlying regressions include main effect terms, site, year, month-of-year, and day-of-week fixed effects, and weather controls (Section IIB). Range bars represent 95% confidence intervals constructed using standard errors clustered at the county level.

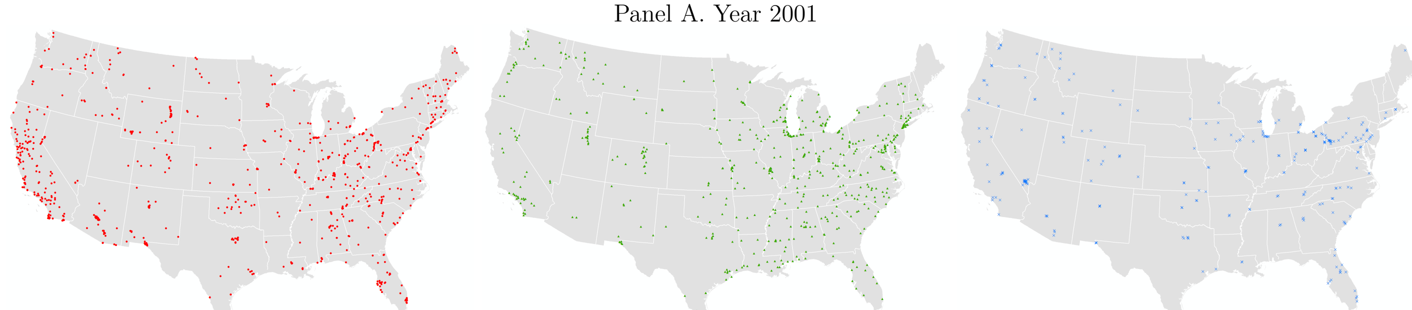
Figure D.2: Distribution of Forecasted AQI on Action Days



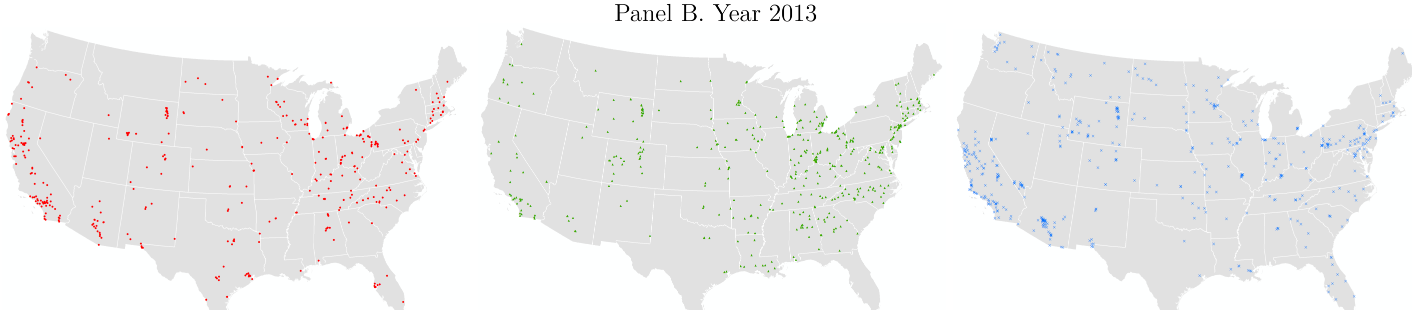
Notes: Distribution of forecasted AQI on days with Action Day advisory.

Figure D.3: Location of 1-in-6-Day (Left), 1-in-3-Day (Middle), 1-in-1-Day (Right) Monitors

Panel A. Year 2001

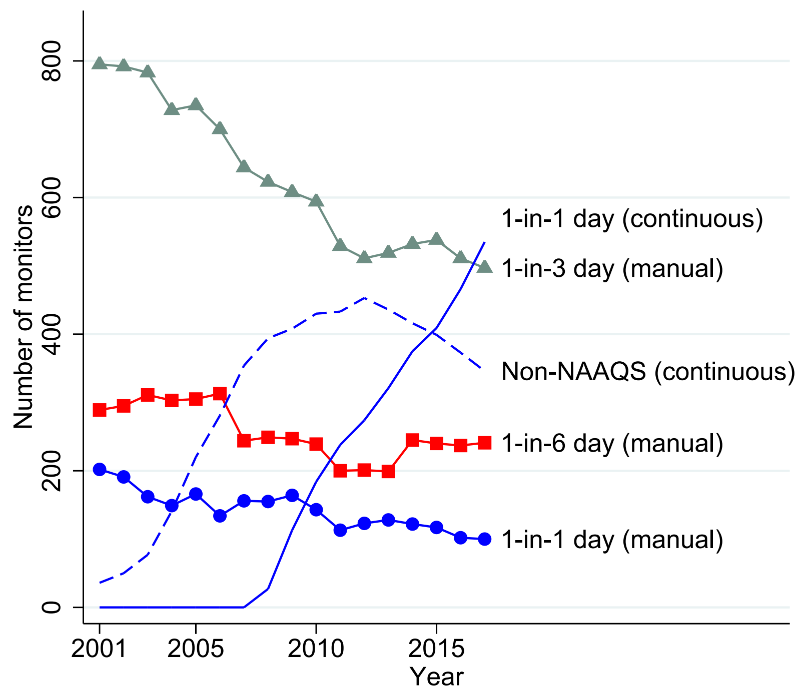


Panel B. Year 2013



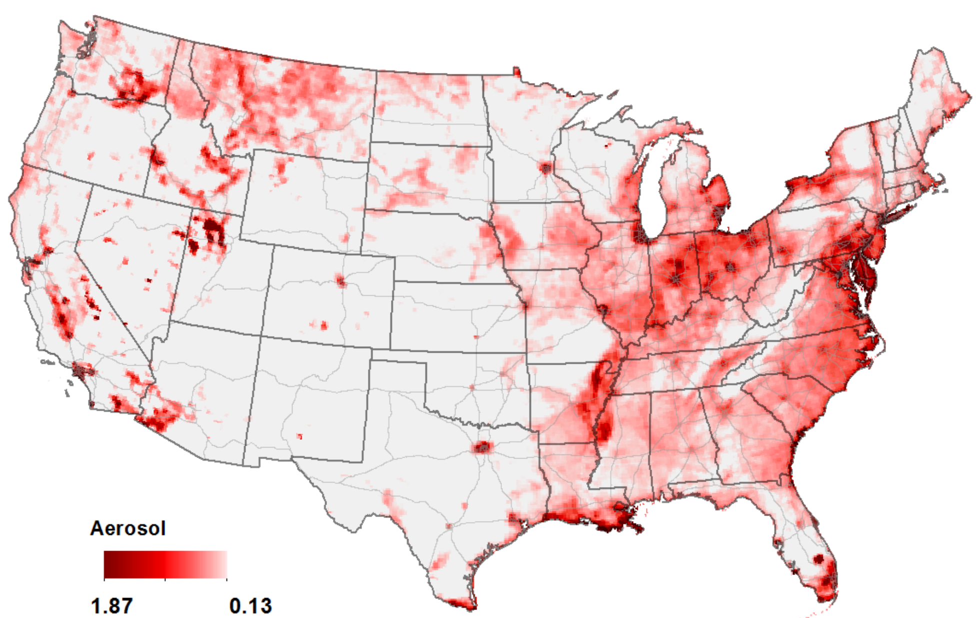
Notes: Map plots the 2001 (panel A) and 2013 (panel B) snapshot of the location of all PM monitors in the lower 48 states.

Figure D.4: Number of PM2.5 monitors by monitoring frequency



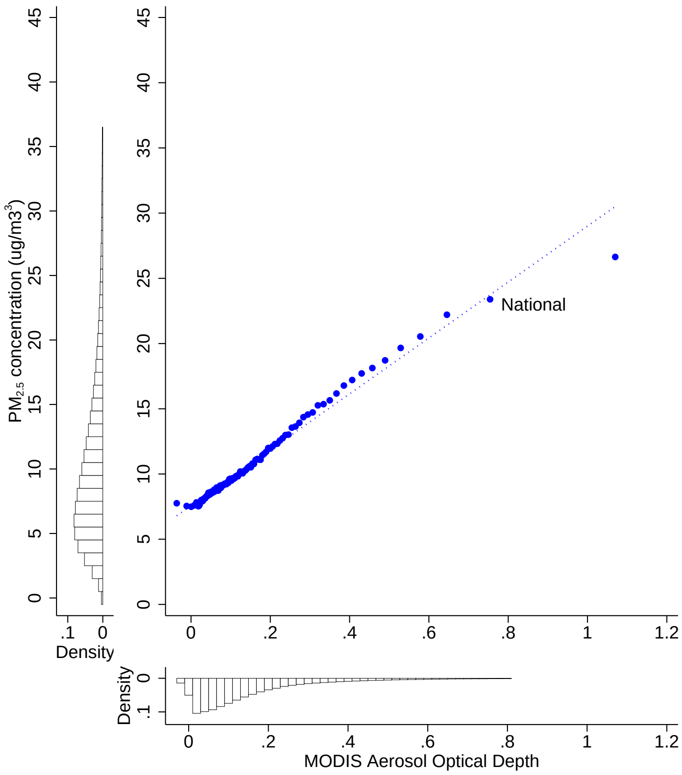
Notes: This figure plots annual number of PM2.5 monitors that follow different monitoring frequencies. "1-in-1-day (continuous)" represents 1-in-1-day monitors that adopt a Class III FEM technique that automates PM2.5 monitoring. "Non-NAAQS (continuous)" represents 1-in-1-day monitors that adopt non-FEM techniques; these monitors are not used in determining NAAQS nonattainment status designation.

Figure D.5: 10km×10km Aerosol Concentration, 2001-2013 Average



Notes: This map shows 13 year (2001-2013) average 10km×10km pixel level aerosol optical depth, among cells with above average concentration value. Light-gray lines represent major highways.

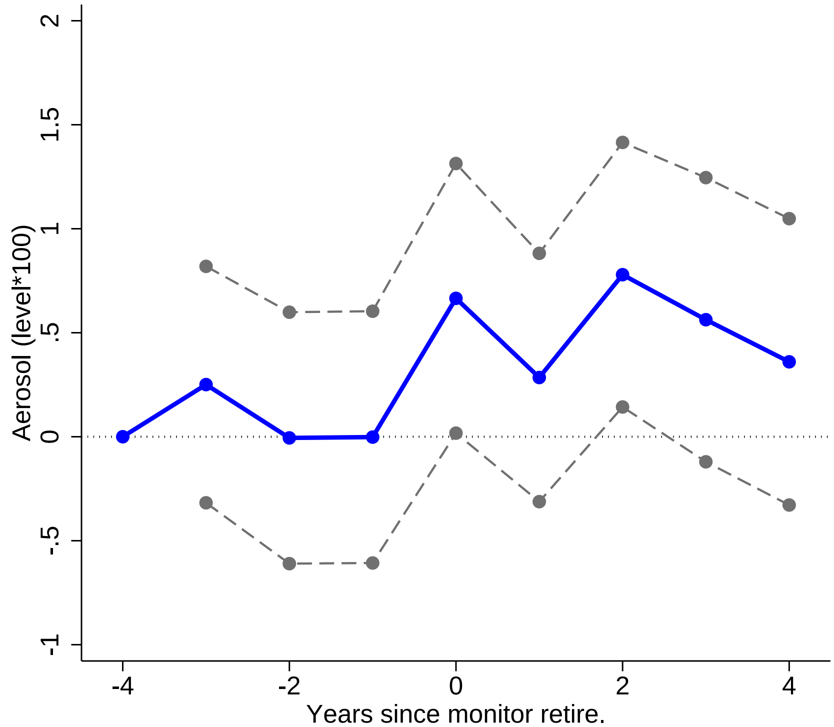
Figure D.6: PM2.5 and Aerosol Correlation, 2001-2013



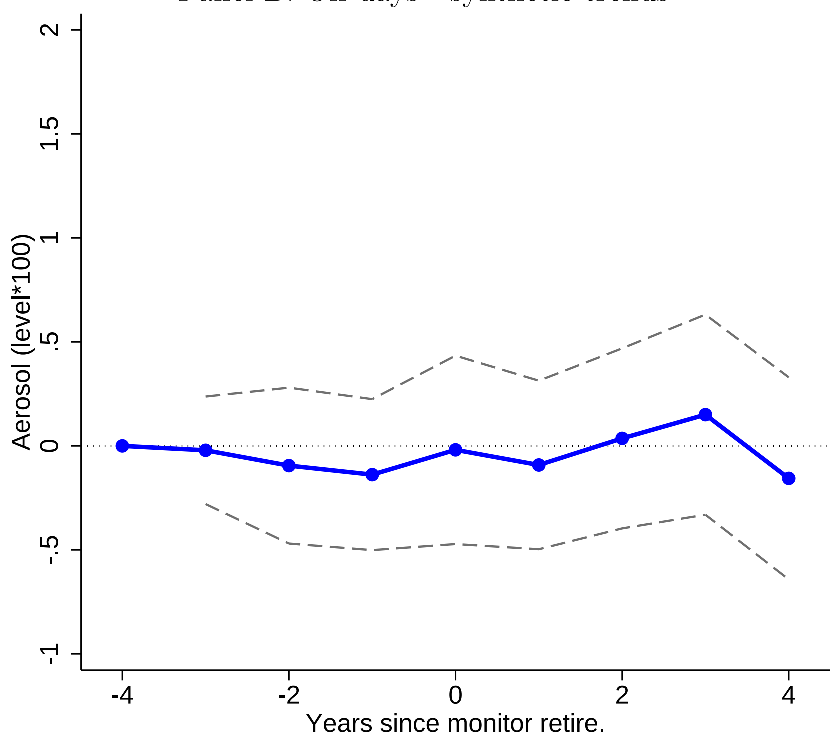
Notes: This figure presents the correlation between monitor PM2.5 readings and the satellite aerosol optical depth measure, defined as the aerosol level within the 10km×10km area where the monitor lives in. Dots show average PM2.5 within 100 equally-sized aerosol bin. Histograms show density of the raw PM2.5 and aerosol data.

Figure D.7: Changes in Levels of Aerosol Pollution by Years Relative to Monitoring Site Retirement

Panel A. On-days - synthetic trends

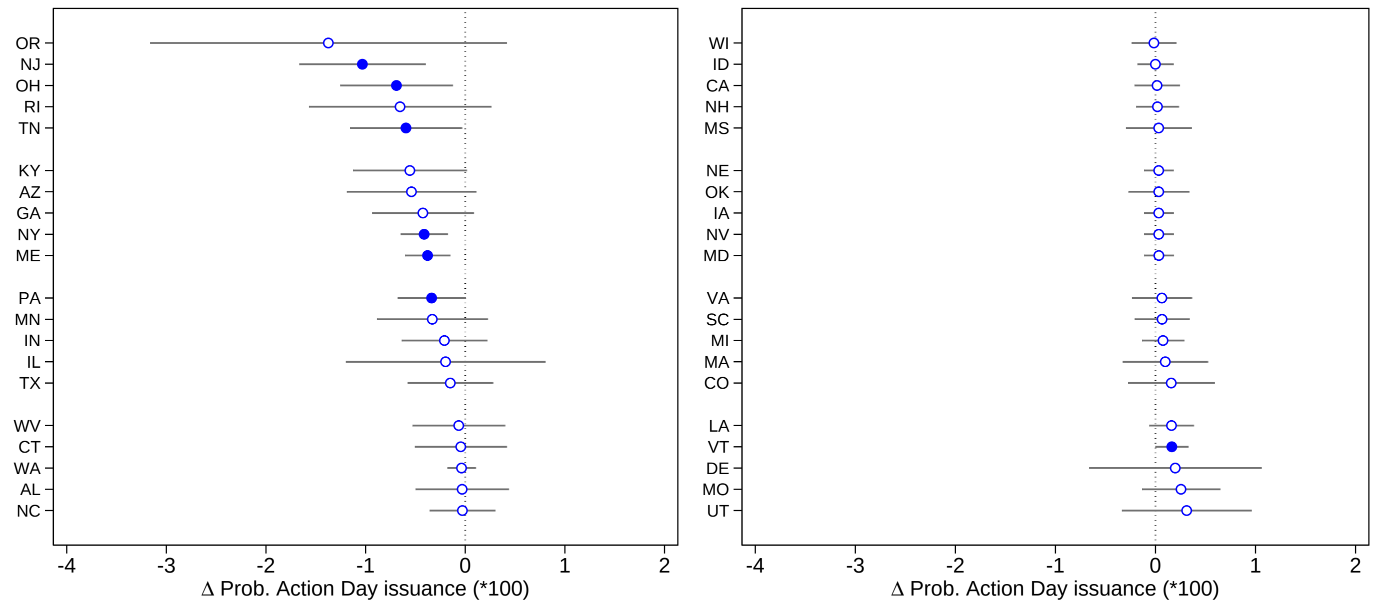


Panel B. Off-days - synthetic trends



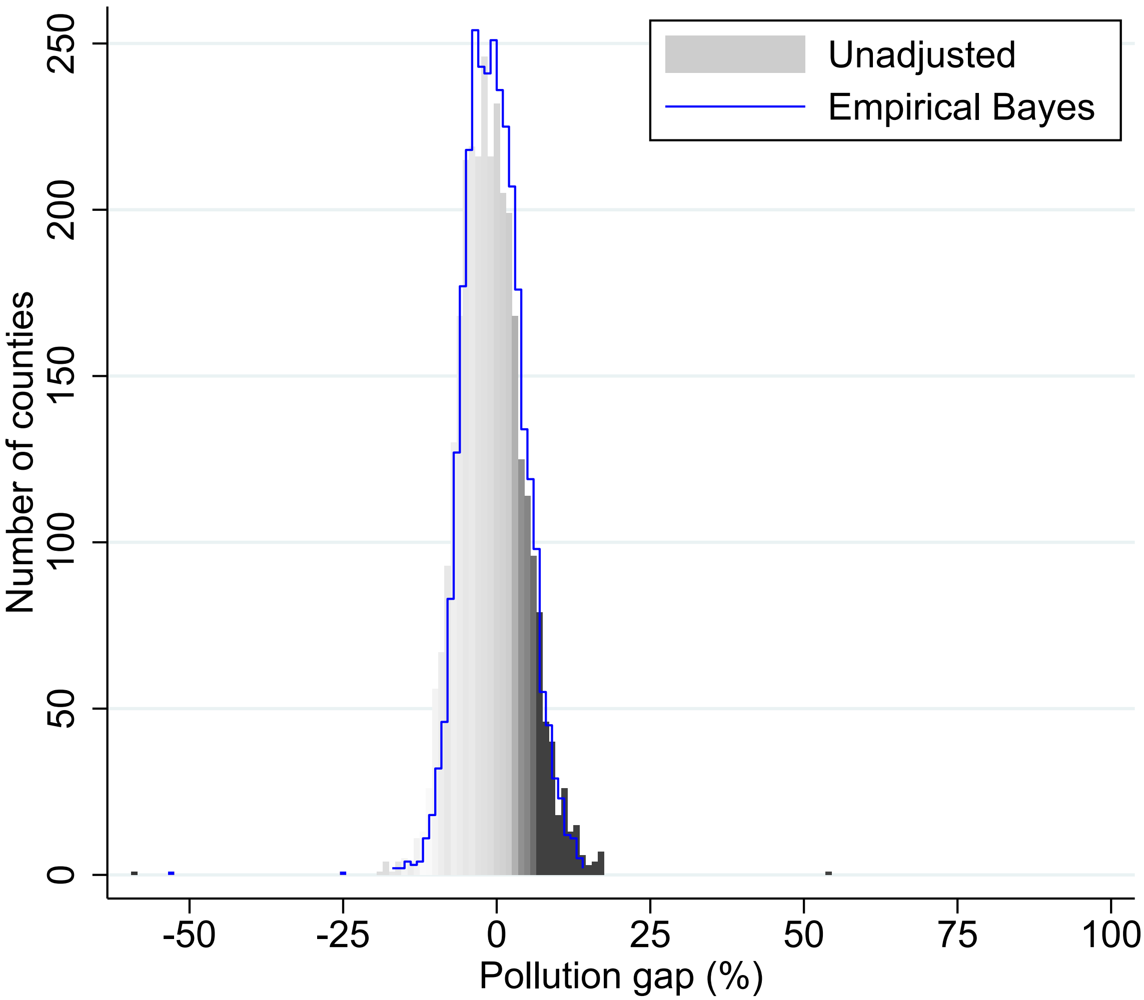
Notes: This chart shows regression estimates comparing on-day (panel A) and off-day (panel B) pollution and synthetic controls. Event year -4 is normalized to 0. The underlying regression controls for group fixed effects (retiring site and matched non-retiring sites within the same state), and no other controls. Dashed lines represent 95% confidence intervals constructed using standard errors clustered at the county level.

Figure D.8: Strategic "Pollution Action Day" Declarations: State Heterogeneity



Notes: This graph shows state-specific estimate of the differences in the odds that off-day and on-days coincide with Action Day advisory issuances. Coefficients are obtained from separate regressions by states that control for CBSA, year, month-of-year, and day-of-week fixed effects. Range bars show 95 percent confidence intervals.

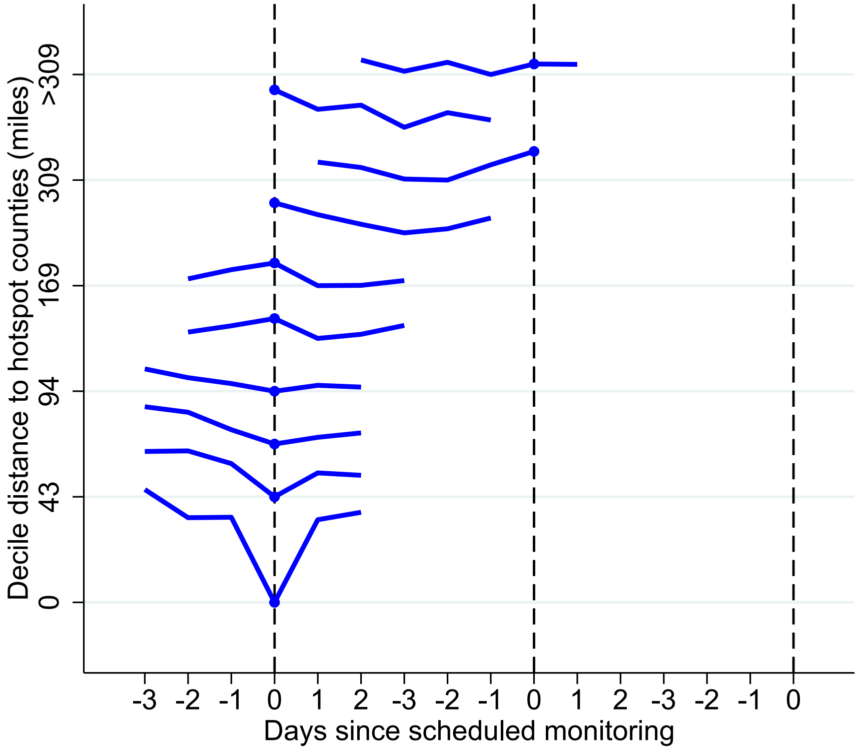
Figure D.9: Distribution of Raw vs. Empirical Bayes Estimates of County-Level 1-in-6-Day Pollution Gap



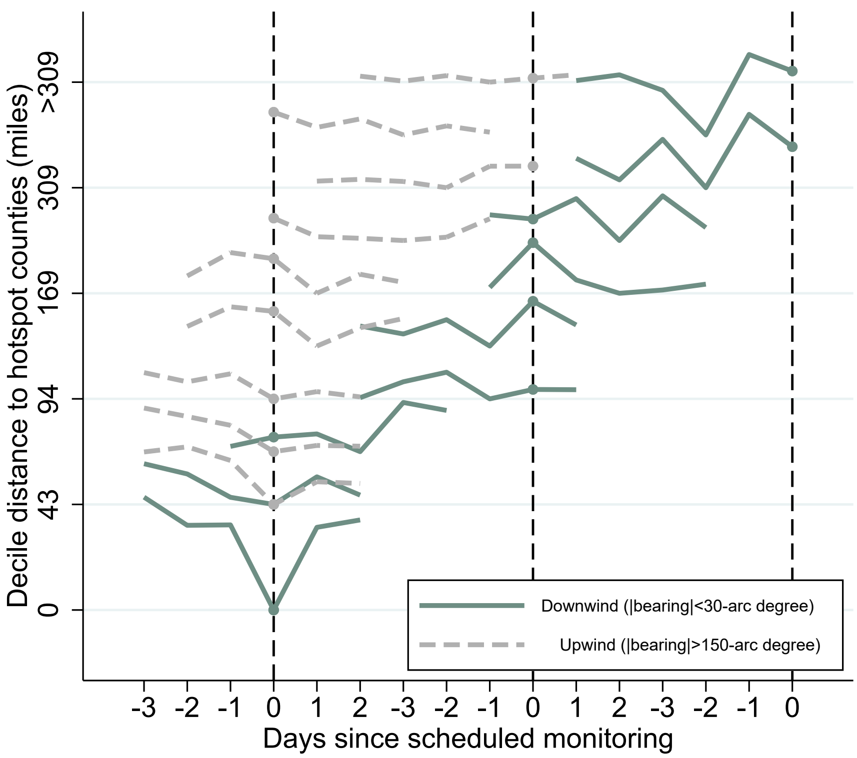
Notes: This graph plots a histogram of the raw (unadjusted) county-level pollution gap estimates and the Empirical Bayes-adjusted estimates. Raw estimates are painted so that darker colors indicate counties closer to the hot-spot (top 10% largest pollution gap) counties.

Figure D.10: Wind-Shifts of Pollution Gaps

Panel A. All counties

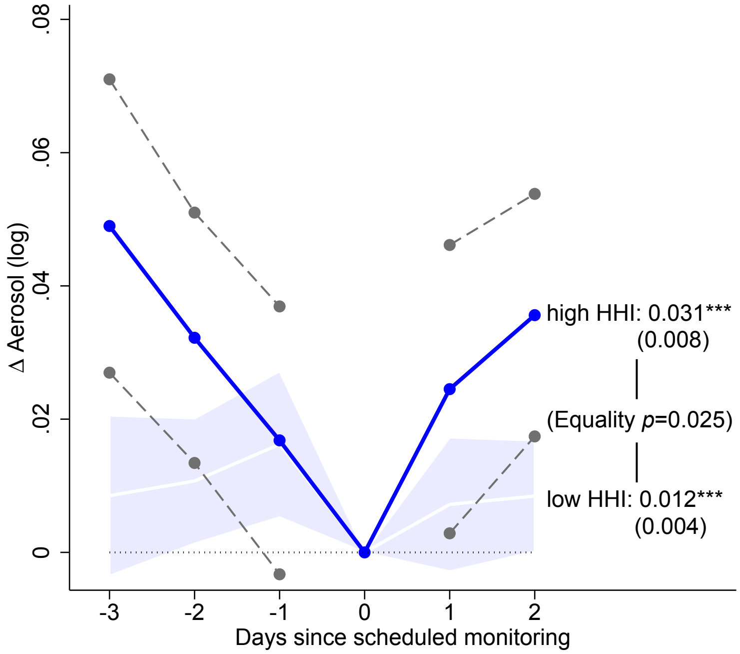


Panel B. Downwind vs. upwind counties



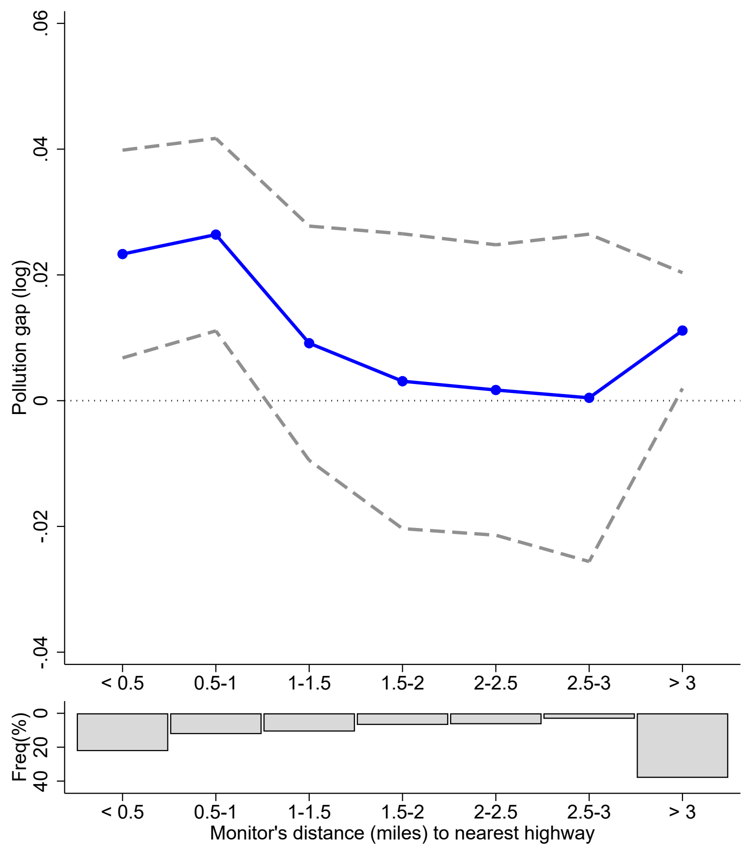
Notes: This chart shows the "shape" of the 1-in-6-day pollution gap, estimated separately for counties in each decile of distance to hot-spot counties indicated by the y-axis. Thus, the "distance = 0 mile" group represents the hot-spot counties themselves; the shape represents average aerosol levels by six day of monitoring cycle, indicated by the x-axis. The left panel shows data from all counties. The right panel repeats the same analysis, but does so separately for downwind counties (those with an absolute wind bearing of less than 30 arc-degrees relative to hot-spot counties) and upwind counties (those with an absolute wind bearing of over 150 degrees).

Figure D.11: Heterogeneous Pollution Gap by County's Emission Herfindahl-Hirschman Index



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

Figure D.12: Heterogeneous Pollution Gap by Monitor's Distance to Highway



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

Table D.1: Off-days vs. On-days Pollution Gap: Robustness to Sample Restrictions

Dep. var. = Aerosol concentration (log)				
	(1) Sample: sites w. any 1/6d monitor	(2) Sample: sites w. any 1/6d monitor	(3) Sample: sites w. only 1/6d monitor	(4) Sample: counties w. only 1/6d monitor
1(off-days)	0.016***	0.016***	0.018***	0.018***
	(0.004)	(0.004)	(0.004)	(0.006)
Ctrls		✓	✓	✓
N	685,060	685,060	427,846	176,225
N (site)	1,193	1,193	899	489
	(5) Sample: sites w. any 1/3d monitor	(6) Sample: sites w. any 1/3d monitor	(7) Sample: sites w. only 1/3d monitor	(8) Sample: counties w. only 1/3d monitor
1(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. "1(off-days)" indicates days when PM monitoring is not scheduled. "Controls" include site, year, month-of-year, and day-of-week fixed effects, and weather covariates (Section IIB). Standard errors are clusterd at the county level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Table D.2: Additional Summary Statistics

	(1) 1/1d	(2) 1/3d	(3) 1/6d
Ever in non-attainment (odds)	0.509	0.334	0.295
Daily traffic volume (count)	19,217	16,496	16,968
Industrial emissions concentration (HHI)	0.460	0.502	0.497
Employment in manufacturing (share)	0.092	0.115	0.097
Employment in utility (share)	0.0077	0.0065	0.0086
Employment in mining (share)	0.018	0.012	0.027
Employment in construction (share)	0.108	0.102	0.101
Employment in government (share)	0.120	0.121	0.132
State-level conservation scorecard	49.9	47.5	46.0
State-level Democrats affiliation (share)	0.499	0.508	0.500

Notes: One unit of observation is a monitor, and the monitored county's characteristics are reported. HHI is computed using the EPA's Toxic Release Inventory. Employment share is computed as employment counts per county's population from the Census County Business Pattern. Conservation score is computed using the League of Conservation Voter's score. Party affiliation is based on a 2006 Gallup Poll.

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

Dep. var. = Aerosol concentration (log)			
	(1) Sample: retired 1/6d sites	(2) Sample: 1/1d sites	(3) Sample: Non-PM 1/6d sites (HAPs)
1(off-days)	-0.0020 (0.0046)	-0.0050 (0.0077)	0.0023 (0.0044)
Ctrls	✓	✓	✓
N	372,989	231,532	370,020
N (site)	490	556	792

Notes: Each column reports a separate regression. The column names indicates the sample used. Column 1 includes areas that had 1-in-6-day PM monitoring sites that retired. Column 2 includes 1-in-1-day sites. Column 3 includes 1-in-6-day HAPs sites. "1(off-days)" indicates days when PM monitoring is not scheduled. "Controls" include site, year, month-of-year, and day-of-week fixed effects, and weather covariates (Section IIB). Standard errors are clusterd at the county level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Table D.4: Off-days vs. On-days Pollution Gap: Robustness to Specification Choices

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Baseline specification							
1(off-days)	0.0160***	0.0162***	0.0134***	0.0130***	0.0130***	0.0169***	0.0129***
	(0.0040)	(0.0035)	(0.0035)	(0.0036)	(0.0036)	(0.0034)	(0.0036)
N	685,060	685,060	685,059	684,961	684,735	685,060	685,060
Panel B: Difference-in-differences (1-in-1-day sites as control group)							
1(off-days)×1(1-in-6-day)	0.0173*	0.0149**	0.0125*	0.0138**	0.0142**	0.0154**	0.0140*
	(0.0103)	(0.0075)	(0.0071)	(0.0067)	(0.0065)	(0.0074)	(0.0078)
N	916,592	916,592	916,591	916,477	916,212	916,592	916,592
No ctrls.	✓						
Baseline ctrls.		✓	✓	✓	✓	✓	✓
Time FEs × state			✓				
Time FEs × county				✓			
Time FEs × site					✓		
Month-of-sample FEs						✓	
Week-of-sample FEs							✓
Number of FEs ctrls. (Panel A)	0	1,239	2,730	15,437	29,729	1,370	1,886
Number of FEs ctrls. (Panel B)	0	1,576	3,095	18,362	38,401	1,707	2,223

Notes: Each panel-column represents a separate regression. "1(off-days)" indicates days when PM monitoring is not scheduled. "Baseline controls" include site, year, month-of-year, and day-of-week fixed effects, and weather covariates (Section IIB). Standard errors are clustered at the county level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Table D.5: Pollution gap estimates by Pollution "Action Day" declaration

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

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

Table D.6: Particulate Matter Monitor Sampling Compliance

	(1) Samples required	(2) Samples taken	(3) Fraction taking ≥90% required samples	(4) Fraction taking 100% required samples
1/6day monitors	60 or 61	58.4 [2.2]	94.27%	19.21%
1/3day monitors	121 or 122	115.6 [4.4]	90.75%	5.42%
1/1day monitors	365 or 366	349.1 [13.0]	91.81%	6.33%

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

Table D.7: Heterogeneous Strategic Pollution "Action Day" Warnings by State Characteristics

Dep. var. = 1(warning)×100	(1)	(2)	(3)	(4)	(5)
State characteristics	Government size	Democrats affiliation	Conservation scorecard	Corruption index	NAAQS complaints
1(off-days)	-0.130**	-0.109*	-0.117**	-0.029	-0.040
	(0.052)	(0.059)	(0.052)	(0.040)	(0.045)
1(off-days)×1(>median states)	0.045	-0.001	0.018	-0.154**	-0.125*
	(0.069)	(0.073)	(0.069)	(0.068)	(0.068)
N	624,150	620,500	624,150	624,150	624,150

Notes: Each column represents a separate regression. 1(>median states) indicates states with above-median value of the corresponding characteristic. Standard errors are clustered at the CBSA level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.