

# How States Comply with Federal Regulations: Strategic Ambient Pollution Monitoring\*

Corbett Grainger<sup>†‡</sup>, Andrew Schreiber<sup>‡</sup>, and Wonjun Chang<sup>‡</sup>

<sup>‡</sup>Department of Agricultural and Applied Economics, University of Wisconsin-Madison, USA

October 9, 2016

## Abstract

In the United States ambient air quality is regulated through National Ambient Air Quality Standards (NAAQS), which are set by the EPA. Enforcement of these standards is delegated to state and sub-state regulators who are also tasked with designing their own monitoring networks for ambient pollution. A state's monitoring network is then used to determine whether a region, usually a county, is in attainment or nonattainment with the NAAQS. We develop an analytical model to study the incentives a local regulator faces with respect to siting new pollution monitors. We show that, for marginal counties (i.e. those that may be "close" to the NAAQS threshold), the regulator has an incentive to avoid siting pollution monitors in dirty locations. On the other hand, for counties already in nonattainment, the local regulator has an incentive to target pollution. To test for this type of behavior, we employ monitoring and satellite-derived pollution estimates to characterize pollution at non-monitored locations. We find that, on average, newly-sited monitors in marginal counties are placed in relatively clean areas, which suggests that local regulators strategically avoid pollution hotspots when siting monitors. We conclude with implications for the use of monitoring data as well as policy.

*JEL Codes: P48, Q52, Q53, Q56*

*Keywords: Strategic Regulator Behavior; Air Pollution Monitoring; Clean Air Act; Nonattainment; Distributional Impacts*

---

\*For comments and helpful conversations, the authors thank Max Auffhammer, Christopher Costello, Matthew Kahn, Gary Libecap, Nick Parker, Daniel Phaneuf, and Bill Provencher. The authors benefitted from feedback at the Occasional Workshop at UC Santa Barbara, the CU Workshop, the AERE Conference, and seminars at Wisconsin and UCSB. Any errors are our own.

<sup>†</sup>**Corresponding Author.** Email: [corbett.grainger@wisc.edu](mailto:corbett.grainger@wisc.edu)

# 1 Introduction

Under the Clean Air Act and its Amendments (CAAA)<sup>1</sup>, states, tribes and local governments are charged with ensuring that local ambient air quality complies with the National Ambient Air Quality Standards (NAAQS) set by the EPA. These standards, which are set based on the best available science to protect human health, are achieved through a variety of federal and local regulations. Counties (or other areas) found out of compliance with the NAAQS are designated as "nonattainment" areas. The economics literature has focused on enforcement of the Clean Air Act, but an understudied feature of this federal-state arrangement is *how* ambient pollution is actually monitored (a notable exception is a working paper by [Muller and Ruud \(2016\)](#), which we discuss in the next section). We show in this paper that states may face perverse incentives in collecting data for compliance with ambient standards. States are charged with monitoring their own ambient pollution levels, and as we detail in the following section, there is considerable flexibility in exactly how a state chooses the *locations* for its ambient pollution monitors. Depending on relative costs, this flexibility gives states an incentive for strategic behavior; this incentive is particularly strong for counties with subregions at risk for exceeding the NAAQS threshold.

Nonattainment designation, and the resulting reduction in pollution due to additional regulatory pressure, has received some attention in the economics literature. Subject matter spans from human health implications to reactionary firm behavior. [Auffhammer, Bento, and Lowe \(2009\)](#) show that nonattainment designation leads to targeted reductions in air pollution within nonattainment counties.<sup>2</sup> Nonattainment designation itself has been used as an instrumental variable (and the threshold has been used in a regression discontinuity design) by several researchers to identify hedonic models and distributional effects of regulations ([Chay and Greenstone 2005](#); [Grainger 2012](#)) or to identify impacts on health or mortality ([Chay and Greenstone 2003](#)). Nonattainment status has been shown to be costly<sup>3</sup>

---

<sup>1</sup>40 C.F.R. Subchapter C, Parts 50-97

<sup>2</sup>In addition, [Grainger \(2012\)](#) and [Bento et al. \(2015\)](#) both use the within-county variation in the pollution reduction to look at distributional impacts of the regulations.

<sup>3</sup>Central estimates from the EPA suggest that the benefits of the CAAA between 1990 and 2020 are orders of magnitude larger than the costs. Damages are estimated at \$56 Billion, whereas benefits (mostly from health and avoided mortality) are roughly 30 times higher ([USEPA 1999](#)).

for polluting firms (Becker 2005; Greenstone 2002), who face additional regulatory pressure to reduce emissions. Other literature focuses on how firms choose their location decisions in response to county-level nonattainment. For example, Kahn (1997), Henderson (1996), List et al. (2003) and Becker and Henderson (2000) study impacts of regulations, including on new plant siting choices.

Nonattainment status is also costly for state, local or tribal governments; roughly two-thirds of the administrative costs are borne by EPA, but the remaining costs fall on local governments. Under the CAAA, counties found to be out of compliance with the NAAQS are required to come back into compliance by submitting a State Implementation Plan, which details how the state plans to bring the air quality in the violating county back to acceptable levels. If the county fails to come back into attainment, EPA has authority to impose penalties on states such as withholding federal highway funding.

To study the incentives local regulators face in their monitor siting decision, we begin with an analytical model. We show that, under reasonable conditions, a regulator would target pollution when siting a new monitor in a county that is already in nonattainment.<sup>4</sup> However, for marginal counties (counties not designated as being in nonattainment, but where ambient pollution levels are reasonably high), the regulator has an incentive to avoid detecting pollution hotspots. These predictions are then tested empirically using a novel dataset and approach.

Because the ambient pollution monitoring data only show pollution levels at locations with monitors, we use remote-sensing data from several satellites to learn about pollution levels throughout the United States, with a particular focus on the areas surrounding pollution monitors.<sup>5</sup> We then compare ambient pollution (as measured remotely) at the monitor site to pollution levels in the surrounding area. To do so, we calculate a local z-score to com-

---

<sup>4</sup>In counties that are already designated nonattainment, the "penalty" has already been imposed, so to speak. In these counties, the regulator faces an incentive to come back into compliance, so she may wish to target polluted areas for source attribution.

<sup>5</sup>An alternative approach would be to use an atmospheric model to estimate pollution at locations without monitors, such as the Community Multiscale Air Quality model. These models are typically calibrated to the AQS monitors, however, so if monitors are strategically sited, it is possible that the models would miss important variation in pollution hotspots.

pare pollution at any grid cell to the surrounding "zone". We show that there is a significant difference in the z-scores for pollution at monitoring sites in "marginal" counties compared to areas already in nonattainment. Our empirical strategy then focuses on new monitor sitings that have taken place since 2006, and we demonstrate that monitoring decisions are significantly different between county "types".

The implications for our findings span disciplines and have important policy implications. First, depending on the type of county, our measurements of ambient pollution from monitoring networks are likely systematically biased. Monitor readings are not only used for compliance purposes, but they are used in atmospheric model calibration/validation, in economic studies, and in the public health literature. Second, there are areas of the country that are likely out of compliance with federal standards but, because they are not actively monitored, remain classified as being in attainment. Third, there are likely important environmental justice implications of our findings. A central finding in the environmental justice literature is that poorer populations tend to live in more polluted areas. As we discuss in this paper, because some pollution hotspots have gone undetected by ambient monitors, there has been a less-than-efficient targeting of pollution across space, and poorer populations are likely disproportionately affected. Finally, our results suggest that there are potentially large gains from additional oversight or guidance from the federal government when local regulators choose where to site new monitors. Given the recent availability of remote sensing data for ambient air pollution, this information could be leveraged to better target pollution hotspots.

## 2 Background

Regulations for air quality could be classified into two general categories: emissions regulations, and ambient air quality regulations. In this paper, we are concerned with the latter, in particular nonattainment designation and the monitor placement choices of local regulators. We are aware of only one other paper that studies ambient pollution monitoring choices by the regulator. [Muller and Ruud \(2016\)](#) study how monitored ozone levels in one period affect the siting or removal of monitors in subsequent periods. A major difference between their

approach and the current paper is our use of satellite data to detect strategic behavior in the placement of monitors. Under the CAAA, the local regulator<sup>6</sup> is charged with ensuring compliance with federal standards for air quality, but the regulator is also charged with maintaining a monitoring network to record local ambient air pollution levels.

Ambient pollution monitors, which are sited and maintained by local regulatory authorities, are used to provide continuous monitoring of ambient air pollution.<sup>7</sup> The fixed setup costs of a monitor are non-negligible to the local authority, and in general monitors are only placed when mandated, because the cost of a new monitor is high.<sup>8</sup> Monitors are generally placed on trailers or on rooftops, and they must have adequate space for the associated electrical and computer equipment. Furthermore, a site must have a source of power for heating and cooling the instruments and computers, and the instruments must be calibrated and maintained by engineers. The EPA provides guidance on the standards that a monitor must meet. Furthermore, once placed, a monitor is difficult to remove—the state (or local regulatory agency) must demonstrate that the historical pollution at the site is well below the NAAQS threshold and that the probability of future nonattainment at that monitor location is very low. In practice, the monitor’s location is viewed as a permanent decision by the local regulator.

Importantly, the local regulator is given considerable discretion in determining its mon-

---

<sup>6</sup>This is usually a state-level agency, but in some cases substate organizations may have authority, such as the case in California.

<sup>7</sup>There are many possible objectives that a regulator faces when siting monitors; in this paper, we focus on incentives for "discovering" or "avoiding" pollution hotspots. In a separate working paper, [Chang and Grainger \(2016\)](#) develop a model of optimal monitoring for ambient pollution under alternative objectives. The World Health Organization ([WHO 1999](#)) lists multiple objectives that should be met by a monitoring system, including population exposure and health impact assessment; identifying threats to natural ecosystems; determining compliance with national or international standards; informing the public about air quality and establishing alert systems; providing objective input to air quality management and to transport and land-use planning; identifying and apportioning sources; developing policies and setting priorities for management action; developing and validating management tools such as models and geographical information systems; quantifying trends to identify future problems or progress in achieving management or control targets.

<sup>8</sup>Furthermore, because nonattainment can be triggered by a "bad day" at a single location, there is little incentive to increase the network size within a county.

itoring strategy. For the criteria pollutants regulated under the Clean Air Act, the Code of Federal Regulations sets guidelines for how a monitoring network should be established.<sup>9</sup> Our discussion here focuses on  $NO_2$  because our remote sensing data do particularly well at detecting  $NO_2$ , but the same flexibility is present in the rules guiding the placement of monitors for all criteria pollutants. In practice there are few guidelines from the federal government regarding the placement of ambient pollution monitors, but important for this study, states (or tribal, or substate regulatory agencies) are required to establish a plan to place, or identify, an area-wide  $NO_2$  monitor;  $NO_2$  monitors should be placed to characterize vulnerable and susceptible populations; and core-based statistical areas (CBSAs) are required to have varying numbers of near-road monitors, depending on the population and traffic density along major freeways.<sup>10</sup> Area-wide monitors, which are meant to be representative of a larger spatial scale, have additional rules in the Code of Federal Regulations. Specifically, the agency shall "monitor a location of expected highest  $NO_2$  concentrations representing the neighborhood or larger spatial scales". Furthermore, emissions inventories and meteorological analysis should be used to identify the appropriate locations within a CBSA for locating an area-wide monitoring station.

While there are guidelines and rules to establish a network of pollution monitors, there exists considerable flexibility in determining the exact location of each monitor in the network. This discretion allows the regulator to choose locations strategically, depending on local conditions. The analytical model in the following section demonstrates the conditions under which the regulator would have an incentive to target or avoid pollution when siting a new monitor.

### 3 Model

To begin, we assume that a state (or substate, or tribal) regulator is charged with complying with federal ambient pollution standards, such as the NAAQS discussed above. The federal

---

<sup>9</sup>For  $NO_2$  monitors, for example, see 40 U.S.C. 1(C)§58.

<sup>10</sup>Specifically, CBSAs with 1,000,000 people or 500,000 people and significant traffic require a single monitor, and a second monitor is required if the CBSA population exceeds 2.5 Million (or if thru-traffic increases substantially).

government sets ambient pollution standards and nonattainment penalties, and the local regulator, taking the standards and penalties as given, chooses how to monitor ambient air pollution.

Our approach is similar in spirit to the analytical model of Rabassa (2008), who looks at local enforcement of the CAAA through emissions inspections. However, while he focuses on emissions inspections, we hold constant enforcement effort and instead focus on the incentives to monitor *ambient* pollution through monitor siting. Following Rabassa’s model, the local regulator’s objective is to maximize political support, net of enforcement costs and penalties. There is a representative firm that emits pollution, which in turn causes a degradation in ambient air quality. We also allow for other exogenous pollution and a random component, so that the ambient pollution experienced by the local population is uncertain.

### 3.1 Model Setup

The regulator’s decision variables are monitoring efforts,  $D$ , and other air pollution regulation efforts,  $e$ . To focus on optimal ambient air pollution monitoring efforts we assume that the level of other enforcement  $e$  is held fixed at level  $\bar{e}$ .

The regulator’s objective function contains several arguments, which we consider in turn. First,  $g(A)$  is the health damage from ambient pollution  $A$ , which is increasing in  $A$  at an increasing rate:  $g' > 0$ ,  $g'' > 0$ .  $h(E)$  is the support signal sent by point sources (industry), which depends on emissions  $E$ , for which we have  $h' > 0$  and  $h'' < 0$ . Emissions  $E(\bar{e}, D)$  is a function of monitoring efforts,<sup>11</sup>  $D$  and all other emissions regulation efforts,  $e$ . We assume that  $E' (= \frac{\partial E}{\partial D}) < 0$  and  $E'' > 0$ . Lastly,  $c(\bar{e}, D)$  is the monetary cost of overall enforcement efforts of the regulator, where  $c' > 0$ , and  $c'' > 0$ .

Ambient pollution,  $A$ , is a function of point source emissions,  $E$ , exogenous non-point source emissions,  $\gamma$ , and a stochastic variable,  $x$ , whereas *monitored* ambient pollution is given by  $\tilde{A}$ .  $\tilde{A}$  is used to enforce the nonattainment designation for ambient air pollution.

---

<sup>11</sup>There is a vast literature discussing the complicated relationship between monitoring, enforcement and compliance; see Shimshack (2014) for a recent review.

That is, if  $\tilde{A} > A_c$ , there is a penalty,  $Z$ , incurred by the state.

$$A = \alpha E(\bar{e}, D) + \gamma + x$$

$$\tilde{A} = D\alpha E(\bar{e}, D) + \gamma + x$$

The influence firm level emissions have on ambient pollution concentrations is captured in  $\alpha$ , a pollution transport coefficient. For instance, if  $\alpha = 0$ , all local emissions travel outside the study area and do not affect local ambient concentrations. Conversely if  $\alpha$  is high, a majority of firm level emissions contribute to ambient levels.

Monitoring effort,  $D$ , measures how well the monitored ambient air quality represents the incidence of emissions on local residents. We assume that non-point source emissions,  $\gamma$ , and the stochastic term,  $x$ , are exogenous, realized with certainty and are fully detectable, which allows monitoring to be focused entirely on point source emissions,  $E$ . A higher  $D$  can represent the location of an ambient air quality monitor that better captures the true externalities of air pollution emissions. Assume that  $0 \leq D \leq 1$ .

Uncertainty in ambient pollution concentrations is captured by the random variable  $x \sim N(0, \sigma_x)$ . We let  $f(x)$  denote the probability density function and  $F(x)$  be the corresponding cumulative distribution function.

### 3.2 Analytical Model

The local regulator's objective is to maximize political support, net of enforcement costs and penalties. Therefore, given the definitions above, the regulator faces the following maximization problem:

$$\begin{aligned} \max_D \mathbb{E}[U(A)] &= \int_{-\infty}^{A_c - D\alpha E(\bar{e}, D) - \gamma} \left[ -g(A) + h(E) - c(\bar{e}, D) \right] f(x) dx \\ &\quad + \int_{A_c - D\alpha E(\bar{e}, D) - \gamma}^{\infty} \left[ -g(A) + h(E) - c(\bar{e}, D) - Z \right] f(x) dx \\ &= \mathbb{E}[-g(A)] + h(E) - c(\bar{e}, D) - (1 - F(A_c - D\alpha E(\bar{e}, D) - \gamma))Z. \end{aligned} \tag{1}$$

Differentiating with respect to  $D$ , the first-order condition is given by

$$\alpha \mathbb{E}[-g']E' + h'E' - c' - \alpha f(A_c - D\alpha E(\bar{e}, D) - \gamma)(E + DE')Z = 0. \tag{2}$$



Rearranging terms, equation (2) can be written to express the change in the marginal benefits of monitoring:

$$\underbrace{\alpha \frac{\partial E}{\partial D} \mathbb{E}[g'] + \alpha f(A_c - D\alpha E(\bar{e}, D) - \gamma) \underbrace{\left( E(\bar{e}, D) + D \frac{\partial E}{\partial D} \right)}_{MCM}}_{\text{marginal nonattainment effect}} Z = \underbrace{h' \frac{\partial E}{\partial D} - c'}_{MC}. \quad (3)$$

We denote the net marginal contribution of monitoring efforts to the monitored ambient pollution levels,  $(E + DE')$ , as the *marginal contribution of monitoring* ( $MCM$ ). This term represents the increase in pollution exposure net of the decrease in point source emissions due to increased regulation (monitoring).

The model can then be used to develop the following propositions.

**Proposition 1.** *If the marginal contribution of monitoring is non-negative ( $MCM \geq 0$ ), expected ambient pollution is below or at the nonattainment threshold, and  $MCM$  is weakly increasing in  $D$  (i.e.  $\partial MCM / \partial D \geq 0$ ), then optimal monitoring efforts are decreasing in expected pollution levels; i.e.;  $\frac{\partial D^*}{\partial \mathbb{E}[A]} < 0$ .*

*Proof.* The corresponding second order condition of the model is:

$$\begin{aligned} \alpha \mathbb{E}[-g'] E'' + \alpha^2 \mathbb{E}[-g''] E'^2 + h'' E'^2 + h' E'' - c'' \\ + \alpha^2 Z f' \cdot (E + DE')^2 \\ - \alpha Z f \cdot (2E' + DE'') \end{aligned} \quad (4)$$

From equation (2) we have that:

$$\alpha \mathbb{E}[-g'] + h' = \frac{\alpha f(E + DE') Z + c'}{E'},$$

which is substituted into the second-order condition above, yielding

$$\begin{aligned} \alpha E'' \left( \frac{\alpha f \cdot (E + DE') Z + c'}{E'} \right) + \alpha^2 \mathbb{E}[-g''] E'^2 + h'' E'^2 - c'' \\ + \alpha^2 Z f' \cdot (E + DE')^2 \\ - \alpha Z f \cdot (2E' + DE''). \end{aligned} \quad (5)$$

By assumption  $MCM$  is positive and is a weakly increasing function of monitoring effort, so the following two conditions must hold

$$MCM \equiv E + DE' \geq 0 \quad (6)$$

$$MCM' \equiv 2E' + DE'' \geq 0.$$

Through this condition, we assume emissions levels are lower bounded by the lowest achievable emissions rate<sup>12</sup> ( $LAER$ ) and/or that  $\frac{\partial E}{\partial D}$  has an upper bound. We also assume that as monitoring efforts increase, the rate at which pollution is exposed exceeds the rate at which firm emissions decrease.

As expected ambient pollution levels do not exceed the threshold  $\mathbb{E}[A] \leq A_c$  (and hence  $\tilde{A} \leq A_c$ ), it is the case that  $f'(\cdot) \leq 0$ . Note that under such conditions, the second-order condition is non-positive, which allows for the first-order condition to be sufficient for optimality.

Now given equation (3) we have the following:

$$\underbrace{\alpha \frac{\partial E}{\partial D} \mathbb{E}[g'] + \alpha f(A_c - D\alpha E(\bar{e}, D) - \gamma) \underbrace{\left( E(\bar{e}, D) + D \frac{\partial E}{\partial D} \right)}_{MCM}}_{\text{marginal benefit of monitoring}} \underbrace{Z}_{MC} = \underbrace{h' \frac{\partial E}{\partial D} - c'}_{MC}. \quad (7)$$

Optimal monitoring effort,  $D^*$ , is determined where marginal benefits equal marginal costs. The *marginal non-attainment effect* (MNE)—the marginal increase of expected penalty payments due to a unit increase in pollution exposure—decreases the benefits of an additional unit of monitoring effort, due to a positive MCM. This means that the marginal benefits of monitoring air pollution is *smaller* than in the case when we ignore the penalty for noncompliance. In other words,  $E, \gamma$  and  $Z$  all contribute toward a positive MNE, and as a result,  $\frac{\partial D^*}{\partial E} < 0, \frac{\partial D^*}{\partial \gamma} < 0, \frac{\partial D^*}{\partial \mathbb{E}[A]} < 0, \frac{\partial D^*}{\partial Z} < 0$ .

□

**Proposition 2.** *Under the same assumptions, optimal monitoring efforts are decreasing in the penalty value  $Z$ ,  $\frac{\partial D^*}{\partial Z} < 0$ .*

---

<sup>12</sup>LAER refers to the most restrictive possible emissions standards defined in the State Implementation Plans in the case of nonattainment designation.

*Proof.* This follows directly from proof of Proposition 1.  $\square$

**Proposition 3.** *If expected ambient pollution levels exceed threshold  $A_c$  and the nonattainment penalty ( $Z$ ) is incurred with certainty, then optimal monitoring efforts are increasing in expected pollution levels; i.e.,  $\frac{\partial D^*}{\partial \mathbb{E}[A]} > 0$ .*

*Proof.* As the nonattainment penalty is now effectively a sunk cost, the model simply becomes:<sup>13</sup>

$$\begin{aligned} \max_D \mathbb{E}[U(A)] &= \int_{-\infty}^{\infty} \left[ -g(A) + h(E) - c(\bar{e}, D) - Z \right] f(x) dx \\ &= \mathbb{E}[-g(A)] + h(E) - c(\bar{e}, D) - Z \end{aligned} \quad (8)$$

The first-order condition with respect to  $D$  is:

$$\alpha \mathbb{E}[-g']E' + h'E' - c' = 0. \quad (9)$$

The second order condition is given by:

$$-\alpha^2 \mathbb{E}[g'']E'^2 - \alpha \mathbb{E}[g']E'' + h''E'^2 + h'E''. \quad (10)$$

From equation (9), we have  $\mathbb{E}[g'] = h'/\alpha - c'/\alpha E'$  which is used to replace terms in equation (10) to obtain:

$$\frac{c'E''}{E'} + E'^2 \{ \alpha^2 \mathbb{E}[g''] + h'' \} < 0. \quad (11)$$

Concavity of the optimization problem guarantees optimality through the first-order condition, which can be written in terms of marginal benefits and costs. Therefore,

$$\underbrace{\alpha \mathbb{E}[g'(A)]E'}_{MB} = \underbrace{h'E' - c'}_{MC}. \quad (12)$$

Since an increase in  $\mathbb{E}[g'(A)]$ , through an increase in  $E(\bar{e}, D^*)$  and/or  $\gamma$ , would shift the marginal benefits of monitoring upwards for each monitoring level, the optimal level of monitoring increases for marginal costs to match marginal benefits. As a result, we have  $\frac{\partial D^*}{\partial \mathbb{E}[A]} > 0$ .  $\square$

---

<sup>13</sup>Though there is still a random component determining the ambient pollution level  $A$ , the area in question could have already incurred the penalty. For example, a county could currently be designated as nonattainment when the regulator is choosing its monitoring effort.

Thus we move to the empirical section of the paper with testable hypotheses: counties already in nonattainment will place monitors in relatively dirty areas, targeting pollution, whereas marginal counties will avoid pollution hotspots by siting new monitors in relatively clean locations. Now we briefly discuss the data before proceeding to our empirical approach.

## 4 Data

### 4.1 AQS Monitoring Data

To examine the regulator’s choice of monitoring location, we use the Air Quality System (AQS) data from the US Environmental Protection Agency. The AQS data contain the universe of ambient pollution monitors that are used for compliance in the United States and come from the network of monitors placed by state, local and tribal governments. The monitoring data provide the longitude and latitude positioning and hourly pollution readings of each active monitor (we focus on  $NO_2$  and ozone). The spatial distributions of both ozone and  $NO_2$  monitors locations<sup>14</sup> for the United States are given in figure 1. The figure describes all active monitors for at least one year from 2005-2012. More monitors are used to regulate ambient ozone concentrations than for  $NO_2$  and both have spatial concentrations that are more dense in urban areas due to federal guidelines.

Table 1 provides descriptive statistics for the AQS data. The table provides the number of monitors active, newly sited, retired, and re-sited to within 3 kilometers of the location of a previously retired monitor in a given year for both ozone and  $NO_2$ . Sited monitors are defined as locations that did not have a monitor in the previous year. Retired monitors are defined as locations that did not have a monitor in the following year<sup>15</sup>. Overall there are very few re-sited monitors in the dataset<sup>16</sup>.

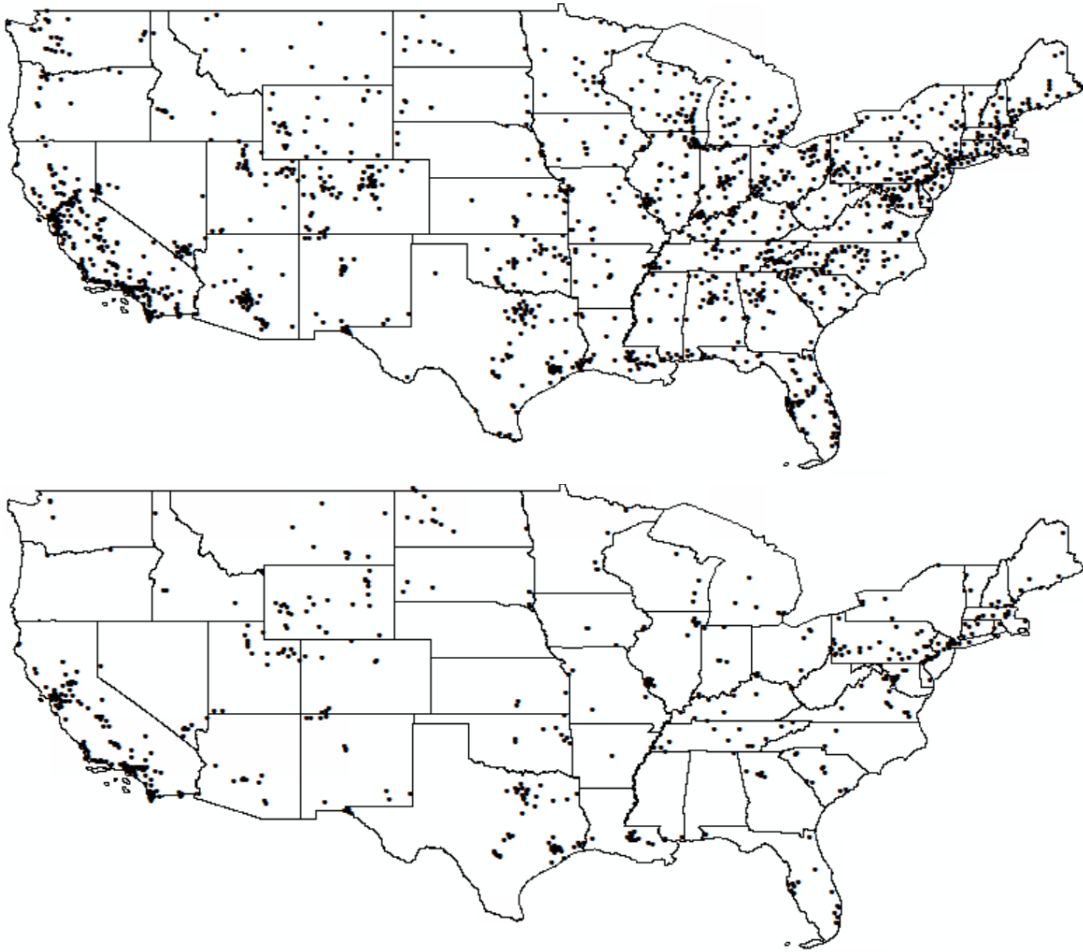
---

<sup>14</sup>All coordinates are translated to the WGS84 coordinate system using ArcMap. Many (but not all) monitors are for multiple pollutants.

<sup>15</sup>In this paper we focus on sitings of new monitors; in a subsequent paper we will examine retirements and pollution dynamics. Also note that we are unable to calculate the number of sited monitors for 2005 and the number of monitors retired in 2012 as they represent the first and last years of collected data.

<sup>16</sup>While 3 kilometers is chosen as a reasonable distance to consider a monitor as re-sited, we’ve computed this number for up to 10 kilometers away. The number of re-sited monitors using larger bandwidths remains

Figure 1: Air Quality System Pollution Monitors



Notes: Black dots denote ambient air pollution monitors. The top map represents the number and distribution of Ozone monitors in the AQS database from 2005-2012. The bottom map represents the number of  $NO_2$  monitors in the AQS database for the equivalent time period.

The monitoring data are generally reported hourly, but annual tables are available which include one-hour averages, 8-hour averages, and several percentiles following the NAAQS for each pollutant. The NAAQS for  $NO_2$  include a primary and secondary annual standard of 53 ppb, and a primary one-hour standard of 100 ppb (98th percentile of one-hour daily maximum concentration, averaged over three years). For ozone, the 2008 standard<sup>17</sup> dictates

---

roughly constant.

<sup>17</sup>The nonattainment designation in our data follows the contemporaneous standard, so 2005 nonattainment, for example, follows the NAAQS in place in 2005, which was the 1997 standard of 0.08 ppm

Table 1: AQS Monitor Sitings

<i>Year</i>	<i>NO<sub>2</sub></i>				<b>Ozone</b>			
	<i>Total</i>	<i>Sited</i>	<i>Retired</i>	<i>Re-Sited</i>	<i>Total</i>	<i>Sited</i>	<i>Retired</i>	<i>Re-Sited</i>
2005	441		30		1201		41	
2006	434	23	25	1	1207	47	41	1
2007	432	23	31		1232	66	50	1
2008	422	21	34	1	1235	53	42	1
2009	407	19	18		1250	57	33	1
2010	417	28	39	1	1274	57	47	1
2011	400	22	19		1331	104	42	1
2012	410	29		1	1324	35		1

that the annual fourth-highest daily maximum 8-hour concentration averaged over three years is 0.075 ppm for both the primary and secondary standards.

## 4.2 Satellite Data: BEHR and OMI

Our characterization of the problem necessitates information about the entire pollution distribution over the United States. We employ a state-of-the-science satellite dataset from the Berkeley High Resolution (BEHR) group (Russell et al. 2011) for nitrogen dioxide data and formaldehyde data<sup>18</sup> from NASA’s OMI (Ozone Monitoring Instrument) from 2005-2012. The NASA OMI project seeks to describe pollution levels over the entire world by employing satellites. NASA data is available for a variety of criteria pollutants (ozone, sulfur dioxide, etc.) while the BEHR data, though based on NASA OMI satellite imagery, is only available for  $NO_2$  for the contiguous United States. Relative to NASA, BEHR provides finer resolution pollution estimates regridded to a  $0.05^\circ \times 0.05^\circ$  grid cell level (roughly equaling a  $5 \times 5km$  cell).

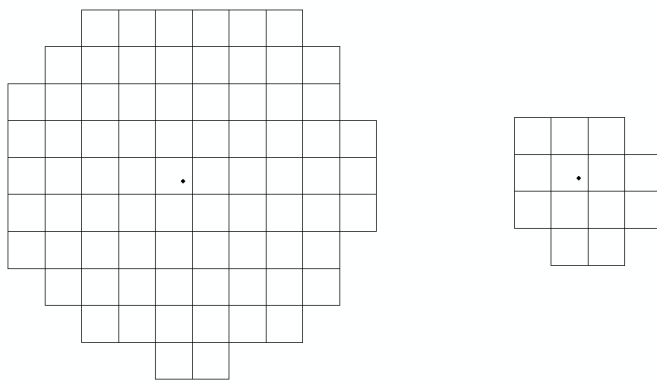
Extracting satellite data<sup>19</sup> presents computational challenges because of its sheer size and

<sup>18</sup>Satellite data for ozone is not as reliable as for aerosols and  $NO_2$ . Moreover, NASA provides the only known remote sensing data for ozone, which is at too large a resolution to provide a sufficient means of study. As detailed in a later section, we can use a combination of  $NO_2$  and formaldehyde to proxy for true ozone levels. We discuss this in greater detail later.

<sup>19</sup>A variety of quality control checks were built into our data extraction programs. Aside from following basic quality flags, we followed the BEHR recommendations for dropping some observations and only

the time frames we are interested in. For this reason, we restrict data collected for pollution estimates only around monitors for years 2005-2012 at a radius of 25km (and decreasing iterations thereafter as a robustness check). The 25km radius was chosen as an upper bound on the relevant range of options available to the regulator at the time of siting. Consider figure 2 for a visual of our spatial configuration. We designate the area around each monitor as a *zone* which describes the set of potentially viable monitoring sites.

Figure 2: BEHR Grid Zones



Notes: The black dot represents the air quality monitor. BEHR data were extracted for 25 kilometers and 10 kilometers from the monitored site. The left side represents grid centroids within 25 kilometers of the monitor and the right the equivalent 10 kilometer case. The shape of the zone will depend on the monitor's positioning in the BEHR grid cell. Our main specifications rely on the 25-km data, as they provide more variation in pollution estimates, but our results are robust to the chosen zone radius. Results at the 10-km level are available in an online appendix.

Satellite pollution estimates of  $NO_2$  perform well when ground-truthed to ground level estimates of pollution, as has been documented thoroughly in the atmospheric science literature (i.e. [Bechle et al. \(2013\)](#)). The BEHR  $NO_2$  data do particularly well when comparing concentrations within a smaller region, such as at the city level, which we do in our empirical application. But even at larger spatial scales, a simple linear regression of the AQS  $NO_2$  concentrations on the BEHR estimates of  $NO_2$  explains 30-35% of the variation at the country level with positive correlation. Within a smaller area, the  $R^2$  is comparable or

---

extracted data with limited cloud coverage.

higher<sup>20</sup>.

One issue with using satellite pollution data concerns the mismatch in comparable units. The satellite data are measured as column densities (molecules/ $cm^2$ ), while ground level pollution is monitored in either parts per billion (ppb) for  $NO_2$  or parts per million (ppm) for ozone. Converting the column densities to ground level estimates would require the use of an atmospheric model which considers things like average meteorology, topography, etc. As such, we just use the raw column densities as our pollution estimates which is not problematic if considering *within-zone* variation of the data points around each sensor, as opposed to variation across sensors. We circumvent the issue with the introduction of localized z-scores. Because we are interested in how pollution at any given grid cell compares to the surrounding region, we take the observed value in a grid cell  $i$  in zone  $z$  and year  $t$ , subtract the average for that zone,  $z$ , and scale it by the standard deviation for the same zone. That is, the z-score for pollutant  $p$  is given by  $zscore_{izt}^p = (p_{izt} - \bar{p}_{zt})/\sigma_{zt}$ .

Consider figure 3 for an example of z-score calculations for the metropolitan area of Minneapolis, Minnesota. The underlying boundary lines denote census tracts. As is evident, the relatively most polluted part of the city is located over the most populated areas. The figure plots estimates for ozone z-scores for 2006 around locations with active monitors anywhere from 2005-2012. Monitored grid cells, relative to their surrounding zones tend to be nearer to the mean and on the boundaries of the city.

### 4.3 $NO_2$ and Ozone Overview

The AQS data include both  $NO_2$  and ozone ( $O_3$ ). Because ozone is not directly measured in the BEHR data and NASA OMI ozone data are reported at too low a resolution to capture the variation of pollution in localized areas, we propose a combination of BEHR and NASA’s OMI aerosol data to proxy for ozone. As such, a brief overview of ozone formation is useful in order to understand our empirical approach.

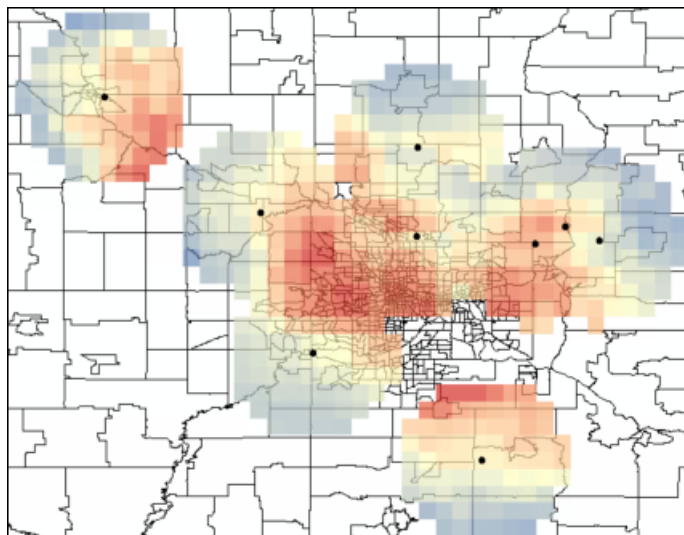
Ozone is not directly emitted, but tropospheric ozone is formed through a combination of

---

<sup>20</sup>Regressions detailing the correlation between BEHR and AQS data show  $R^2$  estimates of roughly 40% for New York City county, 30% for Cook county (Chicago) and 15% for the Greater Los Angeles Area all with statistically significant positive correlations.



Figure 3: Z-Score Map: Minneapolis (2006)



Notes: Black dots represent ozone monitors as part of the AQS. Monitors are encircled by zones which are used to calculate z-scores for ozone. Red grid cells are associated with high z-scores while blue cells are representative of low negative scores. Tan coloring represents z-scores of roughly zero (mean valued pollution cell). Z-scores are provided for 2006 around monitors that were active any time during 2005-2012.

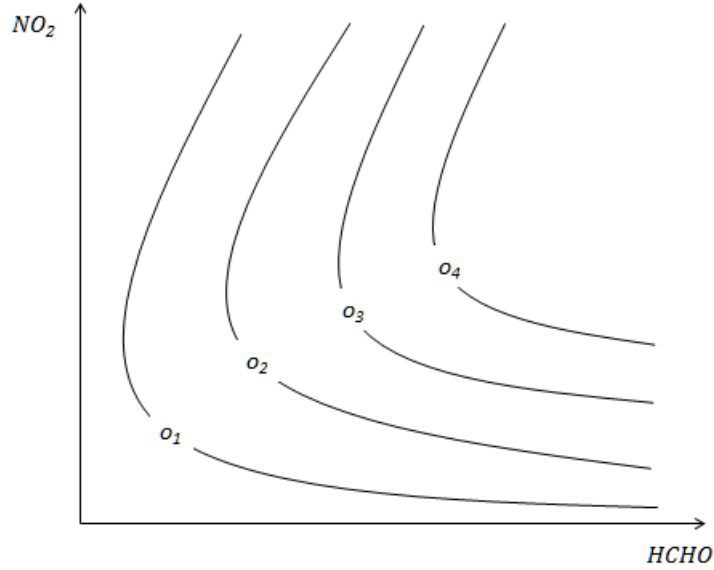
volatile organic compounds (VOCs) and nitrogen oxides ( $NO_x$ ) in the presence of sunlight, particularly ultraviolet light. The BEHR data include  $NO_2$  (a subset of nitrogen oxides), while NASA OMI data include formaldehyde ( $HCHO$ , a proxy for VOCs). The atmospheric science literature often uses these data to look at ozone formation.<sup>21</sup>

In general, more formaldehyde and more  $NO_x$  (plus sunlight) translate into higher levels of tropospheric ozone, but at low levels of either VOCs or  $NO_x$  (i.e. a VOC-limited environment, or a  $NO_x$ -limited environment) the relationship between ozone and its inputs is more complicated. An example of the level curves for ozone is given in figure 4.

We use information on  $NO_2$ ,  $HCHO$  and ozone measured by AQS monitors and satellites in our empirical section. As a rough proxy for ozone, in some specifications we take the product of  $HCHO * NO_2$ . In order to get a spatial sense on our proxy for ozone and other pollution estimates, consider figure 5. The top left figure represents BEHR data for  $NO_2$  column density estimates. On average, darker red portions of the map indicate higher levels

<sup>21</sup>For example, see [Jin and Holloway \(2015\)](#).

Figure 4: Ozone Formation Example



Notes: This illustrates the general relationship between  $NO_x$  and VOCs (proxied by  $NO_2$  and  $HCHO$ ). Moving out to the upper-right leads to higher levels of ozone, indicated by level curves  $O_1$  to  $O_4$ . Exact values and shapes of these curves are determined by local factors such as sunlight and atmospheric chemistry.

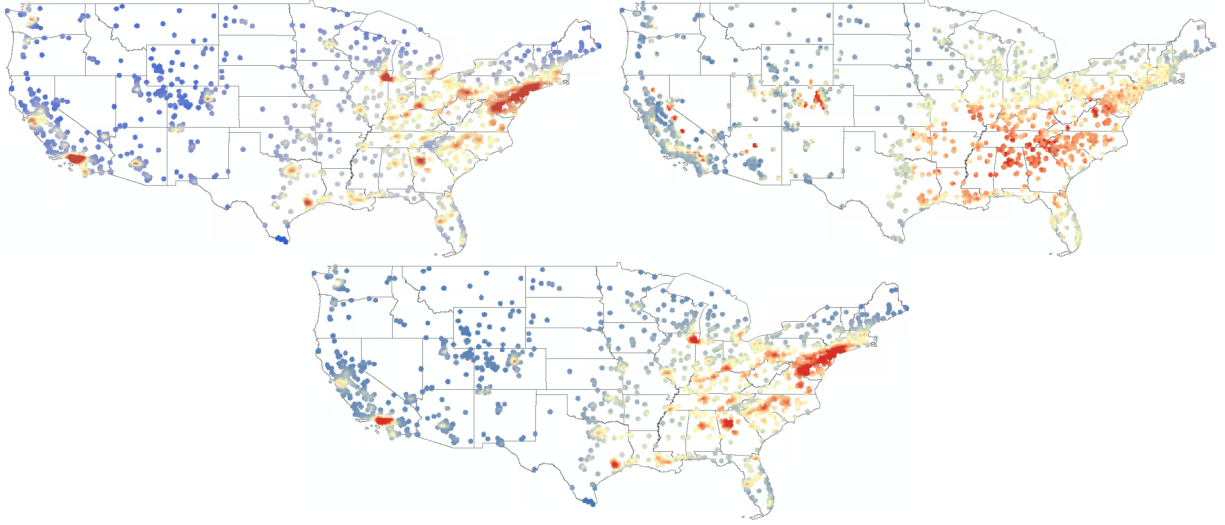
of BEHR pollution and conversely, darker blue portions represent lower levels of pollution. However, cardinal values of column densities cannot be directly compared across the country because they are effected by land use, altitude, temperature, etc. Highly polluted areas tend to be concentrated around urban areas like New York City, Los Angeles and Chicago. The top right map denotes  $HCHO$  estimates mapped into BEHR grid cells from NASA's OMI. The most heavily polluted areas for formaldehyde tends to be in the southeast. The combination of both in the form of an interaction term is displayed on the bottom map.

#### 4.4 Summary Statistics

There are several other data sources used in the empirical model. Other data can be characterized either as spatial data or variables used to described federal guideline incentives for regulators.

Due to federal guidelines, monitors should partly be placed to characterize vulnerable

Figure 5: Ozone Formation (2006)



Notes: This illustrates satellite pollution data and the formation of our ozone proxy for 2006. Only areas surrounding pollution monitors (i.e. a 25 km radius) are shown. The top left image concerns the distribution of  $NO_2$  pollution, the top right is for  $HCHO$  (formaldehyde) and the bottom is the interaction of both to form the proxy for ozone. Red indicates higher levels of pollution while blue represents low levels.

and susceptible populations. For this reason, we include a variety of controls from the American Community Survey (ACS)<sup>22</sup>. Notably, because our main regression specifications will effectively be cross-sectional, we are able to use these data even though these variables are time invariant. Variables from the ACS include the average poverty rate, median age, median rent, population levels, race, and percent small children (less than 5). We are also concerned with the number of point sources for emissions throughout our area of study. We choose the Toxic Release Inventory (TRI) data, which includes the number of polluting firms, each with a precise geocode to allow use of GIS to calculate the number of TRI firms within a given grid cell and year. In addition to polluting firms, we are interested in economic activity of a region. We use nighttime data from NOAA, which includes the amount of ambient light as measured from satellites. We include nonroad light as a proxy for economic activity (i.e. Doll et al. (2006)) at the grid cell level. Roads were excluded in GIS using

---

<sup>22</sup>We use the 5 year estimate tables for 2006-2010 at the Census tract level. We map these data to our zonal grid cells by averaging over tracts that fall in the same grid cell.

a shapefile from the National Highway Planning Network (NHPN). The highway shapefile contains information on the National Highway System, the Eisenhower Interstate System, the Strategic Highway Network, and National Highway System Intermodal Connectors.

We also include other spatial information from GIS shapefiles including County borders, State borders, Air District shapes, the amount of water in a grid cell, and whether a particular area is classified as Urban or Tribal. Details are available from the authors.

Table 2: Summary Statistics

<i>Variable</i>	<b>All</b>			<b>New Sitings</b>		
	<i>Obs.</i>	<i>Mean</i>	<i>SD</i>	<i>Obs.</i>	<i>Mean</i>	<i>SD</i>
Monitor	995,606	0.03	0.17	30,447	0.03	0.16
BEHR $NO_2$	992,004	4.15	2.76	30,289	3.05	2.11
HCHO OMI	995,606	14.21	5.00	30,447	14.84	4.63
$NO_2$ *HCHO	992,004	59.22	41.00	30,289	46.23	35.79
AQS 8 Hour	31,331	0.04	0.01	808	0.04	0.01
AQS 4th Max	31,331	0.07	0.01	808	0.07	0.01
Nonroad Light	995,606	17.59	17.15	30,447	12.30	15.36
TRI Facilities	995,606	1.60	8.51	30,447	0.88	6.20
Median Age	995,606	37.59	10.15	30,447	38.72	8.86
Total Population	995,606	4228.28	1892.05	30,447	4148.49	1805.10
Median Rent	995,606	779.15	338.49	30,447	746.23	313.02
Poverty Rate	952,288	0.12	0.07	29,698	0.12	0.07
Percent Children ( $\leq 5$ )	952,296	0.08	0.02	29,699	0.08	0.02
Percent White	952,296	0.83	0.18	29,699	0.85	0.16
Percent Black	952,296	0.07	0.14	29,699	0.05	0.11

Consider table 2 for a table of summary statistics. Statistics are computed for both the entire dataset (at the 25km buffer around each monitor) as well as for just zones with new monitor sitings. Monitor represents a dummy variable for if a monitor is active within a grid cell for a given year over the course of our time window. With the entire dataset and considering zones with newly sited monitors, roughly 3% of observations contain an ambient pollution monitor. The next three variables are the satellite pollution estimates ( $NO_2$ , formaldehyde and our ozone proxy). Such estimates are scaled to  $10^{15}$  molecules/ $cm^2$  which as detailed above, is not comparable to standard parts per billion/million units. However, they do show that there exists variation in the data which we rely on when studying within-

zone variation of pollution. The table also provides summary statistics on the AQS data for an annual 8-hour mean and the annual 4th largest pollution reading for ozone which shows that the average 4th largest reading for ozone is at the NAAQS.

The next series of data represent our additional control variables. Our non-road light variable ranges from 0 to 63 where a higher number reflects more light on average in a grid cell. The Appendix shows a visual example of the light data for the Chicago region. Table 2 shows that the average levels of light not attributed to traffic is smaller for newly sited monitors. The same trend holds for our TRI variables which describes the number of polluting firms in a given grid cell. Summary statistics across samples for the ACS variables are roughly comparable.

## 5 Empirical Evidence

Following the analytical model's predictions, we want to test the hypothesis that regulators in "marginal" counties act strategically in choosing the location of new ambient pollution monitors. Recall that the prediction from the analytical model is that a regulator in a marginal county would place monitors in relatively "clean" areas in order to avoid nonattainment designation. On the other hand, the regulator in a nonattainment county would site a new monitor in a relatively "dirty" area, targeting polluted areas. In order to do so, we first need to define empirically what constitutes a "nonattainment", "marginal", or "clean" area. We then begin with descriptive statistics and preliminary evidence before proceeding to the econometric specifications.

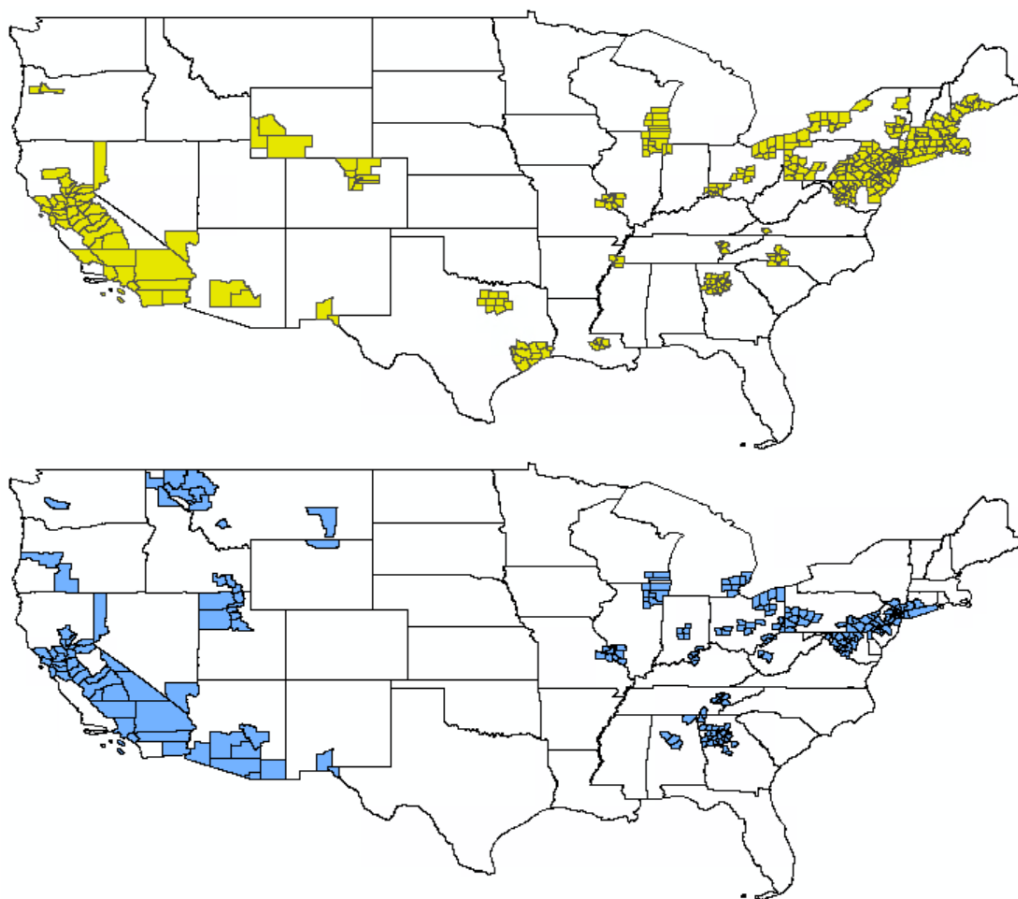
The nonattainment designation occurs at the county-year level when a primary or secondary standard is violated for one of the criteria pollutants. We use the "Greenbook" from EPA, which provides data on the year(s) of violation by pollutant for each county in the United States. For our purposes, we are primarily interested in ozone, because  $NO_2$  violations are quite rare.<sup>23</sup> Because  $NO_x$  is a precursor to ozone, we will rely primarily on ozone nonattainment, but we also consider violations of the standards for particulate matter ( $PM_{10}$  and  $PM_{2.5}$ ) in alternative specifications. To help visualize which counties are designated as

---

<sup>23</sup>There were four cases in the data, all of which occurred long in the past.

nonattainment for ozone and particulates, we present two maps for the most recent year in our data (2012) in figure 6.<sup>24</sup>

Figure 6: Non-Attainment Maps (2012): Ozone and Particulate Matter



Notes: The top map represents nonattainment and partial nonattainment counties for ozone and the bottom for particulate matter, both for the year 2012. We have excluded showing nonattainment maps for  $NO_2$  because very few counties have ever been characterized as nonattainment for this pollutant.

To characterize "clean" counties, we attempt to approximate the rule used by EPA in defining a clean monitor. Roughly stated, under the Code of Federal Regulations, a monitor

---

<sup>24</sup>We note that our specifications take into account contemporaneous nonattainment designation, so the maps presented here are only for one cross-section. There is some change from year-to-year in nonattainment status, which we consider in our empirical models. In specifications not shown here, we also consider counties that have "recently" (e.g. in the past five years) been in nonattainment for any pollutant. The qualitative results are unchanged by alternative definitions of nonattainment. See the online appendix for details.

may only be removed if it is clean and the probability of nonattainment in the near future is very low. We use the AQS pollution monitor readings for the primary and secondary standards to empirically define which monitors are "clean" in any given year. Because a monitor could be placed in a relatively clean location in a dirty area, we also consider the standard deviation of the remote sensing estimates for the zone where a monitor is located. We then define a "clean" monitor as having both a low monitoring value for the primary and secondary standards and a low standard deviation for that region.<sup>25</sup> For ozone, the primary and secondary standards state that the annual fourth-highest daily maximum 8-hour concentration, averaged over 3 years, must not exceed 0.07 parts per million. We take half of the NAAQS (i.e. 0.035 ppm) as the cutoff for a "clean" area if the standard deviation for our remote sensing proxy for ozone (i.e.  $HCHO * NO_2$ ) is below the fifth percentile. This threshold for defining "clean" is somewhat arbitrary, so we experiment with alternative definitions. Overall, few monitors are defined as clean, and our results are very robust to our definition.

"Marginal" areas are then defined as areas that are neither in nonattainment nor "clean" according to our definition above. Consider table 3 for the breakdown of different types. The left side of the table reports the proportions of the total data sample that is characterized as either in nonattainment, "marginal" or "clean". In the full sample, roughly 45% of counties are characterized as nonattainment, only 2% as clean and 53% as marginal. These proportions change if considering grid cells around only newly-sited monitors.

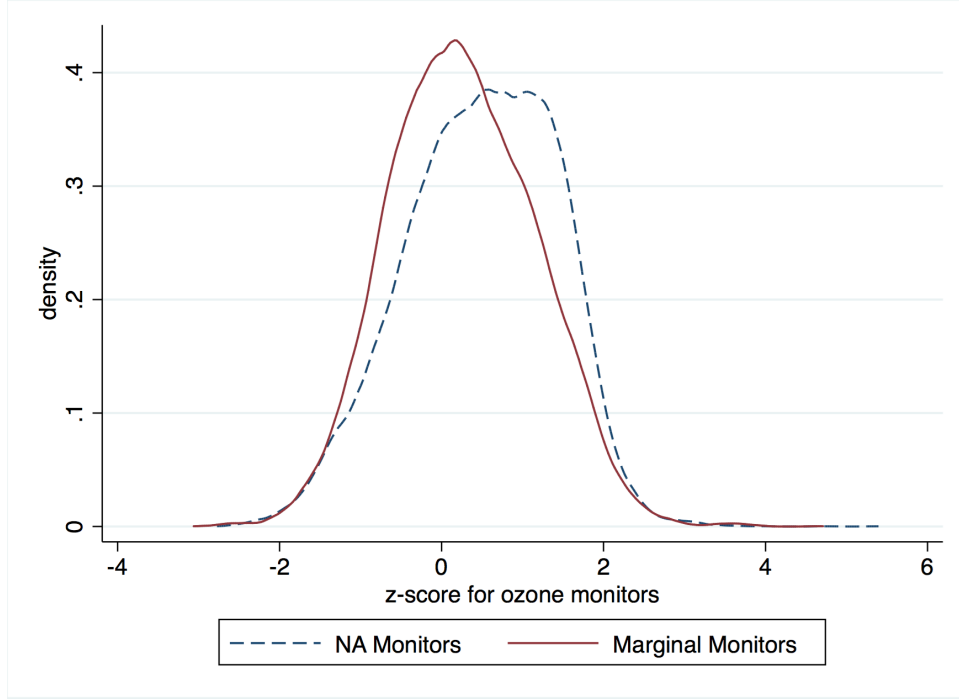
Table 3: County Types

<i>Variable</i>	<b>All</b>			<b>New Sitings</b>		
	<i>Obs.</i>	<i>Mean</i>	<i>SD</i>	<i>Obs.</i>	<i>Mean</i>	<i>SD</i>
Non-attainment	995,606	0.45	0.50	30,447	0.29	0.45
Clean	995,606	0.02	0.12	30,447	0.03	0.17
Marginal	995,606	0.53	0.50	30,447	0.68	0.47

---

<sup>25</sup>This definition is conservative, and as a result we will have very few monitors designated as "clean." Under alternative definitions of "clean" there would be more clean and fewer marginal monitors, so our current definition should create a bias against finding an effect for marginal cases. We also note that alternative definitions of "clean" may exclude the use of the remote sensing standard deviation, as that information may not be readily available to the regulator.

Figure 7: Distribution of Pollution Z-Scores by Monitor Type



Notes: Kernel density plots are shown for the z-scores for ozone (or more accurately, our ozone proxy,  $HCHO * NO_2$ ) for all monitors in our dataset from 2005-2012. "Marginal" monitors are defined as those not in nonattainment counties and not meeting our definition of a "clean" monitor as defined in the text.

## 5.1 Comparing Nonattainment and Marginal Areas

Because few monitors are defined as clean, and because the analytical predictions regarding clean counties are ambiguous, we focus on the empirical differences between the pollution distributions in areas already in nonattainment versus areas that we define as marginal. To begin with, we have just under one million observations using the 25 km-radius data, where an observation is a grid cell-by-year. We focus on the z-scores for our ozone proxy ( $HCHO * NO_2$ , as discussed in the data section) at the monitor level to test for differences between marginal and nonattainment counties. Recall that z-scores are calculated using the data within the 25 km radius (or 10 km in alternative specifications). Therefore a lower z-score (i.e. more negative) represents pollution avoidance whereas a larger z-score represents targeting the most polluted grid cell in a zone. We first plot the kernel densities of z-scores for marginal and nonattainment counties using all years (2005-2012).



As shown in figure 7, the distribution of z-scores for monitors in marginal counties lies to the left of the distribution of z-scores for monitors in nonattainment counties. The mean z-scores for nonattainment counties are 0.483 for ozone and 0.545 for  $NO_2$ ; for marginal counties, the mean z-scores are 0.263 and 0.366 for ozone and  $NO_2$ , respectively. The difference between nonattainment and marginal counties is significant in both cases at the 0.1% level.<sup>26</sup>

In what follows, we focus on *new* monitor sitings to test for differences in regulator behavior. For that, we introduce a simple econometric model.

## 5.2 New Monitor Sitings

How does the local regulator (i.e. state, substate or tribal) choose a monitor location, and is there evidence that regulators in marginal counties target cleaner areas than nonattainment counties? Over our period of study, there are 379 new AQS monitors for ozone that were sited<sup>27</sup>, and we rely on these sitings for our hypothesis testing. As discussed in the data section, the monitors were placed relatively uniformly over our time of study, 2006-2012. For each monitor, we use the z-score in the year it was sited as the main independent variable, and we test the null hypothesis that marginal and nonattainment monitors are sited based on similar criteria.

Before continuing into the econometric specification, consider figure 8 which provides a map of all county level new monitor siting decisions relative to their surrounding areas. Lined counties in the map denote nonattainment counties at time of new monitor sitings. Using our localized z-scores, we calculate average county level values for counties with new monitor sitings from 2006-2012. Blue denotes positive z-scores while red denotes negative z-scores (where darker shading denotes larger magnitudes in absolute value). The bottom three images are three representative urban areas (Detroit, Houston and Charlotte, from left to right). As is evident in these urban areas, counties in nonattainment at the time of siting placed monitors in areas with relatively high pollution levels than the surrounding

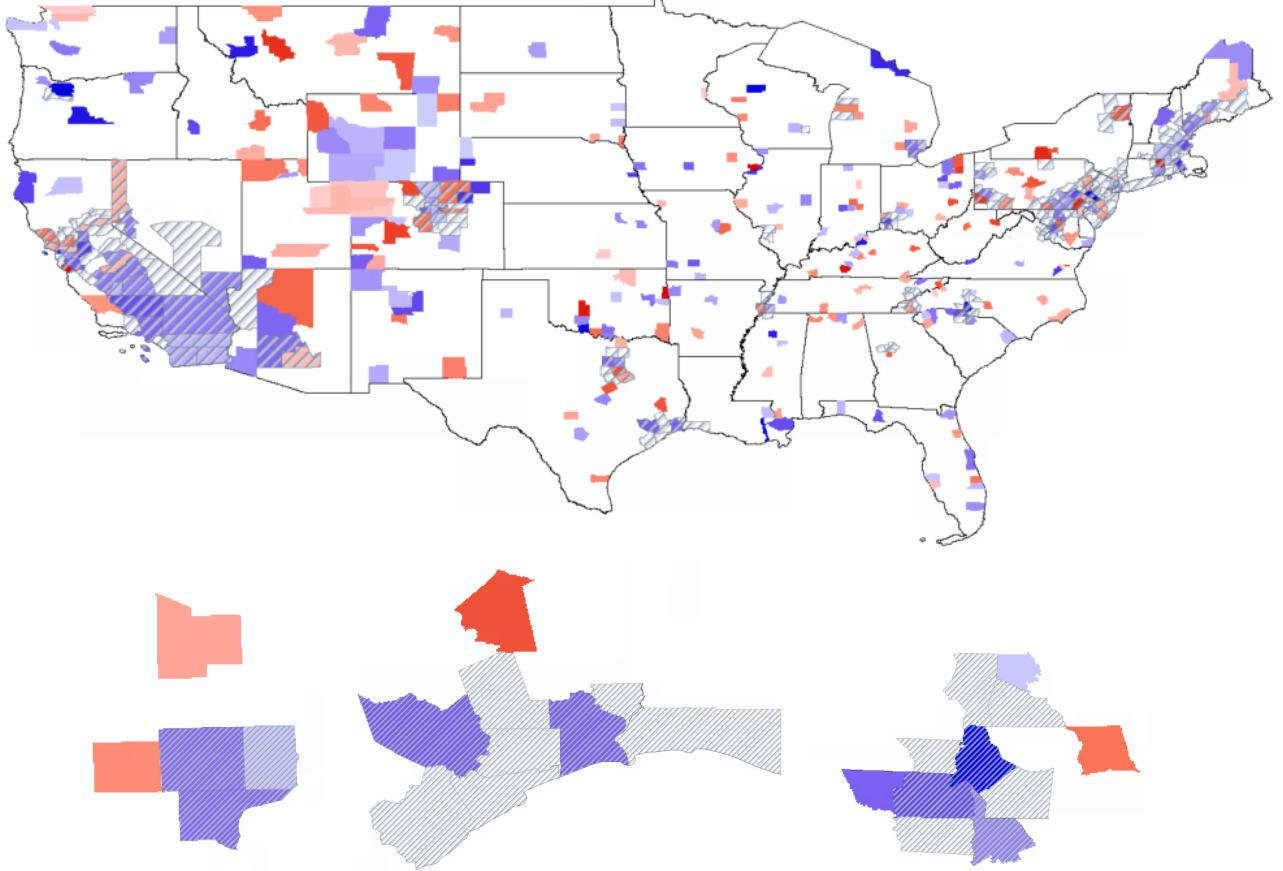
---

<sup>26</sup>For clean counties, the average monitor z-scores are 0.464 and 0.473 for ozone and  $NO_2$ , respectively.

<sup>27</sup>This differs slightly from the number of actual new monitors because we observe new "monitors" here as grid cells that were previously unmonitored.

area, while nearby counties not facing nonattainment status placed monitors in relatively cleaner areas.

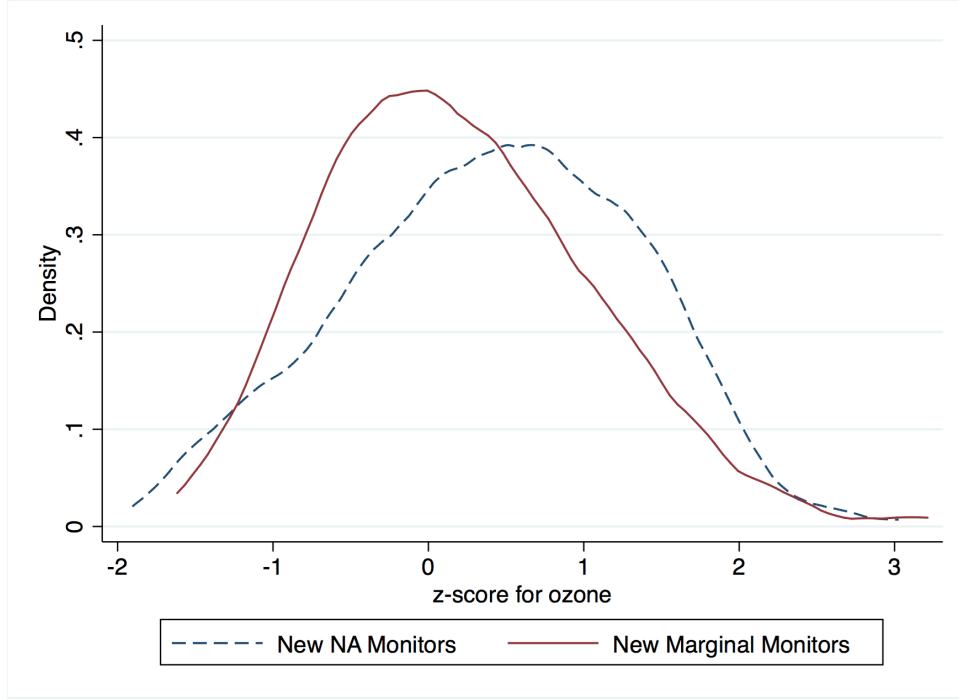
Figure 8: New Monitor Positioning



Notes: The top figure describes the average z-score of a county when siting a monitor. Blue shading denotes a positive z-score (darker being more positive) and red denotes a negative z-score (darker being more negative). The lined shading represents nonattainment counties in our collapsed new monitor data set. The bottom three figures are zoomed in versions of Detroit, Houston and Charlotte.

To test whether the z-scores for new monitors are significantly different in marginal counties, we simply restrict the sample to the new monitors in our dataset and regress the z-score for our ozone proxy on an indicator for marginal. The restricted sample includes all new monitor sitings defined above and all associated grid cells in the same zone at the time of siting. The coefficient is -0.231 and is significant at the 0.1% level, indicating that the average z-score for new monitors in marginal counties is smaller than for new monitors in

Figure 9: Distribution of Z-Scores by County Type for New Monitors



Notes: The sample is restricted to newly-sited monitors in the dataset. Kernel density plots are shown for the z-scores for ozone (or more accurately, our ozone proxy,  $HCHO * NO_2$ ) for all monitors in our dataset from 2005-2012. "Marginal" monitors are defined as those not in nonattainment counties and not meeting our definition of a "clean" monitor as defined in the text.

other counties. A kernel density plot of the z-scores for new marginal monitors versus new nonattainment monitors reveals a similar trend in figure 9.

As shown in figure 9, the distribution of new marginal monitors lies to the left of the distribution for new nonattainment monitors, suggesting that regulator behavior is indeed different at the time of siting. To formalize this observation, we turn to an econometric model of the probability of choosing any given grid cell for a monitor siting. Here we assume that the relevant area for comparison is the surrounding "zone"; that is, we assume that the regulator would place a monitor in the zone defined by the actual placement of the monitor. We then test whether the probability of a monitor siting varies with county type, conditional on relevant covariates.

First we define  $Monitor_{izt}$  as taking a value of 1 if a monitor is placed in grid  $i$ , which

is in zone  $z$ , in year  $t$ ; otherwise it takes a value of zero. We then define *Marginal* and *Clean* as before, leaving the excluded category as nonattainment monitors. The econometric specification is given by

$$\begin{aligned} Monitor_{izt} = & \alpha_0 + \beta_1 * zscore_{izt}^{FN} + \beta_2 * zscore_{izt}^{FN} * Marginal_{zt} \\ & + \beta_3 * zscore_{izt}^{FN} * Clean_{zt} + \delta_{marginal} + \eta_{clean} + \sum_j \theta_j X_j + \epsilon_{izt}, \end{aligned}$$

where  $zscore^{FN}$  is the z-score for our proxy for ozone,  $HCHO * NO_2$  as described in the data section,  $X_j$  represents other control variables, and  $\epsilon_{izt}$  is a random disturbance term. In alternative specifications we add the number of *Toxic Release Inventory* (TRI) facilities present in grid  $i$ , the poverty rate in grid  $i$  and non-road, nighttime light as measured from space (another proxy for economic activity).

This specification effectively collapses the data into a cross-section of zones that received a new monitor. Because the comparisons are all being made *within a zone*, we are able to estimate how the probability of a monitor siting is affected by *relative* changes in pollution. That is, because we are calculating z-scores at the grid cell level within a specified zone, we can determine how regulator siting behavior within a zone is affected by how relatively dirty or clean a given location is. The results are in Tables 4 and 5, which show estimates for the linear probability model and probit specifications with a 25 km radius, respectively.

The results for our ozone regressions fully support our analytical predictions, namely that newly-sited monitors in marginal counties have lower z-scores for pollution, on average, than do newly-sited monitors in nonattainment counties. The negative and significant coefficient on the interaction term between the z-score and *Marginal* indicates avoidance behavior—local regulators are placing monitors in less polluted areas if the county has not already been designated as being in nonattainment.

In specification (4) in Table 4, for example, a one standard deviation increase in the z-score in a marginal county reduces the probability of siting a new monitor by almost one percentage point. That is, the less polluted a site is in a marginal county, the higher the probability of a new monitor siting. The probit results for ozone are similarly signed, but the magnitudes of the marginal effects are slightly smaller in specification (4).

The control variables are also generally significant and take the expected sign. For ex-

Table 4: Linear Probability Model for Monitor Placement

	(1)	(2)	(3)	(4)
$zscore^{ozone}$	0.0183*** (0.00366)	0.0113*** (0.00305)	0.0169*** (0.00395)	0.0121*** (0.00312)
$zscore^{ozone}$ *Marginal	-0.0147*** (0.00380)	-0.0110*** (0.00327)	-0.0136*** (0.00408)	-0.0110*** (0.00333)
Marginal	-0.0210*** (0.00402)	-0.00197 (0.00359)	-0.0148* (0.00784)	-0.000867 (0.00570)
$zscore^{ozone}$ *Clean	-0.00188 (0.0100)	0.00294 (0.00890)	-0.00725 (0.00731)	-0.00207 (0.00686)
Clean	-0.00964 (0.0112)	0.00707 (0.00870)	-0.0633*** (0.0153)	-0.0439*** (0.0117)
Ozone Proxy			0.000143* (8.57e-05)	-0.000135* (7.87e-05)
Marginal*Ozone			-7.18e-05 (9.23e-05)	-8.86e-05 (8.35e-05)
Clean*Ozone			0.00398*** (0.00113)	0.00302*** (0.000871)
TRI Facilities		0.000706* (0.000385)		0.000751* (0.000404)
Nonroad Light		0.00152*** (0.000195)		0.00163*** (0.000204)
Median Age		-0.000401* (0.000222)		-0.000456** (0.000217)
Total Population		-1.56e-06** (7.06e-07)		-1.36e-06** (6.76e-07)
Median Rent		6.91e-06 (6.76e-06)		5.52e-06 (6.68e-06)
Poverty Rate		0.124*** (0.0227)		0.127*** (0.0232)
Percent Children ( $\leq 5$ )		-0.172*** (0.0567)		-0.189*** (0.0577)
Percent White		-0.0376*** (0.0135)		-0.0269** (0.0131)
Percent Black		-0.0766*** (0.0200)		-0.0484** (0.0196)
Constant	0.0409*** (0.00391)	0.0611*** (0.0183)	0.0319*** (0.00768)	0.0607*** (0.0182)
Observations	30,289	29,561	30,289	29,561

Notes: The dependent variable is a new monitor siting. There are 379 new monitor sitings over the years 2006-2012. *Marginal* and *Clean* are defined in the text; the excluded category is a Nonattainment county. The ozone proxy ( $HCHO * NO_2$ ) is also described in the data section. The z-scores are calculated for each grid cell in a 25 km circle from the newly-sited monitor. Standard errors are clustered at the county level. \*\*\*, \*\* and \* denote significance at the 1%, 5% and 10% levels, respectively.

Table 5: Probit Marginal Effects for Monitor Placement

	(1)	(2)	(3)	(4)
$zscore^{ozone}$	0.01177*** (0.00196)	0.00575*** (0.00144)	0.01079*** (0.00216)	0.00612*** (0.00144)
$zscore^{ozone}$ *Marginal	-0.00734*** (0.00221)	-0.00505*** (0.00175)	-0.00683*** (0.00241)	-0.00501*** (0.00176)
Marginal	-0.01792*** (0.00348)	-0.00089 (0.00252)	-0.01551** (0.00622)	0.00050 (0.00371)
$zscore^{ozone}$ *Clean	-0.00685 (0.00642)	0.00347 (0.00561)	-0.02350*** (0.00188)	-0.01377*** (0.00257)
Clean	0.00162 (0.00566)	0.00280 (0.00362)	-0.00267 (0.00475)	0.00024 (0.00316)
Ozone Proxy			0.00008* (0.00004)	-0.00006* (0.00003)
Marginal*Ozone			0.00000 (0.00006)	-0.00006 (0.00005)
Clean*Ozone			0.00264*** (0.00046)	0.00118*** (0.00032)
TRI Facilities		0.00016* (0.00010)		0.00019* (0.00011)
Nonroad Light		0.00091*** (0.00007)		0.00094*** (0.00007)
Median Age		-0.00033* (0.00018)		-0.00035** (0.00018)
Total Population		-0.00000 (0.00000)		-0.00000 (0.00000)
Median Rent		0.00001 (0.00001)		0.00001 (0.00001)
Poverty Rate		0.08503*** (0.01359)		0.08413*** (0.01386)
Percent Children ( $\leq 5$ )		-0.12126*** (0.04459)		-0.12785*** (0.04462)
Percent White		-0.02202*** (0.00776)		-0.01833** (0.00787)
Percent Black		-0.04678*** (0.01054)		-0.03506*** (0.01076)
Observations	30289	29561	30289	29561

Notes: Marginal effects from a probit estimation are shown. The dependent variable is a new monitor siting. There are 379 new monitor sitings over the years 2006-2012. *Marginal* and *Clean* are defined in the text; the excluded category is a Nonattainment county. The ozone proxy ( $HCHO * NO_2$ ) is also described in the data section. The z-scores are calculated for each grid cell in a 25 km circle from the newly-sited monitor. Standard errors are clustered at the county level. \*\*\*, \*\* and \* denote significance at the 1%, 5% and 10% levels, respectively.

ample, monitors are placed in areas with higher-than-average poverty rates, more polluting sources (TRI facilities), and more economic activity.<sup>28</sup>

As a robustness check, we estimate the same equations, substituting the z-score for ozone with a z-score for  $NO_2$ , as measured in the BEHR remote sensing data. These results are found in Table 6 for the linear probability model, and in Table 7 for the probit estimates. Again, the pattern clearly emerges where newly-sited monitors in marginal counties are placed in cleaner areas, on average, when compared to the area surrounding that monitor. Results are robust across specifications. As we saw with our ozone regressions, the "dirtier" a grid cell is in a marginal county, the lower the probability of a new monitor siting.

### 5.3 Pollution and Equity Implications

We have shown that monitors in marginal areas tend to have lower z-scores for ozone and  $NO_2$ , which means that, compared to the surrounding zone, the pollution level at the monitor is relatively low. Of the counties that we classify as "marginal", we find 11 observations where the monitor's annual 8-hour average for ozone is at least 5 ppb (the NAAQS primary and secondary standards are 7 ppb), and where the z-score is less than -1.96. If we relax the criterion for z-scores to be less than -1.645, we find 37 observations where the ozone reading at the monitor is greater than 5 ppb. These observations can be seen in figure 10. Note that because many observations from the same county in multiple years, we relax our cutoff to -1.28 (10% level for a one tailed test) for new monitor sitings as seen in red. Our more stringent definition displays counties in yellow. Though we cannot directly measure pollution in the surrounding region (because the units of measurement are different between the satellite and monitoring data), the negative and significant z-scores for these monitors suggests that many of these counties would be in nonattainment if the monitor were placed elsewhere in the surrounding region.

---

<sup>28</sup>In specifications not shown here, we also investigated whether the political environment (as measured by the 2008 election results) affected the siting of monitors in the previous regression results. Conditional on the other covariates, Democrat was insignificant across specifications, as were interaction terms. It is possible that we are not capturing the political preferences of the local regulator with such a measure, but it suggests that local political majorities do not affect the siting decision as much as other factors.

Table 6: Linear Probability Model for Monitor Placement ( $NO_2$ )

	(1)	(2)	(3)	(4)
$zscore^{NO_2}$	0.0286*** (0.00560)	0.0162*** (0.00452)	0.0253*** (0.00601)	0.0158*** (0.00444)
$zscore^{NO_2} * \text{Marginal}$	-0.0220*** (0.00577)	-0.0145*** (0.00476)	-0.0191*** (0.00617)	-0.0134*** (0.00467)
$\text{Marginal}$	-0.0221*** (0.00394)	-0.00355 (0.00372)	-0.0108 (0.00752)	0.00618 (0.00587)
$zscore^{NO_2} * \text{Clean}$	-0.00800 (0.0154)	0.000560 (0.0138)	-0.0150 (0.0111)	-0.00592 (0.0104)
$\text{Clean}$	-0.0114 (0.0104)	0.00498 (0.00808)	-0.0927*** (0.0258)	-0.0625*** (0.0218)
$NO_2$			0.00371*** (0.00128)	-0.000638 (0.00124)
$\text{Marginal} * NO_2$			-0.00180 (0.00146)	-0.00402*** (0.00147)
$\text{Clean} * NO_2$			0.0712*** (0.0194)	0.0514*** (0.0164)
TRI Facilities		0.000694* (0.000378)		0.000707* (0.000399)
Nonroad Light		0.00147*** (0.000185)		0.00157*** (0.000192)
Median Age		-0.000338 (0.000230)		-0.000394* (0.000224)
Total Population		-1.43e-06** (6.99e-07)		-1.23e-06* (6.83e-07)
Median Rent		6.23e-06 (6.68e-06)		5.22e-06 (6.71e-06)
Poverty Rate		0.120*** (0.0221)		0.128*** (0.0224)
Percent Children ( $\leq 5$ )		-0.175*** (0.0572)		-0.198*** (0.0581)
Percent White		-0.0358*** (0.0133)		-0.0276** (0.0129)
Percent Black		-0.0763*** (0.0198)		-0.0546*** (0.0201)
Constant	0.0421*** (0.00383)	0.0598*** (0.0186)	0.0259*** (0.00729)	0.0559*** (0.0178)
Observations	30,287	29,559	30,287	29,559

Notes: The dependent variable is a new monitor siting. There are 379 new monitor sitings over the years 2006-2012. *Marginal* and *Clean* are defined in the text; the excluded category is a Nonattainment county.  $NO_2$  is estimated remotely as described in the data section. z-scores are calculated for each grid cell in a 25 km circle from the newly-sited monitor. Standard errors are clustered at the county level. \*\*\*, \*\* and \* denote significance at the 1%, 5% and 10% levels, respectively.

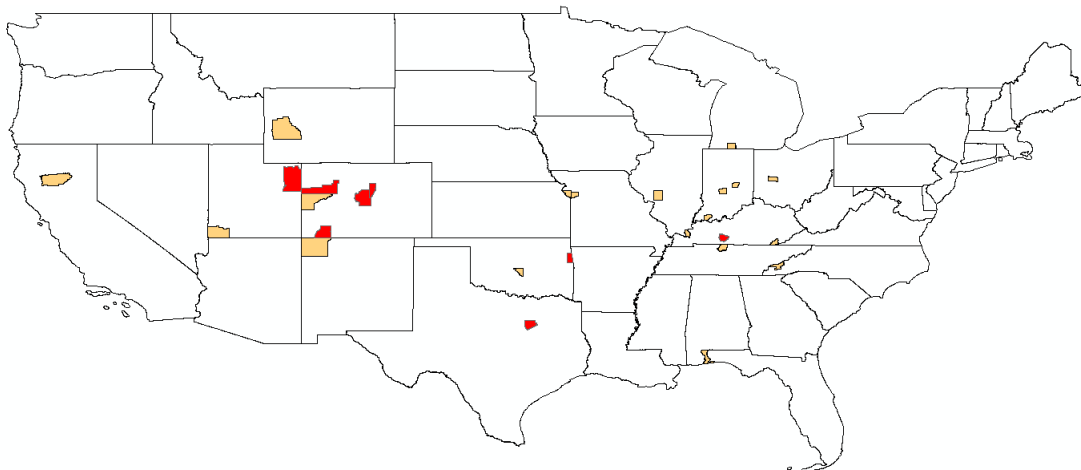


Table 7: Probit Marginal Effects for Monitor Placement ( $NO_2$ )

	(1)	(2)	(3)	(4)
$zscore^{NO_2}$	0.01846*** (0.00282)	0.00850*** (0.00207)	0.01642*** (0.00315)	0.00833*** (0.00200)
$zscore^{NO_2} * \text{Marginal}$	-0.01083*** (0.00309)	-0.00589** (0.00231)	-0.00937*** (0.00340)	-0.00533** (0.00226)
Marginal	-0.01698*** (0.00307)	-0.00140 (0.00254)	-0.01326** (0.00540)	0.00379 (0.00361)
$zscore^{NO_2} * \text{Clean}$	-0.00097* (0.00913)	0.00186 (0.00617)	-0.00777 (0.00726)	-0.00259 (0.00509)
Clean	-0.00775 (0.00421)	0.00197 (0.00426)	-0.02480*** (0.00151)	-0.01722*** (0.00264)
$NO_2$			0.00180*** (0.00060)	-0.00030 (0.00046)
Marginal* $NO_2$			0.00024 (0.00093)	-0.00205** (0.00085)
Clean* $NO_2$			0.04820*** (0.00906)	0.02257*** (0.00646)
TRI Facilities		0.00016* (0.00009)		0.00017* (0.00010)
Nonroad Light		0.00087*** (0.00007)		0.00090*** (0.00007)
Median Age		-0.00029 (0.00018)		-0.00032* (0.00018)
Total Population		-0.00000 (0.00000)		-0.00000 (0.00000)
Median Rent		0.00001 (0.00001)		0.00001 (0.00001)
Poverty Rate		0.08156*** (0.01310)		0.08301*** (0.01317)
Percent Children ( $\leq 5$ )		-0.11822** (0.04567)		-0.12839*** (0.04561)
Percent White		-0.02071*** (0.00790)		-0.01786** (0.00785)
Percent Black		-0.04682*** (0.01062)		-0.03823*** (0.01099)
Observations	30287	29559	30287	29559

Notes: Marginal effects from probit estimations are shown. The dependent variable is a new monitor siting. There are 379 new monitor sitings over the years 2006-2012. *Marginal* and *Clean* are defined in the text; the excluded category is a Nonattainment county.  $NO_2$  is estimated remotely as described in the data section. z-scores are calculated for each grid cell in a 25 km circle from the newly-sited monitor. Standard errors are clustered at the county level. \*\*\*, \*\* and \* denote significance at the 1%, 5% and 10% levels, respectively.

Figure 10: Strategic Marginal Counties



Notes: Red counties represent marginal counties with pollution levels greater than 5 *ppb* and a z-score of less than -1.28 (10% level for a one tailed test) at the time of siting. Yellow counties are marginal counties with pollution levels greater than 5 *ppb* and z-scores less than -1.645 any time during the monitor's existence.

Overall this strongly suggests that there are dirty areas in some counties that are not being subjected to further regulatory scrutiny because these counties remain in attainment. Who, then, is being harmed by strategic monitor siting? To address this question, we note that a simple regression of the natural log of the poverty rate on  $zscore_{ozone}$  yields a coefficient of 0.026, with a t-statistic of roughly 47. This means that an increase in our ozone proxy of one standard deviation is associated with an increase in the poverty rate by roughly 2.6%. Therefore, as documented throughout the environmental justice literature, poorer areas tend to be more polluted. In this setting, not reducing pollution in the dirtiest areas could be harming the poorest members of society.<sup>29</sup>

---

<sup>29</sup>As discussed in Grainger (2012), who wins and loses from a targeted pollution reduction depends on many factors, including the extent to which rents respond to the air quality improvement. If the capitalization of the amenity is fully passed forward, low-income renters may pay for the air quality improvement while the landowners benefit.

## 6 Conclusion

Under the Clean Air Act and its Amendments, states must comply with federal National Ambient Air Quality Standards, but states (or tribal, or other local regulators) are also charged with developing and maintaining their own monitoring network that is used for compliance with the federal standards. We begin with an analytical model that shows how this creates an incentive for strategically siting new pollution monitors, depending on whether the county is in attainment or nonattainment with the NAAQS. Most importantly, a regulator in a "marginal" area has an incentive to avoid pollution when siting monitors, whereas a regulator for a nonattainment area may want to target pollution in her siting decision.

To test the theory, we use a novel approach that utilizes both monitoring data and remote sensing estimates of local air pollution. Focusing on new monitor sitings, we calculate a local z-score for pollution (either  $NO_2$  or a proxy for ozone) that compares pollution in the grid cell where a new monitor is placed to pollution in the surrounding region. We show that new monitors in "marginal" areas are placed in relatively clean areas, whereas new monitors in nonattainment areas tend to target pollution.

Our findings have several practical implications worth further discussion. First, existing estimates of air pollution may be systematically biased because of the incentive structure facing local regulators. Taking this further, there are likely counties that are not in compliance with federal air quality standards, but because of strategic siting decisions, they remain in attainment.

There are also implications for applied econometrics. Because regulators can strategically site air quality monitors, the use of Nonattainment status as an instrumental variable (or the NAAQS threshold in a regression discontinuity setting) may be invalid. This suggests that economists studying pollution should exercise caution in interpreting findings based on nonattainment status.

Finally, since nonattainment status is used to trigger additional regulatory pressure in order to force a pollution reduction, the mis-siting of monitors is likely to disadvantage those populations residing in hotspots. It is well documented that these populations are

disproportionately poor, which is further supported by our data. This suggests that there are some marginal counties that would experience large pollution reductions if only the monitor were sited differently, and this reduction would disproportionately affect low-income groups.

What can be done to improve the siting of monitors? Assuming states will maintain their own monitoring programs, the EPA could mandate that remote sensing estimates of pollution be used in guiding which sites are chosen by local authorities. Rather than complete autonomy in the siting decision, remote sensing estimates of local pollution could be used to determine a subset of the region where a monitor ought to be placed. This would remove some of the ability of local regulators to strategically site monitors, and it would improve the monitoring network's capacity to detect pollution and protect human health.

## References

- Auffhammer, M., A. M. Bento, and S. E. Lowe (2009). Measuring the Effects of the Clean Air Act Amendments on Ambient PM10 concentrations: The Critical Importance of a Spatially Disaggregated Analysis. *Journal of Environmental Economics and Management* 58(1), 15–26.
- Bechle, M. J., D. B. Millet, and J. D. Marshall (2013). Remote Sensing of Exposure to NO2: Satellite Versus Ground-Based Measurement in a Large Urban Area. *Atmospheric Environment* 69, 345–353.
- Becker, R. and V. Henderson (2000). Effects of Air Quality Regulations on Polluting Industries. *Journal of Political Economy* 108(2), 379–421.
- Becker, R. A. (2005). Air Pollution Abatement Costs Under the Clean Air Act: Evidence from the PACE Survey. *Journal of Environmental Economics and Management* 50(1), 144–169.
- Bento, A., M. Freedman, and C. Lang (2015). Who Benefits from Environmental Regulation? Evidence from the Clean Air Act Amendments. *Review of Economics and Statistics* 97(3), 610–622.
- Chang, W. and C. Grainger (2016). The Optimal Siting of Ambient Pollution Monitors. *Working Paper, University of Wisconsin - Madison*.
- Chay, K. Y. and M. Greenstone (2003). Air Quality, Infant Mortality, and the Clean Air Act of 1970. *National Bureau of Economic Research Working Paper* 10053.
- Chay, K. Y. and M. Greenstone (2005). Does Air Quality Matter? Evidence from the Housing Market. *Journal of Political Economy* 113(2), 376–424.
- Doll, C. N., J.-P. Muller, and J. G. Morley (2006). Mapping Regional Economic Activity from Night-Time Light Satellite Imagery. *Ecological Economics* 57(1), 75–92.
- Grainger, C. A. (2012). The Distributional Effects of Pollution Regulations: Do Renters Fully Pay for Cleaner Air? *Journal of Public Economics* 96(9), 840–852.
- Greenstone, M. (2002). The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufacturers. *Journal of Political Economy* 110(6), 1175–1219.
- Henderson, J. V. (1996). Effects of Air Quality Regulation. *The American Economic Review*, 789–813.
- Jin, X. and T. Holloway (2015). Spatial and Temporal Variability of Ozone Sensitivity over China Observed from the Ozone Monitoring Instrument. *Journal of Geophysical Research: Atmospheres*.
- Kahn, M. E. (1997). Particulate Pollution Trends in the United States. *Regional Science and Urban Economics* 27(1), 87–107.
- List, J. A., D. L. Millimet, P. G. Fredriksson, and W. W. McHone (2003). Effects of environmental regulations on manufacturing plant births: evidence from a propensity score matching estimator. *Review of Economics and Statistics* 85(4), 944–952.

- Muller, N. Z. and P. Ruud (2016). What Forces Dictate the Design of Pollution Monitoring Networks?
- Rabassa, M. J. (2008). Ambient Air Pollution and the Allocation of Environmental Enforcement Effort. *Unpublished Manuscript*.
- Russell, A., A. Perring, L. Valin, E. Bucsela, E. Browne, P. Wooldridge, and R. Cohen (2011). A High Spatial Resolution Retrieval of NO<sub>2</sub> Column Densities from OMI: Method and Evaluation. *Atmospheric Chemistry and Physics* 11(16), 8543–8554.
- Shimshack, J. P. (2014). The Economics of Environmental Monitoring and Enforcement. *Annual Review of Resource Economics* 6(1), 339–360.
- USEPA (1999). The Benefits and Costs of the Clean Air Act from 1990 to 2020. In *Report, EPA-410-R-99-001, Final Report to US Congress*.
- WHO (1999). Monitoring Ambient Air Quality for Health Impact Assessment. *Copenhagen: WHO Regional Office for Europe*.

## A Appendix

### A.1 NOAA Nighttime Data

Figure 11: Example of Nighttime Data: Chicago



Notes: Nighttime light data is overlaid on a map of the major highways in the United States. We net out the light on top of these highways to exclude information on traveling cars.