Pollution Monitoring, Strategic Behavior, and Dynamic Representativeness

Lin Yang *

January 26, 2021

Link to the latest version

Abstract

Air quality evaluation in major countries around the world is mainly based on stationary, in situ monitors that aim to provide a representative measure of local air quality. To comply with air quality standards, local governments may take targeted measures to reduce pollution around the monitors. The strategic response could lead to changes in the spatial representativeness of the monitors in the long run. Using high-resolution satellite-based air pollution measures, I examine local governments' strategic behavior and its implications on dynamic representativeness based on the staggered roll-out of the monitoring system in China. My analysis shows that local governments target pollution reductions in areas closer to monitors after monitor installations, leaving pollution elsewhere unchanged or even increased. I also find heterogeneities in the strategic measures taken by local officials with different political incentives, e.g., larger strategic reductions in cities with younger mayors. My results suggest an improved policy design for air quality evaluations, which needs a combination of ground monitoring data and auxiliary pollution information from remote sensing data and public supervision.

Keywords: Air Pollution Monitoring; Environmental Regulations; Strategic Response; China

*Yang: The Dyson School of Applied Economics and Management, Cornell University (email: ly285@cornell.edu). I am very grateful for the suggestions and guidance by Shanjun Li, Cathy Kling, Ariel Ortiz-Bobea, and Ivan Rudik. I would like to thank Todd Gerarden, Binglin Wang, Deyu Rao, Avralt-Od Purevjav, and participants at the Applied Economics & Policy Seminar and the Cornell Dyson Energy and Environmental Economics Workshop for helpful comments. Financial support from the C.V. Starr Fellowship is gratefully acknowledged. All errors are mine.

1 Introduction

Enforcement of and compliance with regulations hinge on accurate measurements of implementation and outcomes.[1] Imperfect monitoring of national regulations can lead to strategic compliance at the local level, which will further bias measurements and cause policy failures. Implementation of national policies at local levels under fiscal and political incentives is a principal-agent problem inherent in the delegation of authority by governments to bureaucratic officials (Aghion and Tirole 1997).[2] Given the ubiquitous information asymmetry between central and local governments, local regulators are likely to implement targeted strategies to meet national policy goals. In the field of environmental regulation, studies have found firms and local governments responding to different regulation stringencies in ways that result in unintended consequences such as pollution spillover (Kahn 2004; Kahn and Mansur 2013; Kahn et al. 2015; Chen et al. 2018; Karplus et al. 2018). For example, Auffhammer et al. (2009) find targeted regulatory efforts in response to nonattainment designations under the Clean Air Act in the U.S., and He et al. (2020) find that Chinese local officials enforce tighter water quality regulations on polluters immediately upstream of monitoring stations. Thus, an accurate measure of environmental quality that accounts for local regulators' strategic behavior is critical for decentralized regulation enforcement.

Air quality evaluation in major countries around the world is mainly based on stationary, in situ monitors that aim to provide a representative measure of local air quality. China launched a nation-wide, real-time air quality monitoring and disclosure program in 2013. Over 1400 monitors in three staggered waves of cities were quickly built, and air quality in China has greatly improved in the past few years. However, the monitors do not cover the entirety of China. The central government intends to use national policy goals to achieve better air quality but only observe the air pollution at monitored areas. Consequently, the local regulation enforcement tends to target "monitor readings" instead of the actual air quality. Studies find data manipulation issues in China's air quality data before this

[1] For instance, crime reduction relies on correct detections of crime activities; tax reform requires precise estimation of population income distribution; transportation and environmental regulations need accurate monitoring of traffic and pollutants.

[2] There exists a rich theoretical literature outlining contracts that align the principal's and agent's incentives (Laffont and Tirole 1993; Bénabou and Tirole 2006). In the political contract between central and local governments, the incentives include monetary incentives such as subsidies and fines, as well as political incentives such as hierarchical assignments of duties and promotions.

real-time monitoring was introduced, indicating the importance of "monitor readings" to local regulators. (Andrews 2008; Chen et al. 2012; Ghanem and Zhang 2014) Although better monitoring technologies help improve data quality significantly (Greenstone et al. 2020a), strategic responses at local levels can still exist. Previous studies by Zou (2020) and Grainger et al. (2019) have shown firms' and local regulators' strategic behaviors in responding to either the intermittent monitoring schedule or choices of new monitor sites for the monitoring system in the U.S. However, there is a lack of empirical analysis of strategic responses to spatial gaps in monitored areas at the local level. Moreover, previous studies have not examined the monitors' spatial representativeness from a dynamic perspective. Even if the monitor siting was representative ex-ante, strategic responses could invalidate the representativeness ex-post.

In this paper, I leverage high-resolution satellite-based air pollution measures to examine local officials' strategic behaviors in pollution reduction and the implications on dynamic spatial representativeness of ground monitors in China. I use a distance-based Difference-in-Differences analysis with treatment intensity to study the strategic behaviors. The staggered roll-out of the new monitoring system allows cities that joined in different waves to serve as treated and control cities for each other. I then examine the strategic pollution reductions by defining a treatment intensity indicator. The areas near monitors are classified as "monitored" areas, and areas far away from monitors are "unmonitored" areas. I then compare the pollution changes before and after a monitor is opened. In order to learn if such strategic behaviors would change regulatory effectiveness, I examine the spatial representativeness of ground air pollution monitors by comparing population-weighted average pollution levels of an entire city to the city's average pollution based on monitored locations. In doing this, I find that most of the monitors represent the city's average air quality well at the years of monitors roll-out. However, the spatial representativeness is changing over years, indicating spatially differentiated pollution changes within a city.

My paper fills the spatial gaps of ground-level monitoring data by using fine-scale grids data to study the pollution changes over space. The satellite images include annual PM2.5 (fine inhalable particles, with diameters that are generally 2.5 micrometers and smaller) grids at the 1km by 1km resolution (over nine million grids for all of China) from 2000 to 2017.[3] Using the annual level data, I avoid concerns about the missing data in most monthly

[3]By combining satellite-based measures of AOD with chemical-transport modeling and land character-

and daily satellite data. Moreover, the fine-scale grids provide rich spatial variations. This satellite-based PM2.5 data is becoming popular in economic studies because it fills the gaps in ground monitoring networks and validates the data quality at the ground level. (Sullivan and Krupnick 2018; Fowlie et al. 2019) To provide evidence supporting the political incentives behind strategic pollution reductions, I collect data on city characteristics such as population, GDP, etc., as well as information about local officials from the China Political Elite data, which records the local officials' career path, age, and education.

The main finding of this paper is that areas adjacent to monitors experience 6.5% lower PM2.5 concentrations than those farther away, and the results are robust to alternative definitions of monitored and unmonitored groups. [4] The baseline impact of monitoring on overall air pollution is positive (pollution increases), showing that the strategic pollution reduction may lead to pollution leakages to unmonitored areas. I use an event study analysis to show that the parallel trends hold for pre-opening periods in general. Moreover, by including post-opening periods, I find that the difference in pollution becomes larger as the final assessment deadline approaches. [5] My results are robust to placebo tests of random monitor locations and random monitor opening dates. To eliminate the concerns about measurement errors in the satellite-derived PM2.5 data, which may correlate with ground monitors spatially, I also run the same analysis using raw daily satellite Aerosol Optical Depth (AOD) readings and find robust results.

One additional identification concern may arise from the fact that most monitors are placed in urban centers with poor air quality, so the political interpretation of the results may not be appropriate. Thus, the difference in pollution reduction patterns between the monitored and unmonitored groups may not necessarily be caused by local regulators' strategic responses to stringent environmental targets. Instead, the results could be driven by pollution transported from polluted areas to the cleaner area. Another possibility is that regulators choose to prioritize more polluted areas first instead of gaming the evaluations. I eliminate this type of concern by conducting a heterogeneity analysis in which I compare

istics, van Donkelaar et al. (2019) derive ground-level concentrations of PM2.5 at high levels of spatial disaggregation.

[4] In the main finding, cells within 3km of a monitor are defied as the monitored area, and cells outside 3km are in the unmonitored group.

[5] According to the Air Pollution Prevention and Control Action Plan announced in 2013, the central government conducted a final assessment of overall pollution reduction at the end of this action plan in 2017.

the strategic pollution reductions for monitors located in dirtier areas to monitors located in cleaner areas of a city. I find no significant impact of monitors being in a polluted area on strategic reductions.

I have conducted heterogeneity analyses to support the political interpretation of strategic pollution reduction. First, I find strong heterogeneity across cities according to the timing of entering the new monitoring program. The later a city joins the monitoring program, the larger strategic responses that are observed. Second, I have also conducted a heterogeneity analysis by cities' pollution compliance levels, where I find a larger strategic reduction in cleaner cities and cities with pollution levels approaching the national standard. Third, cities with younger mayors who have greater promotion chances have larger strategic responses. Lastly, I find that having an economic recession in the previous year shifts local officials' regulation focus from environmental performance to economic growth, and leads to smaller strategic reductions. Taken together, these findings consistently confirm the existence of local officials' strategic pollution reduction, which arises from the misalignment between the national policy goal and local bureaucratic incentives.

Local officials employ a few strategies to reduce pollution near monitors strategically. The next part of the paper discusses the channels through which the spatial differences in pollution reductions occur. The potential channels could include local measures such as directly cleaning the air near monitors or shutting down restaurants and small workshops near monitors, and non-local measures such as relocating polluting sources away from monitors or implementing traffic control. Local pollution reduction measures reduce air pollution in areas adjacent to monitors without increasing pollutions elsewhere, whereas non-local measures will lead to pollution leakages to unmonitored areas. My results suggest that non-local measures dominate, and pollution leaks to areas more than 60km away from monitors. Although there is no data available to test for the mechanisms directly, the political incentives behind the strategic behaviors are strongly supported by government reports, media news, and multiple heterogeneity analyses.

I provide policy suggestions for a better air pollution monitoring system. My analysis of spatial representativeness suggests that most of the monitors are good representations of a city's average air quality at the beginning of monitors roll-out. However, given local officials' strategic responses and the fact that monitor locations are unlikely to change once sited, my simulation of future monitors' representativeness shows that the ground monitoring system will not be representative in the long run. Since ground monitors are costly to build, and the observed strategic response may still exist even with new monitors, it is important for the central government to combine ground monitor readings with external sources of pollution measurements such as satellite, mobile monitors, and public supervision.

This paper makes the following contributions. First, my results highlight the importance of accounting for local regulators' strategic responses when the central government designs national policies. By documenting the gap in pollution reductions for monitored and unmonitored areas, I provide evidence that policies that are ex-ante efficient will not necessarily be efficient with the existence of strategic local responses. My paper is the first empirical study which links the local official's strategic behaviors with the dynamic change in monitor representativeness and examines the underlying political incentives.

My paper adds to the growing literature on the political economy of environmental regulation by highlighting the implementation of national regulations at the local level. (Kahn 2004; Kahn et al. 2015; Jia and Nie 2017; Chen et al. 2018; He et al. 2020) A few of these studies focus on the upstream-downstream gap in China's water pollution regulation. A recent study by He et al. (2020) discusses how imperfect performance monitoring of water pollution in China can break down the central-local alignment. In my paper, I show that the gaps in ground monitoring networks can lead to significant deviation in the local air pollution regulations from what the central government observes.

Second, I contribute to the growing literature on the environmental monitoring regulation and enforcement (Gray and Shimshack 2011; Duflo et al. 2013; Shimshack 2014). While existing literature mainly focuses on the air pollution monitoring system in the U.S. (Grainger et al. 2019; Zou 2020), where they look at either the intermittent monitoring schedule or monitor siting from a static spatial point of view. My paper adds to the limited studies looking at the new air quality monitoring program in China and particularly examines the dynamic changes in monitors' spatial representativeness due to local officials' strategic responses to gaps in monitor coverages. My paper relates closely to two of the concurrent studies. Greenstone et al. (2020a) show the improvement of data quality with the help of the new monitoring system, and Barwick et al. (2020) focus on the relationship between information disclosure in the new program and people's avoidance behaviors. My study complements the previous two in that I reveal the heterogeneous impact of the system on air quality caused by local regulators' strategic responses to gaps in monitoring coverages. With the strategic responses, the information disclosed to the public would be inaccurate, and people's avoidance behavior may be biased (especially for rural households). My study is also widely applicable to monitoring regulation in other countries in both the developed and developing world because they either have monitoring networks that were built decades ago or need to design a new monitoring system.

Third, this paper adds to the literature on the value of satellite data in environmental regulations. Taking advantages of the high-resolution satellite images of air pollution, I am able to fill the gaps in ground monitoring and examine the pollution changes across different regions. In particular, I use satellite measures to evaluate the population-weighted pollution levels in each city and the representativeness of the ground monitoring system. Similar studies in the U.S. context also prove the value of satellite data and show the bias in attainment and non-attainment designations using only ground monitor's readings and the resultant welfare losses (Sullivan and Krupnick 2018; Fowlie et al. 2019). In addition to air pollution regulations, the value of satellite data in fields like climate change, wildfire surveillance (Ruminski et al. 2007), forest land cover (Hansen et al. 2013), and biodiversity (Turner et al. 2015) has been increasingly recognized by regulators and researchers.

Finally, I provide policy implications for an improved air pollution monitoring and enforcement. The central government should use auxiliary pollution information from remote-sensing data and public supervisions, together with the ground-level monitoring data, to evaluate pollution conditions. Although it is difficult to directly test the mechanism of local regulators' strategic pollution reductions due to data limitations, I provide indirect evidence for the role of economic development pressure, local regulators' characteristics, and public pressure. My results support the political incentives behind local officials' strategic behaviors and show the importance of an incentive-compatible enforcement from the central government.

The remainder of the paper is organized as follows. Section 2 provides a brief background on environmental regulations and the monitoring system in China. Section 3 describes the main data sources. Section 4 presents the main identification of local officials' strategic pollution reductions. Section 5 explores channels and mechanisms underlying the strategic behavior. Section 6 discusses policy implications for the air pollution monitoring system. Section 7 concludes.

2 Institutional Background

The benefits of China's unprecedented economic growth in the past decades are built upon the huge cost of a stained environment. China's unprecedented economic growth relies heavily on industrialization and fossil fuels, and lax environmental regulations. Over the last 40 years, China has experienced the fastest economic growth and became the largest consumer of energy and coal while also having many of the most polluted cities in the world.[6] Severe air pollution (known as "smog") in major cities attracted the attention of the international community, putting pressure on the central government of China. In the past decades, public awareness of air pollution rises, and more research has revealed the negative impact of air pollution on human health, both physical and mental. The Chinese government began to shift its policy priority from the long-lasting economic growth to environmental concerns and introduced stringent regulations on air pollution. This section introduces the political system and environmental regulations in China and discusses the underlying nature of local officials' strategic behaviors.

2.1 Political System in China

Political incentives are one of the internal mechanisms of both economic development and environmental protection, especially in China. A salient feature in China's political system is that the central government sets targets and links the local officials' promotion to their performance in these targets. Local officials, in turn, are highly incentivized and are given great flexibility in local regulatory plans to meet the national targets. Studies in political economics have examined the principal-agent problem lies in China's economic development. The incentive-based strategic responses by local governments have led to many unintended consequences such as inequality, collusion, corruption, and cheating, which may undermine the policy goals. (Li and Zhou 2005; Fisman and Wang 2015; Oliva 2015; Jia and Nie 2017; Jia 2017)

The Target Responsibility System launched in the 11$^{\text{th}}$ Five-Year-Plans (FYPs) in 2005

[6] "Helping China Fight Air Pollution", The World Bank. https://www.worldbank.org/en/news/feature/2018/06/11/helping-china-fight-air-pollution

marked an important transformation in China's national policy, where environmental targets were incorporated into the evaluation criteria of local officials.[7] In this system, local leaders who fail to attain environmental performance targets, no matter how successfully they accomplished all other tasks, would receive an unqualified evaluation in their year-end comprehensive assessment, and would not be eligible for any annual bonuses or career advancement. However, such a motivation system has also motivated strategic responses. More recent literature has placed the spotlight on the firms and local governments' behaviors under various water and air pollution regulations. The strategic responses to environmental regulations have led to issues like data manipulations (Chen et al. 2012; Ghanem and Zhang 2014; Karplus et al. 2018) and pollution spillovers (Kahn 2004; Kahn et al. 2015; Chen et al. 2018).

2.2 Environmental Regulations in China

Air pollution regulation has been a top priority of the central government of China in the past decade. It declared "war on air pollution," implementing a series of mitigation actions, such as the "Air Ten" action plan that was announced in 2013, (the Air Pollution Prevention and Control Action Plan). The action plans add detailed pollution control requirements to the 12^{th} FYPs in terms of targets, standards, measures, and technologies. In addition to the plans, a raft of new environmental protection laws and guidance are enacted, which are claimed to be the "strictest ever" environmental policies regulations to show the central government's determination to win this "war".

Under the set of stringent regulations that closely correlate with local official's own incentives, it is not a surprise to see that China has made significant progress in pollution reduction and prevention over the past decade. For example, the "Air Ten" evaluates local officials' performance in pollution reductions on an annual basis. In addition, the central government conducted a final assessment of overall performances at the end of this action plan in 2017. The promotion of local officials is not the only aspect linked with their performance in pollution control. The government budgets and new projects related to air pollution are linked to the local officials' performance as well.

Stringent central regulations have helped improving air quality in China, according to the

[7] China's five-year planning process defines overarching principles to guide national policy and broadly sets forth regulatory objectives for both economic growth and environmental protection.

ground monitor readings. For example, Greenstone et al. (2020b) estimate the air pollution trend since 2013 ("Air Ten") and show that all of the air pollutant concentrations dropped sharply, except for O3, which saw a modest increase. PM2.5 levels dropped by 27.7 $\mu g/m^3$, or about 41 percent from the 2013 level. However, the sharp reduction in air pollution is based on the ground monitor readings, which may be subject to bias due to gaps in spatial coverages. My paper aims to dig deeper into this pollution reduction trend and study the local governments' strategic pollution reduction behaviors using the newly disclosed monitoring system.

2.3 Monitoring Systems for Ambient Pollutants

Evaluating a city's air quality and local officials' performance is mainly based on the stationary, in situ monitors. Along with the evolution of China's environmental regulation and policies, the monitoring system for ambient pollutants evolves significantly. The data quality in China has been criticized a lot, especially for air pollution data before 2013: only 74 major cities had monitors, the data was reported by local governments as a daily air pollution index, and not available to the public. Obviously, local governments have great power to manipulate the reported air pollution data. As shown in Ghanem and Zhang (2014), when the policy goal is the number of "blue sky days" in a year, that is when the air pollution index is less than 100, the air pollution data reported by local governments is bunching at the cut-off. [8]

To win the "war against pollution" after 2013, China launched a nation-wide, real-time air quality monitoring and disclosure program, which quickly built-out over 1400 monitors. Several major improvements have been made in this new monitoring program. Firstly, PM2.5 is listed as a major pollutant. Secondly, the monitored data are uploaded to the cloud automatically, which significantly eliminates the data manipulation issue in the pre-automation self-reported pollution data. A recent study by Greenstone et al. (2020b) shows the improvement in data quality with the new monitoring system, and the increased public awareness of pollution prevention.

There are three types of monitors in China: 1. Monitors controlled by the central government; 2. Monitors controlled by local government; 3. Micro Monitors for specific polluting

[8] "Blue sky day" is a term introduced by the central government in 1998 when Beijing was bidding to host the Olympics, at which the city's Air Pollution Index is less than 100. The number of "Blue sky days" is a critical basis to evaluate a city's air quality condition.

sources. The central government control monitors are the first group of monitors set up before air pollution becomes a society-wide concern. Also, the local government has a relatively low involvement in the central monitors. Most importantly, the performance of local officials in eliminating air pollutions is based on the readings of central monitors. To help better control for polluting sources, the local officials build many local government control monitors, which are not included in evaluating a city's average pollution. [9]

In order to regulate the siting and operation of the monitors, the central government issued guidelines for air quality monitoring. The guidelines include the monitors' location choices, monitoring techniques, management of the monitoring data, and penalties for data manipulation and other human intervention of the monitors. The central government state that only central monitors will be counted into the evaluation of cities' average air quality conditions and local official's performance in pollution reduction. Local monitors, although built under the same guidelines, will only be helping local officials in detecting polluting sources and designing for local regulatory plans.

Three waves of prefectural cities entered the monitoring system successively in each year between 2012 and 2014. Major development regions such as the Jing-Jin-Ji region, the Yangtze River Delta region, and the Pearl River Delta region, as well as a few large cities such as provincial capitals, are the first wave to enter the new monitoring network. In these cities, many of the monitors were built and operated long before the new monitoring system was introduced. Entering the program means upgrading the existing monitors to automation, as well as adding new monitors. By the end of 2012, 496 monitors in 74 cities started to work. The second wave and the third wave then added around 450 and 550 monitors into the network each year. Cities in three waves vary largely in terms of their hierarchy level and overall environmental performances. Figure 1 shows the three roll-out waves of monitors in China. The national monitoring network with 1499 central monitors is designed to serve for urban areas of 336 cities. The number of monitors in each city is based on the population density and a city's pollution level in the past three years.

Since local officials do not have much control over the location choices of central monitors, ideally, as long as the central monitors well-represent local air quality conditions, the monitoring network should be efficient. Moreover, the central government encourages third party

[9] As of 2016, there were more than 2000 monitors in China, including both central and local monitors.

companies to gradually take over the operation and maintenance of these monitors, which greatly eliminates the possibility of direct data falsification, shutting down or destroying the monitoring devices. Data accuracy has been significantly improved after the involvement of third-party organizations (Niu et al. 2020). However, manipulations and strategic responses by local officials never ended. Medias covered several stories of constantly watering the monitored areas with fog cannon trucks, shutting down small-scale workshops, and food trucks near monitors, which burnt coal.

There is a lack of empirical evidence for local officials' pollution reduction strategy facing the new monitoring system. Since the scattered monitors lead to gaps in measuring the pollution exposure for unmonitored areas, bias may still exist due to local officials' strategic responses in having spatially differentiated pollution control measurements in monitored areas and unmonitored areas. The issue is not unique in China. Fowlie et al. (2019) and Sullivan and Krupnick (2018) examine the misclassification of attainment and non-attainment designation of counties due to the gaps in ground air pollution monitors in the U.S, and the potential welfare loss using the satellite-based pollutant data as references. Grainger et al. (2019) also use satellite NOx data to check the strategic siting behaviors of attainment and non-attainment counties. They find avoidance behaviors of local officials in attainment counties near the non-attainment threshold, where they strategically place new monitors at a relatively clean area of the county. Inspired by these studies, I use the remote-sensing data to fill the gaps in ground air pollution monitoring system and find evidence for local officials spatially differentiated pollution control strategies.

3 Data

3.1 Remote Sensing Data

In order to examine the spatial difference in air pollution regulations, this paper fills the gap in the ground monitoring system using high-resolution images of the major air pollutant, PM2.5, which are derived from the original satellite measures of Aerosol Optical Depth (AOD). The satellite AOD data comes from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) algorithm. AOD measures the total vertical distribution of particles and gases within a grid according to the light extinction coefficient. It indicates how much direct sunlight is prevented from reaching the ground by aerosol particles and can be used to infer ground-level pollution, particularly for fine particles such as PM2.5 and PM10. Atmospheric science literature has shown a strong correlation between satellite measure and ground-level pollution data.[10] Since the satellite measures are largely affected by cloud coverages over an area, missing data is a big issue when using remote sensing data with fine spatial and temporal resolutions. Studies of the remote sensing techniques find better correlations between AOD and ground-level PM with coarser spatial and temporal resolutions by month or year (Hoff and Christopher 2009).

The satellite images this paper uses include annual PM2.5 grids (1km by 1km resolution, nine million grids for whole China) from 2000 to 2017. By combining satellite-based measures of AOD with chemical-transport modeling and land characteristics, van Donkelaar et al. (2019) derive ground-level concentrations of PM2.5 at high levels of spatial disaggregation. One concern with the satellite-derived ground-level pollution measure is the measurement errors caused by the calibration of the satellite data using ground monitoring data. Even though that van Donkelaar et al. (2019) use geographical weighting method to give smaller weights to cells further away from ground monitors, and larger weights to cells closed to ground monitors, one may be worried about different measurement errors may occur at cells with different distances to monitors. To address this concern, the authors conducted cross-validation tests, where they remove part to all of the ground monitors from the calibration. The derived PM2.5 data still performs well. [11]

3.2 Spatial Representativeness of Ground Monitors

With the fine-scale pollution data and spatial information of the new ground monitoring network, I examine the spatial representativeness of these monitors. First, I apply the kernel density estimation to compare the pollution distribution of monitored cells with that of unmonitored cells, following the methodology from Grainger et al. (2019), which define a z-score for each grid in each city to measure the within-county variation. [12] I also compare the spatial distributions of different types of ground monitors: central vs. local. The kernel

[10] Liu et al. (2007); Lee et al. (2012); Zhang and Li (2015). Previous economic research using the satellite measure as the proxy for ground-level pollution includes Foster et al. (2009); Chen et al. (2013); Bombardini and Li (2016); Sullivan and Krupnick (2018); Fowlie et al. (2019).

[11] I have also used raw daily AOD data downloaded from NASA's MODIS system to check the robustness of my analysis to potential measurement errors that correlate with locations of ground monitors.

[12] Z-score is calculated by taking the observed value in grid cell i in city c and year t, subtract the average for that city, and scale it by the city level standard deviation.

density estimation result in Figure 2(a) shows that monitors are mostly placed in a relatively more polluted area in a city. This is consistent with the intuition that most monitors are placed in urban areas to cover more population. Figure 2(b) shows that local monitors are placed in a slightly cleaner area comparing to central monitors. This is intriguing because one would expect the local officials to put local monitors nearer to polluting sources in order to regulate air pollution directly.

One thing to notice is that almost all monitors are located in urban areas, and the sparse central monitors are the only base in evaluating the air pollution condition of a city. The gaps in the ground monitoring network might cause the regulation focus to bias toward urban citizens. Instead, the less-monitored places, i.e., the rural areas' pollution, will not be considered in evaluating the local officials' environmental performance. Contrarily, the satellite-based measurements give a highly spatial resolved coverage of the air pollution in the entire city area. To examine the difference between monitor-based and satellite-based city average PM2.5, I use the 1km by 1km gridded population count from 2015 Census to weigh each cell and calculate the weighted average PM2.5 for each city. Taking this as the "true" city-level PM2.5, I then compare it with the monitor-based population-weighted average PM2.5. The map in Figure 3 shows the monitors representation errors in the years that cities joined the system. I regard the cities with errors within $\pm 10\%$ as having well-representative monitors. The warm colors are cities where monitors over-represent the "true" city-level PM2.5, and the cool colors are cities with under-representative monitors. The representation errors exhibit large spatial variations, where two-thirds of cities have over-representative monitors, consistent with the kernel density figures. I have also included a set of interesting correlates in Appendix A2 and A3 to check if the leaders' characteristics, the GDP per capita, or industrial type matters for the "representation errors".

The representation errors in Figure 3 are static at the moment of their openings. If the pollution reduction patterns are even across space, then the representativeness of monitors would not change as long as the monitors' locations do not change. However, though monitors are unlikely to move for a long period, local regulators' strategic responses to the static monitor locations would change the monitor's spatial representativeness overtime. From the representation error maps in each year (Appendix A1), this is indeed the case. Monitors' spatial representativeness exhibits dynamic changes in years after cities joining the program, which greatly motivates my study of local regulators' strategic pollution reduction behaviors.

3.3 Other Data and Summary Statistics

To check if other factors would affect the spatial representativeness of ground monitors and the strategic environmental regulating behaviors, I collect data on city characteristics such as population, GDP, etc., and weather variables, such as temperature, humidity, wind directions, wind speed, etc. I have also collected information about local officials from the China Political Elite data, which includes local officials' career path, age, and education.

The summary statistics are presented in Table 1 and 2. Table 1 presents satellite PM2.5 summary statistics by calendar year. Over the period of study, the PM2.5 level increased significantly before 2013, and then declined. After the declaration of the "war against pollution", there is an overall improvement in air quality. In Table 2 , I present a summary statistic by different waves of cities, where I summarize the population-weighted PM2.5 density using the 2015 population in each grid cell as the weight. I also summarize the population-weighted PM2.5 density at cells containing monitors, which are in general higher than the city average PM2.5 in all three waves of cities. In addition to the PM2.5 densities, I also include a summary of city characteristics such as the population and GDP by the three waves. Comparing the three waves, I find that cities in earlier waves tend to be dirtier and have more population. In terms of GDP, and the GDP in each industry, cities joining the program earlier tend to be more economically developed. The difference in city characteristics among waves may lead to different environmental strategies and regulation outcomes. Because cities in earlier waves are high in the hierarchy rank, city officials' characteristics could be different. From the summary statistics of city mayors' age and education, I find that wave one cities have slightly older mayors and more mayors with PhD degrees. Most mayors in wave two cities own master's degrees, and most city mayors in wave three have bachelor's degrees.

4 Strategic Pollution Reduction After Monitoring

4.1 Empirical Framework

I examine the strategic pollution reductions in monitored areas after monitoring using a Difference in Differences method with a staggered roll-out schedule. Joining the new monitoring program by either having new monitors or automation of existing monitoring data could change local officials' incentives and strategies to meet environmental targets. Thus, once a city joins the program, it will be considered as in the treated group. Within each treated city, there will be different treatment effects by distances away from monitors. I use the following empirical framework to examine the impact of monitoring on overall air quality and the heterogeneous treatment effects by treatment intensity:

$$\ln(PM2.5_{iwt}) = \alpha Open_{wt} + \beta Near_i \times Open_{wt} + Cell_i + Year_t \ + Trend_{wt} + \varepsilon_{it} \qquad (1)$$

The outcome variable, $\ln(PM2.5_{iwt})$, is the logarithm of annual PM2.5 concentration at the 1km$\times$1km grid cell. i is the index for grid cells within cities opened in wave w at year t. In my study, there are over nine million cells' annual PM2.5 from 2000 to 2017 in the raw data. $Open_{wt}$ is the treatment indicator that takes the value of 1 if cell i is in a wave w city after joining the new monitoring program. The treatment intensity is defined by $Near_i$, which equals 1 if the grid cell i is in an area adjacent to a ground monitor (monitored area), and 0 if the cell i is in areas far away from monitors (unmonitored area). In most cases, I am less interested in the causal effect of the monitoring program per se (α), but rather more in the difference in the causal effect in monitored vs. unmonitored areas (β) after monitoring. Due to the large spatial and temporal variations in air pollution, there may be confounders that would bias β from identifying the difference in pollution reductions across space. Especially, cells in monitored and unmonitored areas could have different location attributes that affect air quality. To address these concerns, I report results of estimations with a rich array of controls, including cell fixed effect and year fixed effect. I also include a wave-specific time trend to allow the unobserved time trend in pollution to vary across waves. The identification variation is then from comparing cells in monitored vs. unmonitored areas before vs. after new waves of monitor roll-out. Since pollution observed at a cell is likely driven by emissions elsewhere that also affect nearby cells, I cluster standard errors at the city level.

Cities selected into the program in different waves may be due to wave-specific unobservables that are time-variant. Cities in earlier waves tend to be larger cities with more population, higher GDP per capita, higher levels of air pollution and industrial emissions, etc. I include wave by year fixed effects in Equation (2). Although the fixed effects absorb the baseline impact of monitoring on overall pollution (α from Equation (1)), Equation (2) provides a clearer identification of changes in treatment effect by treatment intensity (β). It also has more flexible controls than the wave-specific time trend. The identification variation now is from comparing monitored vs. unmonitored cells in the same wave cities, before and after monitoring. The key explanatory variable is $Near_{it}$ which is an interaction of $Near_i$ and $Open_{wt}$.

$$\ln(PM2.5_{it}) = \beta Near_{it} + Cell_i + Wave_w \times Year_t \ + \varepsilon_{it} \tag{2}$$

4.2 Baseline Results

In the baseline results, I estimate equation (1) and (2) using my preferred sample from 2007 to 2017, which includes five years prior to monitoring and all post monitoring years to have a relatively balanced panel. [13] The monitored area is defined as grid cells within 3km of a monitor. The results are robust to alternative definitions of monitored areas such as 2km, 5km, and 10km, and unmonitored areas such as outside 3km, 30km, and 50km where I drop the cells in between to address the concern of misclassifying monitoring status. The DID with treatment intensity provides estimates of local effects within the choice of the treatment intensity groups, where results using different monitored areas could represent different pollution control strategies that local officials adopt. I will discuss more in the next section.

Table 3 presents the baseline DID result by estimating Equation (1) and adding controls sequentially. In the first four columns, *Open* captures the baseline impact of joining the monitoring program on air pollution, comparing to control cities. The baseline DID estimates of the causal impact of monitoring on air pollution are positive (pollution increases) and significant across the controls. In Figure 4, I conduct an event study for the causal impact of monitoring, where I replace the treatment indicator $Open_{wt}$ with opening dummies for each year pre and post monitoring. The event study figure shows no pretrend, and significant increases in pollution after cities joined the program. [14]

[13] For cities in the first wave, the sample period is [2007, 2017] with five years pre and post monitoring; for cities in the second wave, the sample period is [2008, 2017] with five years pre and four years post monitoring; for cities in the third wave, the sample period is [2009, 2017] with five years pre and three years post monitoring.

[14] Goodman-Bacon (2018) points out the concern of DID with heterogeneity in treatment timing, which could be a valid concern for my baseline DID estimation of the causal effect of monitoring (α). Thus, an event study is preferred than an average treatment effect. In my paper, the three waves of cities entered the program consecutively within three years. The potential impact of wave-specific factors affecting the pollution in different years has been controlled by the Wave by Year FE. The estimated key parameter of interest (β) is the different pollution changes among treatment intensity groups within a wave of cities after monitoring.

I then include the treatment intensity indicator $\mathbf{1}$(0-3km) in column (5) & (6) to capture the heterogeneous treatment effects of the monitoring program on pollution in monitored (cells within 3km) vs. unmonitored (cells outside 3km) areas.[15] The results from column (6) show that pollution in monitored areas is decreased after monitoring by 2%. Unmonitored areas exhibit 4% higher pollution after a city joins the program, indicating the potential pollution leakages.

In Table 4, I show the baseline DID results are robust to alternative definitions of treatment intensity groups. The first three columns present results for monitored areas defined as cells within 3km of monitors, and compare to different unmonitored areas such as cells outside 30km and 50km of monitors. The areas in between are dropped to have a clearer defination of treatment intensity groups. Column (4)-(6) expand the monitored areas to five distance bins to show how the treatment effect varies over space. Consistent with intuition, the difference in pollution changes between monitored and unmonitored groups are larger when two groups are more apart from each other, and the differences are smaller when monitored areas are further from monitors.

In Table 5, I use wave specific year fixed effects to absorb the baseline causal effect of monitoring and show the relative changes between treatment intensity groups (Equation (2)). Column (1) presents the main finding of my paper. Pollution in monitored areas is 6.5% less than that in unmonitored areas after monitors roll-out. Similar to Table 4, the results are robust to alternative treatment intensity groups. The first three columns present results for monitored areas as cells within 3km to monitors, and compare to different unmonitored areas such as outside 30km and 50km of monitors. The areas in between are dropped to have a clearer definition of treatment intensity groups. Column (4)-(6) expand the monitored areas to five distance bins to show how the treatment effect varies over space. Consistent with intuition, the difference in pollution changes between monitored and unmonitored groups is larger when two groups are more apart from each other, and the differences are smaller when monitored areas are further from monitors.

[15] Without controlling for cell fixed effect, the raw difference between two treatment intensity groups is positive. This result is likely driven by the fact that the urban centers, where most monitors are placed, tend to have higher pollution levels than other areas of a city. Once cell fixed effect is included, the results show that areas near monitors experiences larger pollution reductions after monitors opened.

4.3 Identification

The key assumption is that in the absence of a monitor opening or switching to automation, air quality in the monitored and unmonitored areas follow parallel trends. In other words, I assume that the only reason that ambient air quality might show a significant difference between areas nearby monitors and areas far away from monitors is because that local officials strategically put more efforts into reducing "local" air pollution. As directed by the central government, most monitoring stations are placed in urban centers to cover populated areas. One may be concerned that cells in the unmonitored areas are too far away from the city center and thus would have different pollution trends from those in the monitored areas. While the parallel trend assumption is not directly testable, I conduct a "placebo" test and an event study analysis to support the assumption. To address the identification concern of endogenous monitor locations, I conduct another "placebo" test with random monitor placements.

4.3.1 Placebo Tests

First, I conducted a "placebo" test using only pre-program periods and randomly assign opening years for all monitors at the same locations. The rationale behind the placebo test is that cells in "monitored" and "unmonitored" areas should not be significantly different over a false-opening year in the absence of the monitoring program. For each monitor, I randomly assign an opening year between 2007 to 2011 for 500 times. I then conduct 500 estimations of equation (2) and plot the distribution of the coefficients in Figure 5. Comparing with the observed coefficient, I find that the observed coefficient lies outside of the 99% confidence interval of the coefficients from 500 placebo tests, which center around 0.016. This result shows that before the monitoring program, a false opening would not lead to larger pollution reductions in monitored areas than unmonitored areas.

Second, in order to show that my findings indeed a result of local pollution reductions in monitored areas, I conduct a placebo test with random monitor locations. Keeping the number of monitors and the year of joining the program unchanged, I randomly relocate all the monitors within each city 500 times. The underlying idea is that if local officials only conduct strategic reductions in areas very closed to monitors, then no significantly different pollution reduction should be observed in areas with a false monitor opening compare to other areas in the city. After matching the 500 groups of placebo monitors with the satellite grid cells, I estimate equation (2) and plot the distribution of the coefficients in Figure 6. The observed coefficient lies outside of the 99% confidence interval, suggesting that local pollution reductions happened only at the observed monitored areas.

4.3.2 Event Study

I use event study analysis to show the parallel trends between monitored and unmonitored groups hold for pre-opening periods in general. I divide the years around opening dates into five pre-opening periods $n = -5, -4, ..., -1$, and six post-opening periods $n = 0, 1, ..., 5$ and run the following regression:

$$\ln(PM2.5_{it}) = \sum_{n=-5}^{5} \beta_n \phi(n) \times Near_{it} + Cell_i + Wave_w \times Year_t \ + \varepsilon_{it} \tag{3}$$

where $\phi(n) = \mathbf{1}\left[n \leq t \leq n+1\right]$, indicating interval n. The base interval is the year before the opening year (i.e., $n = -1$). I expand the dataset used in main DID analysis (PM2.5 in 2009-2017) to year of 2007 which allows wave 1 and wave 2 cities to have the same number of five pre-opening periods. However, the number of post-opening periods for cities in different waves would not be the same due to data availability.

Figure 7 (and Column 1 in Table 6) presents the coefficient estimates of $\phi(n)$. The results support the parallel trends assumption in general: compared with the base interval (1-year before opening years), the subsequent changes in air pollution between the monitored and unmonitored areas are not significantly different for the four pre-opening intervals in the specification. In contrast, I find statistically significant different air pollution reduction between the monitored and unmonitored groups in the post-opening intervals for the same specification. The fifth year prior to monitoring exhibits a significant difference, which could be due to more unobserved policy changes in years further before monitoring. Column (2)-(5) in Table 6 show the event study estimation results are robust to alternative definitions of $Near_{it}$.

4.3.3 Eliminate Alternative Explanations

In this subsection, I discuss a few alternative explanations which may generate similar patterns, including monitored area's attributes, and the measurement error in the satellite-derived pollution measures. First, an identification concern may arise from the fact that monitors are in urban centers, which happen to be more polluted area. The difference in pollution reduction patterns between the monitored and unmonitored areas exists due to the nature of pollution transporting from dirty areas to clean areas. If this is the case, then one should expect to see larger differences in pollution changes after monitoring for monitored cells located at dirtier areas than monitored cells located in cleaner areas of a city.

A similar concern lies in the political interpretation of local officials' strategic behaviors. One may argue that local officials choosing to reduce more pollution in monitored areas is not a strategy that they play to gaming the performance evaluation. Instead, they choose a more cost-effective way to reduce pollution in a relatively more polluted area, which happens to be the area adjacent to a monitor. To address this type of concerns of monitored areas being coincident with polluted areas, I examine the heterogeneity of treatment effects where I allow the impact to differ based on the relative pollution levels of cells within the vicinity of monitors as in Equation (4),

$$\ln(PM2.5_{it}) = \beta Near_{it} + \eta Near_{it} \times Dirty_{it} + Cell_i + Wave_w \times Year_t \ + \varepsilon_{it} \qquad (4)$$

where $Dirty_i$ is a dummy variable which equals 1 if the PM2.5 of a cell i is higher than the city average PM2.5 level in year t. This specification examines the potential concern of monitors locating in the dirty area of a city. The coefficient η will show the different pollution gaps between monitors in a dirty area and clean area. The results are reported in Table 7, where I include alternative definitions of monitored and unmonitored areas to show robust results. Cells within 3 km of a monitor are the monitored cells in the first three columns, and I then expand the monitored areas to include more distance bins. From columns (1) to (6), I show that no matter which monitored groups, being in the dirtier area of a city does not lead to large pollution reductions as concerned. In fact, the magnitude of the interaction terms with *Dirtier* is almost zero comparing to the strategic pollution reductions in monitored areas.

Another possibility that may generate similar results is the measurement errors from satellite-derived pollution measures. The PM2.5 data I use is derived from the raw satellite images, which require information from monitor-based sources. The Geographical Weighted Regression method used when deriving PM2.5 from satellite images assigns larger weights to areas closer to ground monitors, and smaller weights to farther areas. One may be concerned that the resulted measurement errors from the data generating process will be correlated with the distances to monitors and also varied over time when more ground monitors are opened. If this is the case, then the spatially different pollution patterns could simply because of the spatially differentiated measurement errors. Although van Donkelaar et al. (2016) have conducted several out-of-sample cross-validation tests to justify their satellite-derived PM2.5 data, I conduct a robustness check using the raw satellite images to further eliminate this possible explanation. Using the raw AOD data from the NASA MODIS product, I manually aggregate the daily AOD images at 3km by 3km resolution into annual AOD, and match with the ground monitors. The grid cells containing monitors are monitored cell, and those do not contain any monitors are unmonitored cells. Estimating Equation (2) using the AOD data shows a similar result. After monitoring, pollution in monitored cells decreases comparing to unmonitored cells. (Table 8)

After eliminating alternative explanations, the empirical results shown in this section suggest that after monitoring, the area adjacent to monitors experience larger pollution reductions relative to areas farther away. So far, I have not claimed that the spatial gaps in pollution changes are due to local officials' strategic responses to central environmental regulations.

5 Heterogeneous Effects and Potential Mechanisms

In this section, I conduct multiple heterogeneity analyses to support the political interpretation of the results. I discuss the potential channels through which the heterogeneous effect by treatment intensity may occur and show how the effect size varies in various circumstances, including a cities compliance level, economic development, leader characteristics, and information transparency.

5.1 Channels for Strategic Reduction

I present the spatial distribution of the impact of monitoring and discuss abundant qualitative evidence of the local officials' pollution control strategies to support the political interpretation of my findings. By replacing the binary indicator of one monitored group and one unmonitored group used in Eq (1) with fifteen treatment intensity groups, I show the spatial distribution of the treatment effect by distances from monitors in Table 11. The changes in the impact of monitoring over space also indicate the potential channels of strategic pollution reductions. The coefficient estimates of *Open* represents the impact of monitoring on air pollution in the base group, which includes cells more than 300km away from the closest monitors. Combining with the interaction terms, the strategic pollution reductions exist within 70km ranges of monitors and are robust in magnitudes. Beyond 70km, the overall impact of monitoring turns positive and continues to increase for cells further away. With more distance bins in the unmonitored groups, Table 11 represents the potential pollution migration patterns across space after monitoring. Note that most of these central monitors are placed in population-dense (urban) areas. Column (2) in Table 11 summarizes the population in each distance bin. Although the monitoring enforcement seems to divert air pollution away from areas near monitors, this does not necessarily lead to policy failure when considering the population exposed to air pollution. However, this could exacerbate inequality issues if pollutions are leaking to rural areas. I provide more discussion in Section 6 on the dynamic changes of monitors' representativeness in population-weighted pollution exposure.

I reviewed numerous policy documents from both the central and local governments in China, collecting reports by national inspections teams, and media newsletters. They show that local governments have strong political incentives in improving air quality readings to meet the centrally designated air quality targets. As I introduced in Section 2, the most direct ways to falsify monitor readings from the devices are difficult to implement with the new monitoring system. Such direct manipulation methods include shutting down monitors during polluted days, blocking up the sensors inside monitoring devices, and deliberately damaging monitors. With the real-time data collecting monitoring system, any of these data manipulations would result in abnormal data patterns and trigger alarms. However, the advanced new system cannot eliminate all possible channels of "manipulating" the monitor readings. As the famous saying in China points out, "when the central government has a policy, the local governments have countermeasures". There are several major strategies that local regulators commonly adopt to "manipulate" the monitor readings.

The first type of strategy directly cleans up the air near ground monitors. Since the monitor locations are known to local regulators, many of them choose to clean up the adjacent areas by spraying water or using fog canon towards either monitors (higher risk of being caught, most effective), or towards trees near monitors (lower risk, less effective). A recent scandal was exposed by the media that in Jan 2018, the building of the Environmental Protection Agency in Shizhuishan, Ningxia Province, where a central monitor is located, was turned into an ice sculpture when the staff tried to reduce monitor readings with fog cannons.

The next set of strategies is the ones causing the largest pollution leakages into unmonitored areas. Short term strategies may include traffic controls in monitored areas, divert food trucks and other mobile polluting sources away from monitors, or restrict operation durations for certain polluters. An inspection report of Tianjin's environmental regulation states that the inspection team found strategic pollution reduction behaviors such as traffic controls and increased water spraying frequency in the monitored areas. Media also revealed temporarily shutting down of gas stations near monitors in Pingdingshan, China. [16]

A more effective strategy in the longer term would be relocating polluting sources from small-scale workshops, restaurants to large industrial plants to suburban or rural areas that are commonly unmonitored. This type of strategy would be preferred considering either economic development or environmental performance (improving monitor readings). However, it would impose the largest environmental damages and bias of central regulations. Based on the baseline DID results in Table 3, relocation of polluting sources seems to be the most common strategy given that unmonitored areas become more polluted after monitoring.

The strategies that local officials use to achieve better monitor readings are hard to test empirically due to data limitations. For example, traffic controls and water spraying in monitored areas are short-term actions which may only be caught by constantly observing the abnormal phenomenon near monitors. Instead, I use several heterogeneity analyses to indirectly support the findings of local regulators' strategic responses.

5.2 Heterogeneity in Strategic Pollution Reductions

I present evidence from heterogeneous analysis to show that the political incentives of local politicians are indeed the driving forces behind my main findings.

a) Roll-out Waves of Entering the Program

In addition to the annual assessment, the local officials face a final assessment of air pollution reductions at the end of 2017. They may use more aggressive strategies to reduce monitor readings when the final assessment approaches. On the contrary, major cities in earlier waves, especially those in the key development regions, face more stringent PM2.5 reduction goals. It is unclear which incentivizes local officials more in taking more aggressive strategic responses, the stringent target or the approaching deadline. In Figure 8, I investigate heterogeneity

[16] Example of news and media coverages of the existing manipulation strategies: Yuqing, People.cn; Bloomberg Law; Guancha; People.cn

in the impact of treatment intensity on pollution reductions by roll-out waves. I find the cities in later waves show larger strategic pollution reductions in monitored areas, indicating more aggressive strategies as the deadline approaches. Another possible explanation is that cities in wave one and two cities are those with monitor readings upgraded from manual to automation in the new system, rather than having new monitors opened. Thus, the strategic pollution reduction might exist before the cities join the new monitoring system.

b) Compliance Levels

Local official's pollution control strategy could be varying with the existing pollution conditions. In Table 9 and Figure 9, I explore the heterogeneity by cities' average pollution levels, using the national annual PM2.5 standard 35 ug/m3 as a reference. I use the population-weighted city average pollution at the monitored cells at the years of monitors roll-out. Cities with average pollution levels below the annual standard are defined as clean cities. I find that clean cities tend to have larger strategic pollution reductions in monitored areas after monitoring. Restricting the sample to cities with average PM2.5 from 30-40 ug/m3 shows similar results. In order to see if the heterogeneity by compliance level varies with roll-out waves, I include additional analysis using subsamples in each wave. Clean cities in wave one tend to have more aggressive strategies. This could be due to the fact that wave one cities are in general dirtier than other cities. Thus, dirtier cities in wave one are the most polluted cities in China and under strict supervision by the central government. To see how strategic response varies by the closeness to the national standard, I include another layer of interaction, *Compliance*, which is the difference between city PM2.5 and 35 ug/m3. For a clean city, when its pollution level approaches the national standard, I find larger strategic pollution reductions. The heterogeneous effect by cities' compliance levels indicates that local officials facing different compliance status choose different strategies to meet the environmental targets.

c) Leader Characteristics

City mayors play an essential role in policy regulation and implementations. I investigate whether a city mayor's characteristics have an impact on the strategic pollution reductions after monitoring. Figure 11 shows the heterogeneity analysis by city mayor's age, where I separate the sample into two subsamples by city mayors' age. A mayor has better chances to be promoted to a higher position at an age younger than 57. Thus, a younger mayor may have larger incentives to perform well in the environmental evaluation and adopt more strategic pollution reduction methods in monitored areas. For mayors older than 57, which means they have little to no promotion opportunities, they would be less incentivized to achieve policy targets. Figure 11 shows such results that cities with mayors younger than 57 tend to have larger strategic reductions in monitored areas. On the other hand, I do not find any significant impact of a mayor's educational background on their strategic behaviors. This may suggest the strategic reduction methods are common knowledge for leaders across education levels and do not require elite training.

d) Economic Growth

In general, there are tradeoffs between economic development and pollution abatement for local regulators. Prioritizing environmental regulations may hurt the local GDP growth and local officials may have different strategic behaviors in pollution control when facing different economic conditions. To examine the role of economic growth pressure, I generate a dummy variable indicating the growth or recession of a city's GDP in the previous year and interact with the DID treatment intensity term. Table 10 shows the results for all cities, and for each wave of cities. I find that no matter in which roll-out wave, when a local official faces downward pressure on economic growth, they tend to reduce strategic measures that improve monitor readings. This set of heterogeneity results suggest that local regulators are balancing both their efforts and performance in economic growth and pollution control. The gap in pollution changes between monitored and unmonitored areas is indeed a result of local regulators' strategic pollution reductions.

e) Information Transparency (Public Pressure)

Local official's strategic behaviors can potentially be captured by residents if they have full information about air pollution monitors, such as locations and readings. With the new monitoring system, information about the central monitors are publicly available through multiple sources, including the MEP's website and third-party online platforms. In addition, a few provinces have launched their own online air pollution disclosure platform. They provide detailed information about the monitor locations, including both central and local monitors. In China, eleven provinces have an online platform, which shows their effort in improving information transparency. Moreover, local residents can perform additional supervision on air pollution monitors and check consistency with online information. In

Figure 12, I investigate such heterogeneity and find that provinces with online pollution disclosure do not show significant strategic pollution reductions in monitored areas after monitoring, which suggests the importance of information transparency and public pressure could potentially reduce local official's strategic behaviors.

6 Policy Implications and Suggestions

6.1 A Well-representative Monitoring System

One would expect to see local regulators to have very different strategic behaviors facing the new monitoring system, because monitors' siting could over-represent, well-represent, or under-represent the average city pollution levels. Even though the central government intended to place the monitors in populated areas to improve the representativeness, the over-representing monitors (monitor-based pollution larger than city average pollution) could exacerbate local government's strategic pollution reductions. I conduct a heterogeneity analysis to show that it is necessary to build a monitoring system that well-represents the average city pollution level. I split the sample into three groups: "over-represent" cities with representation errors greater than 10%, "well-represent" cities with errors between -10% to 10%, and "under-represent" cities with smaller than -10% errors.

Figure 13 shows the event study on three subsamples. I find that cities with over-representing monitors tend to have more aggressive strategic reductions after monitoring, comparing to well-representing cities. It is hard to find a clear trend for cities with under-represented monitored pollution due to the few numbers of "under-represent" cities. The heterogeneous results are intuitive because if the central government places monitors in the dirtiest area of a city, local officials will be more incentivized to reduce the pollution only in the monitored area. However, unmonitored areas could still have more pollution than the national standard due to pollution leakages. Thus, it would be necessary for the central government to evaluate the cities' average pollution thoroughly and use the population-weighted average pollution as references for monitoring sites. The well-represent cities still have a slightly downward trend after monitoring. This indicates that a well-representative monitoring system could, to some extent, reduce the strategic responses at the local level but would not prevent the behaviors from happening. In fact, the strategic responses may change the spatial representativeness in the long run.

6.2 Dynamic Monitors Representativeness

From the representation error map in Figure 3, the current monitoring system in most cities shows good representativeness when the cities first joined the program. However, similar to the monitoring systems in developed countries, monitor locations are unlikely to change once the monitors were placed. For example, the current air quality monitors in the U.S. were built two decades ago, and covered populated areas following federal guidelines. Other than adding new monitors to nonattainment counties, the existing monitor locations have not changed ever since. Thus, even though monitors were sited to be well-representing counties' overall air quality in the 90s, the representativeness can be dynamic due to human interventions in monitored areas. Using my estimates for the relative pollution reductions in monitored areas (cells within 3km have 6.5% more pollution reductions), and the last observed year of pollution in my data in 2017, I calculate the projected pollution levels for five years from 2018 to 2022. I do not conduct simulations for a longer period into the future because there could be large uncertainty and new regulations. I find that in the near future, the over-representative monitors seem to become more representative of a city's overall air quality. However, there are also more cities exhibiting negative representation errors, 42 cities at years of monitoring vs. 52 cities in 2022. Even though the monitoring system works fine in my projected years, it is possible that with the strategic responses, monitors would become less representative in the long term. Moreover, the pollution leakages to unmonitored areas, mostly rural regions, could cause large health impacts and biased evaluation of policy goals.

6.3 The Remote Sensing Data and Other Pollution Information

The key to eliminating or preventing local official's strategic responses to the ground monitoring system is to add referencing data sources into the evaluation. In an ideal world with ground monitors everywhere, local officials are impossible to predict which sets of monitors would be used to evaluate their environmental performance. Thus, the only strategy left is to improve air quality city-wide. This seems unrealistic because ground monitors are large in size and costly to build and maintain. The satellite-based pollution measures can be a good source to fill the gap in ground monitor coverages. As shown in (Sullivan and Krupnick 2018) and Fowlie et al. (2019), remote-sensing data has helped the authors to assess the extent to which the existing U.S. ground monitor-based measurements over- or under-estimate true exposure to PM2.5 pollution. In my context, I have used the satellite-based data to re-evaluate the policy goals set by the "Air Ten" action plan for the end of 2017. Unlike the monitor-based pollution patterns estimated in Greenstone et al. (2020a), PM2.5 decreases by 40% from 2013 to 2018, my estimates find an overall increase in the city-wide pollution level. This suggests that monitor-based evaluation would overstate the environmental performance and distort future policy design.

However, it is important to recognize the limitations of completely relying on satellite images. Satellite-based data is not direct measures of ground pollution levels and is subject to missing data issues that are strongly correlated with cloud coverages. Ground monitors, on the other hand can provide more detailed hourly observations and better accommodate various weather conditions. Additionally, advanced monitoring technologies have provided broader coverages with mobile monitors and micro-monitors that local regulators have less control. Hence, the central government should use this information as supplementary evidence for city-wide pollution evaluation. This is true for any country relying on stationary, in situ monitors in environmental regulations. Overall, a better policy design of monitoring regulation and enforcement would need a mixed contribution from the ground monitoring system, remote-sensing technologies, mobile monitors, as well as public awareness, and third-party auditors.

7 Conclusion

Environmental regulations are often associated with strategic responses, and effective regulation relies on accurate monitoring and measurements. In major countries around the world, local governments face stringent pollution abatement targets, which often link local governments' federal funding or regulators' promotions with their success in achieving these targets. A growing literature has highlighted the unintended consequences of these policies, such as pollution spillover in China's water quality regulation, which undermines policy goals and bias evaluations. This paper adds to these studies by demonstrating strategic responses to central regulations at local levels and extending the literature to air pollution monitoring regulations. Using high-resolution satellite measures of pollution, I have shown that local officials have incentives to improve monitor readings by strategically reducing pollution in monitored areas. Such strategic behaviors will change the spatial representativeness of the current monitoring system and lead to biased policy evaluations.

I find that there exists a significant difference between pollution changes in areas adjacent to monitors and areas far away from monitors after monitoring. This result is robust to different definitions of monitored and unmonitored areas. Although the new ground monitoring network has improved data quality significantly, the gaps in monitor coverages lead to pollution leakages from monitored areas to unmonitored areas. The baseline DID result shows that pollution in unmonitored areas increases after monitors roll-out, which indicates that the underlying mechanism of such strategic reduction is non-local, relocating polluting sources away from monitors. By studying the heterogeneous impact of cities' pollution levels, the characteristics of local leaders, the role of public pressure, and the role of economic growth, I provide evidence supporting the political interpretation of the strategic pollution reductions. Overall, my results are consistent with the expectation that strategic pollution reductions are more likely to arise with larger incentives to improve monitor readings, such as in cities with younger mayors and cities with approaching assessment deadlines.

My results emphasize the importance of accurate and representative measurements in regulations and are widely applicable to any regulations with in situ monitoring systems globally. My paper contributes to the growing literature on environmental monitoring regulation and enforcement by expanding the study to China's air quality monitoring system. I highlight the importance of a monitoring regulation that accounts for local regulators' strategic responses and considers the monitoring network from a dynamic point of view. The results are also widely applicable for building or improving monitoring systems in other countries, both in the developed and developing world. I provide policy suggestions for efficient regulations that require a mixed source of pollution information from ground-level monitors, advanced monitoring techniques, and the public to accurately evaluate local officials' environmental performance and improve air quality city-wide.

References

Aghion, Philippe and Jean Tirole, "Formal and real authority in organizations," *Journal of political economy*, 1997, *105* (1), 1–29.

Andrews, Steven Q, "Inconsistencies in air quality metrics: 'Blue Sky' days and PM10 concentrations in Beijing," *Environmental Research Letters*, July 2008, *3* (3), 034009.

Auffhammer, Maximilian, Antonio M. Bento, and Scott E. Lowe, "Measuring the effects of the Clean Air Act Amendments on ambient concentrations: The critical importance of a spatially disaggregated analysis," *Journal of Environmental Economics and Management*, July 2009, *58* (1), 15–26.

Barwick, Panle Jia, Shanjun Li, Liguo Lin, and Eric Zou, "From Fog to Smog: the Value of Pollution Information," *NBER Working Paper*, 2020, *w26541.*

Bombardini, Matilde and Bingjing Li, "Trade, Pollution and Mortality in China," *NBER Working Paper*, 2016, *w22804.*

Bénabou, Roland and Jean Tirole, "Incentives and Prosocial Behavior," *American Economic Review*, 2006, *96* (5), 1652–1678.

Chen, Yuyu, Ginger Zhe Jin, Naresh Kumar, and Guang Shi, "Gaming in Air Pollution Data? Lessons from China," *The B.E. Journal of Economic Analysis & Policy*, December 2012, *12* (3).

_ , _ , _ , and _ , "The Promise of Beijing: Evaluating the Impact of the 2008 Olympic Games on Air Quality," *Journal of Environmental Economics and Management*, November 2013, *66* (3), 424–443.

Chen, Zhao, Matthew E. Kahn, Yu Liu, and Zhi Wang, "The consequences of spatially differentiated water pollution regulation in China," *Journal of Environmental Economics and Management*, March 2018, *88*, 468–485.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan, "Truth-telling by Third-party Auditors and the Response of Polluting Firms: Experimental Evidence from India*," *The Quarterly Journal of Economics*, November 2013, *128* (4), 1499–1545.

Fisman, Raymond and Yongxiang Wang, "The Mortality Cost of Political Connections," *The Review of Economic Studies*, October 2015, *82* (4), 1346–1382.

Foster, Andrew, Emilio Gutierrez, and Naresh Kumar, "Voluntary Compliance, Pollution Levels, and Infant Mortality in Mexico," *American Economic Review*, April 2009, *99* (2), 191–197.

Fowlie, Meredith, Edward Rubin, and Reed Walker, "Bringing Satellite-Based Air Quality Estimates Down to Earth," 2019, p. 12.

Ghanem, Dalia and Junjie Zhang, "'Effortless Perfection:' Do Chinese cities manipulate air pollution data?," *Journal of Environmental Economics and Management*, September 2014, *68* (2), 203–225.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," Technical Report w25018, National Bureau of Economic Research, Cambridge, MA September 2018.

Grainger, Corbett, Andrew Schreiber, and Wonjun Chang, "Do Regulators Strategically Avoid Pollution Hotspots when Siting Monitors? Evidence from Remote Sensing of Air Pollution," 2019, p. 51.

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

Greenstone, Michael, Guojun He, Ruixue Jia, and Tong Liu, "Can Technology Solve the Principal-Agent Problem? Evidence from China's War on Air Pollution," *SSRN Electronic Journal*, 2020.

_ , _ , Shanjun Li, and Eric Zou, "China's War on Pollution: Evidence from the First Five Years," *Review of Environmental Economics and Policy*, 2020.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend, "High-

Resolution Global Maps of 21st-Century Forest Cover Change," *Science*, November 2013, *342* (6160), 850–853.

He, Guojun, Shaoda Wang, and Bing Zhang, "Watering Down Environmental Regulation in China*," *The Quarterly Journal of Economics*, November 2020, *135* (4), 2135–2185.

Hoff, Raymond M. and Sundar A. Christopher, "Remote Sensing of Particulate Pollution from Space: Have We Reached the Promised Land?," *Journal of the Air & Waste Management Association*, June 2009, *59* (6), 645–675.

Jia, Ruixue, "Pollution for Promotion," *21st Century China Center Research Paper*, 2017, pp. No. 2017–05.

_ **and Huihua Nie**, "Decentralization, Collusion, and Coal Mine Deaths," *Review of Economics and Statistics*, March 2017, *99* (1), 105–118.

Kahn, Matthew E, "Domestic pollution havens: evidence from cancer deaths in border counties," *Journal of Urban Economics*, July 2004, *56* (1), 51–69.

Kahn, Matthew E. and Erin T. Mansur, "Do local energy prices and regulation affect the geographic concentration of employment?," *Journal of Public Economics*, May 2013, *101*, 105–114.

_ **, Pei Li, and Daxuan Zhao**, "Water Pollution Progress at Borders: The Role of Changes in China's Political Promotion Incentives," *American Economic Journal: Economic Policy*, November 2015, *7* (4), 223–242.

Karplus, Valerie J., Shuang Zhang, and Douglas Almond, "Quantifying coal power plant responses to tighter SO_2 emissions standards in China," *Proceedings of the National Academy of Sciences*, July 2018, *115* (27), 7004–7009.

Laffont, Jean-Jacques and Jean Tirole, "A theory of incentives in procurement and regulation.," *MIT press.*, 1993.

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

Li, Hongbin and Li-An Zhou, "Political turnover and economic performance: the incentive role of personnel control in China," *Journal of Public Economics*, September 2005, *89* (9-10), 1743–1762.

Liu, Yang, Meredith Franklin, Ralph Kahn, and Petros Koutrakis, "Using aerosol optical thickness to predict ground-level PM2.5 concentrations in the St. Louis area: A comparison between MISR and MODIS," *Remote Sensing of Environment*, March 2007, *107* (1-2), 33–44.

Niu, XueJiao, XiaoHu Wang, Jie Gao, and XueJun Wang, "Has third-party monitoring improved environmental data quality? An analysis of air pollution data in China," *Journal of Environmental Management*, January 2020, *253*, 109698.

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

Ruminski, Mark, Shobha Kondragunta, Roland Draxler, and Glenn Rolph, "Use of environmental satellite imagery for smoke depiction and transport model initialization," 01 2007.

Shimshack, Jay P., "The Economics of Environmental Monitoring and Enforcement," *Annual Review of Resource Economics*, November 2014, *6* (1), 339–360.

Sullivan, Daniel M and Alan Krupnick, "Using Satellite Data to Fill the Gaps in the US Air Pollution Monitoring Network," 2018, p. 33.

Turner, W., C. Rondinini, N. Pettorelli, B. Mora, A.K. Leidner, Z. Szantoi, G. Buchanan, S. Dech, J. Dwyer, M. Herold, L.P. Koh, P. Leimgruber, H. Taubenboeck, M. Wegmann, M. Wikelski, and C. Woodcock, "Free and open-access satellite data are key to biodiversity conservation," *Biological Conservation*, February 2015, *182*, 173–176.

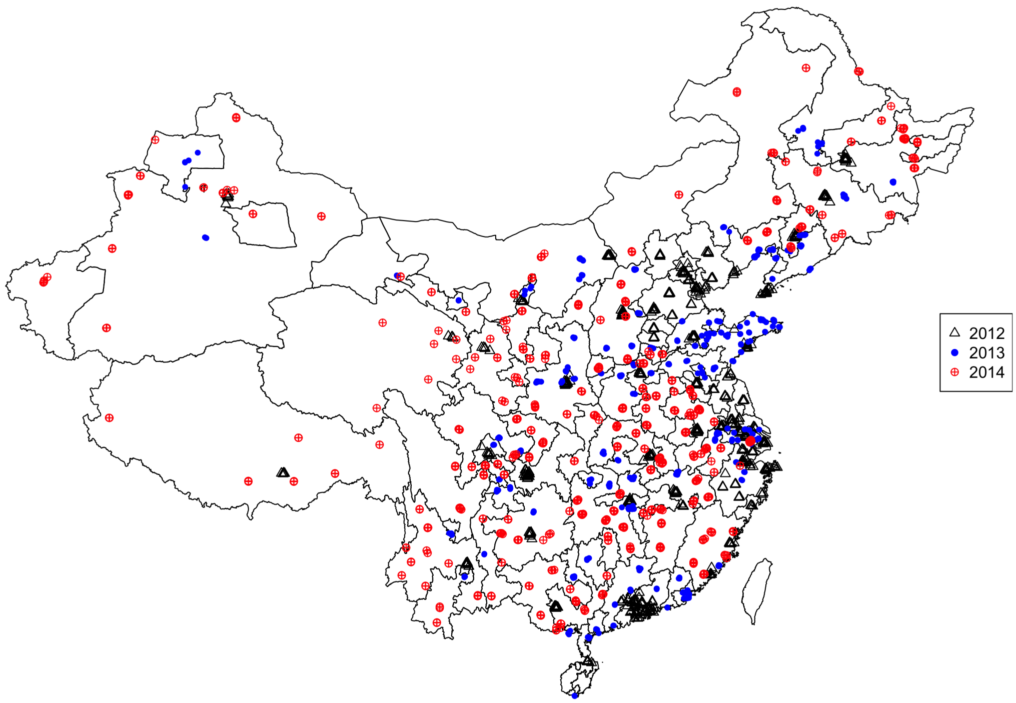
van Donkelaar, Aaron, Randall V. Martin, Chi Li, and Richard T. Burnett, "Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors," *Environmental Science & Technology*, March 2019, *53* (5), 2595–2611.

_ , _ , **Michael Brauer, N. Christina Hsu, Ralph A. Kahn, Robert C. Levy, Alexei Lyapustin, Andrew M. Sayer, and David M. Winker**, "Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors," *Environmental Science & Technology*, 2016, *50* (7), 3762–3772.

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

Zou, Eric, "Unwatched Pollution: The Effect of Intermittent Monitoring on Air Quality," *American Economic Review*, 2020.

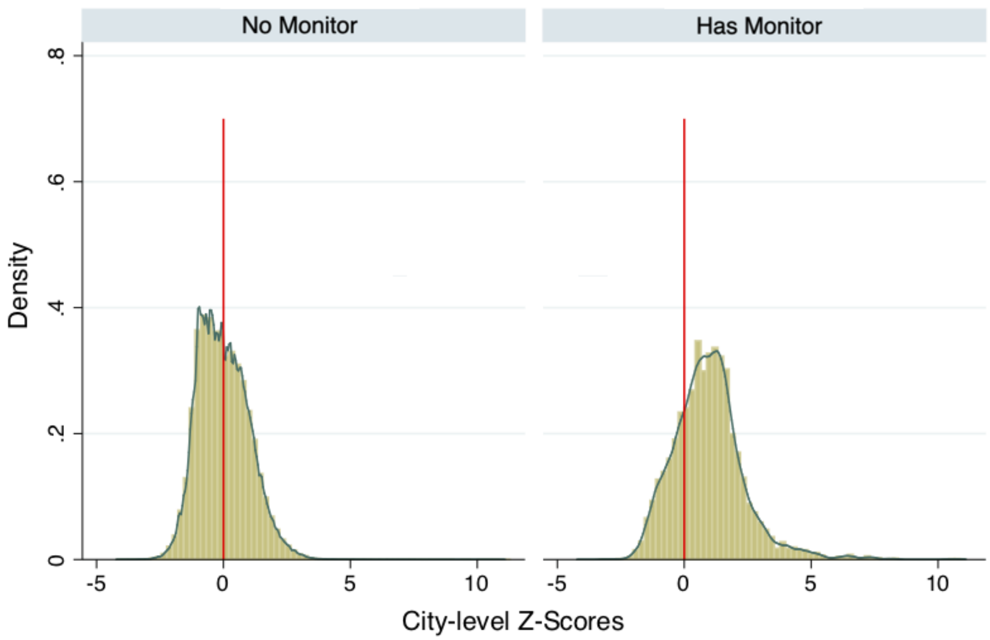
Figure 1: Roll-out of Monitoring Stations in China



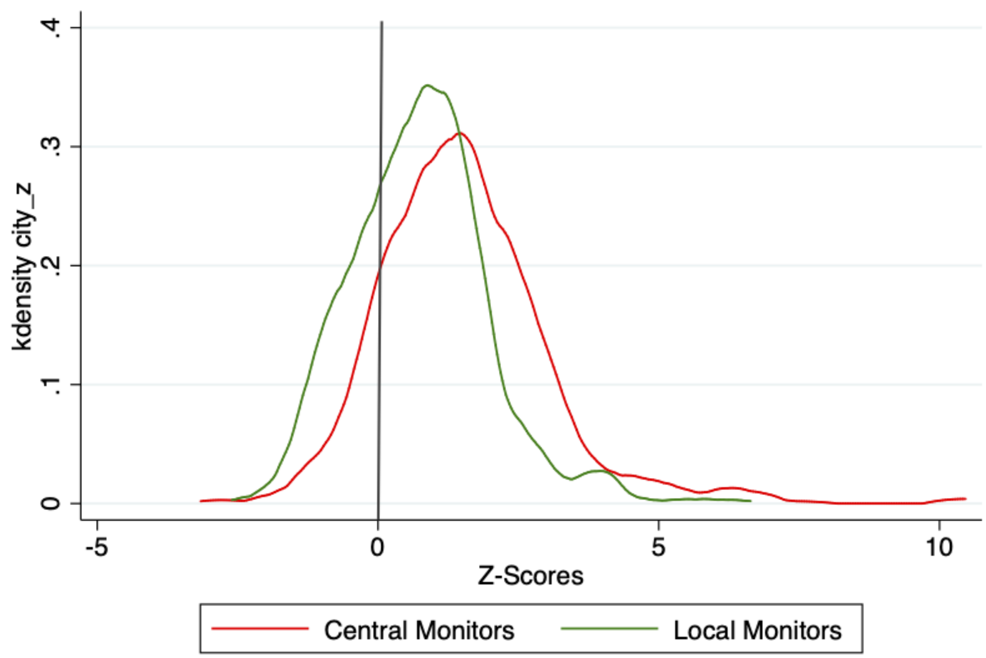
Note: This figure shows the roll-out of air pollution monitoring stations in China by three waves from 2012 to 2014. All monitors on the map are central government-controlled monitors.

Figure 2: Kernel Densities for PM2.5 Z-Scores

(a) Kernel Densities for PM2.5 Z-Scores: Central monitors vs. No monitor

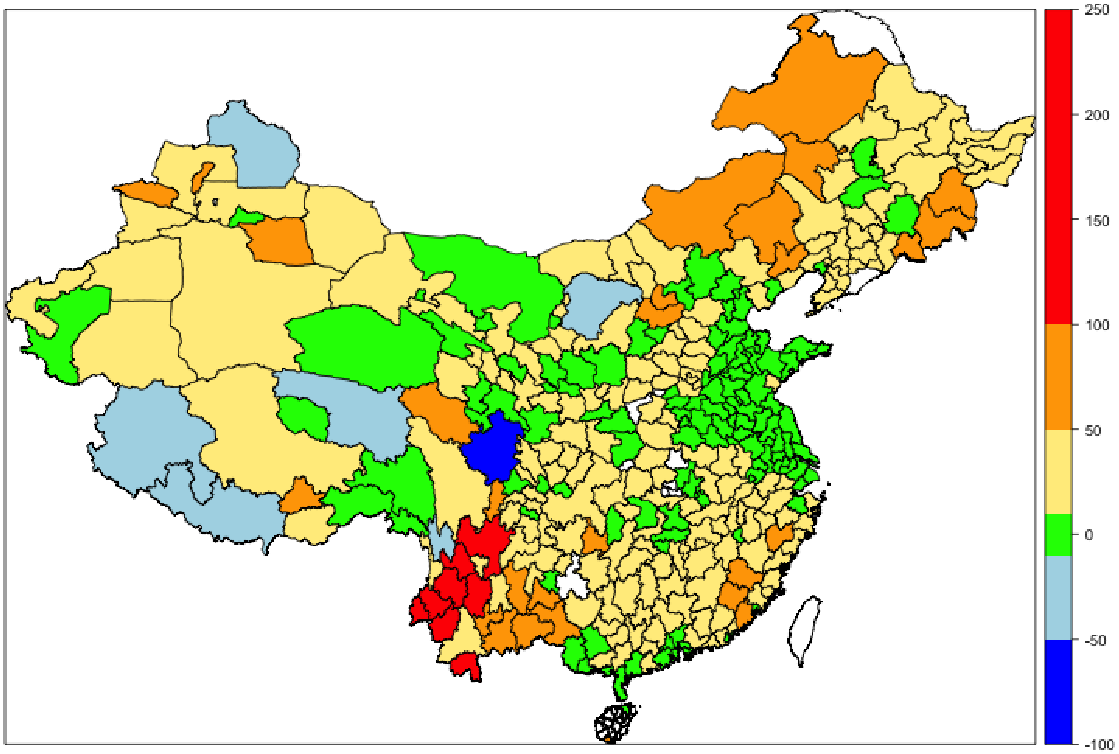


(b) Kernel Densities for PM2.5 Z-Scores: Central monitors vs. Local monitors



Note: Each figure shows the kernel density estimate for the distribution of city-level z-scores. Z-score is calculated by taking the observed value in grid cell i in city c and year t, subtract the average for that city, and scale it by the city level standard deviation. Figure (a) compares the distribution of city-level z-scores at cells containing central monitors to cells without monitors using data from 2009 to 2017. Figure (b) compares the distribution of city-level z-scores at cells containing central monitors to cells containing local monitors in 2016.

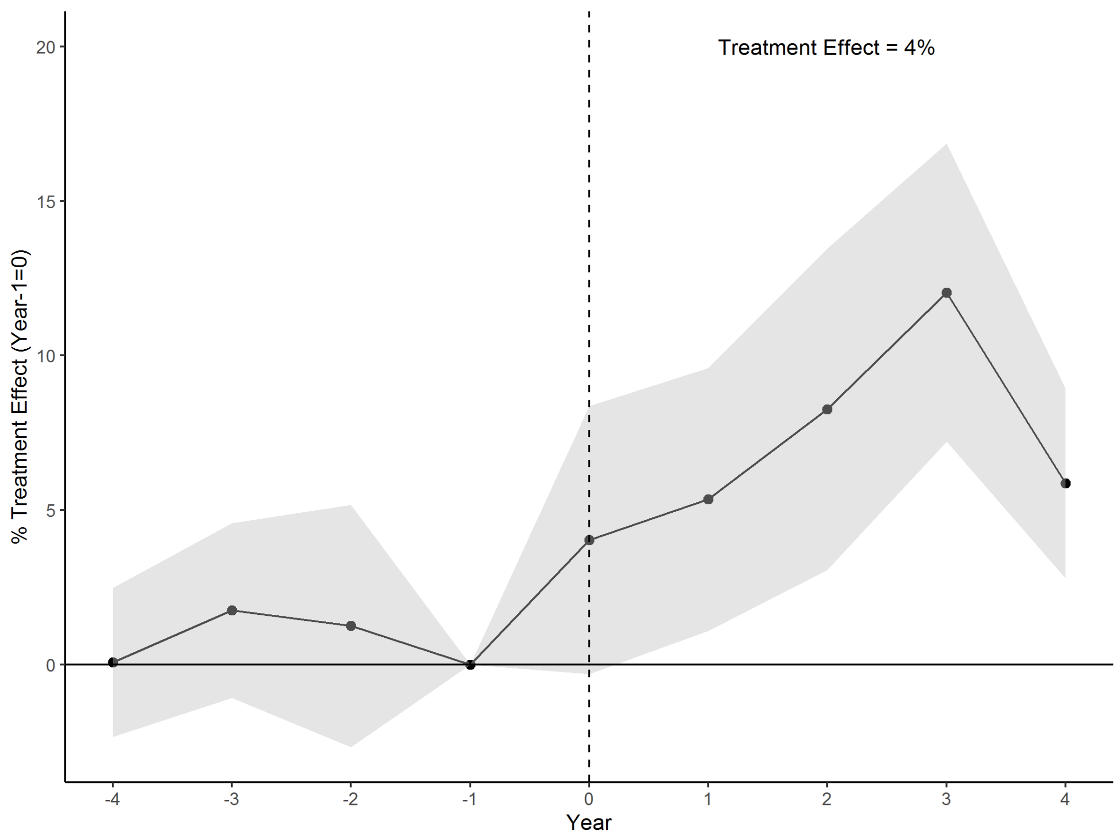
Figure 3: Monitor Representation Errors at Opening Years: All Cells vs. Monitored Cells



Error=(monitor avg-city avg)%
Error > 0: Monitors Over-Estimate City Avg Pollution;
Error < 0: Monitors Under-Estimate City Avg Pollution

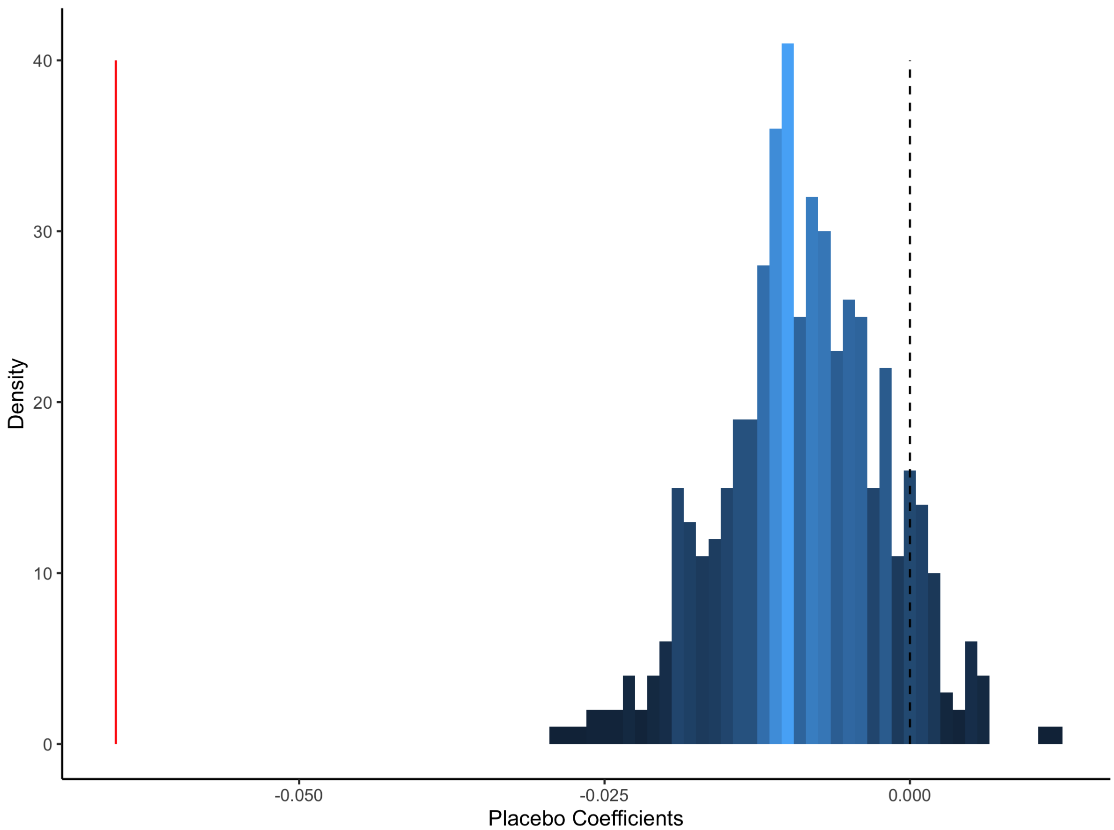
Note: This figure shows the monitors representation errors in the years of joining the new monitoring program. The representation error is defined as the percentage difference between city average pollution level calculated based on only monitored cells and city average pollution based on all cells. All the pollution levels are weighted by the 2015 grid-level population count. Cities in green means the monitors well-represent city average PM2.5, with representation errors in $[-10\%, 10\%]$. Cities in warm colors (error $> 10\%$) have monitors over-representing the city average pollution, and those in cool colors (error $< -10\%$) meansthat the monitors under-present city average pollution level. The map is based on raw data and presented at the city level. Representation error maps for each year from 2012 to 2017 are in Appendix A1.

Figure 4: Event Study of Monitor Opening on Air Pollution



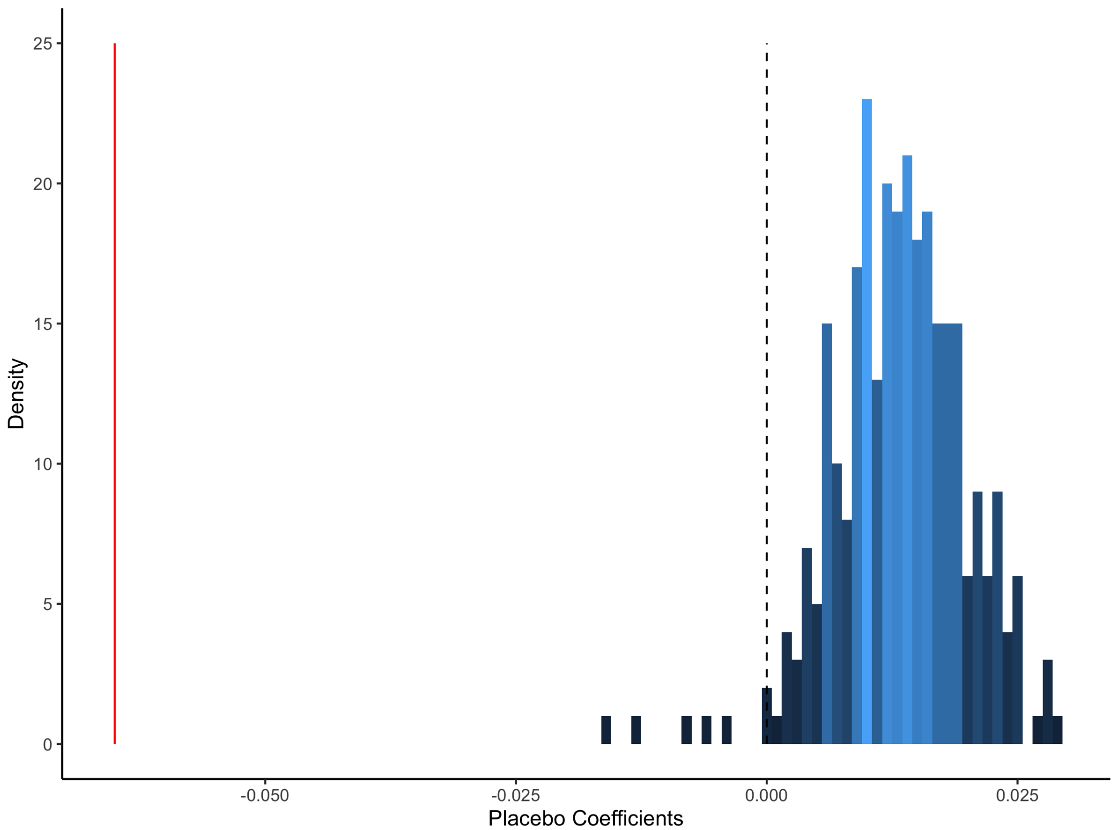
Note: This figure shows the event study results of monitor opening on air pollution controlling for cell fixed effects, year fixed effects, and wave-specific time trend. I regress the PM2.5 on four pre-opening indicators and four post-opening indicators. The year before monitoring is the base interval. Standard errors are clustered at city level.

Figure 5: Placebo Test with Random Opening Years in Pre-Monitoring Periods



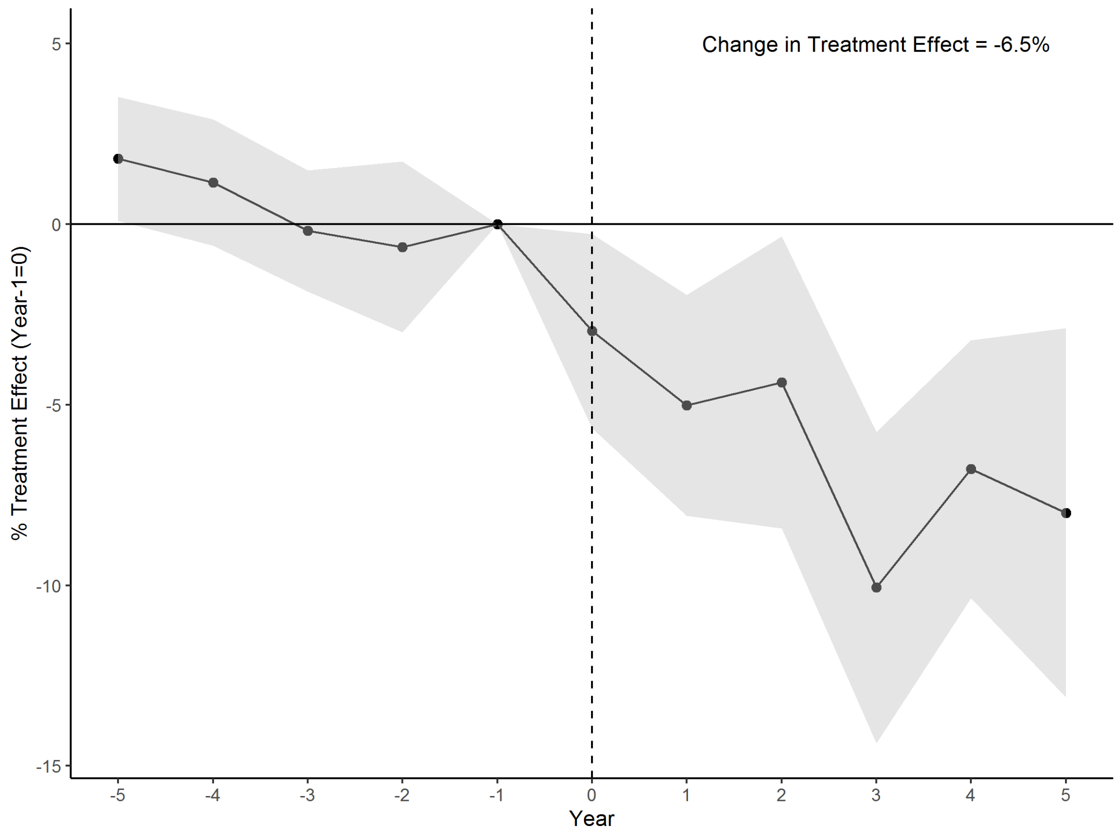
Note: This figure shows the results of a “placebo” test using only pre-program periods and randomly assign each monitor an opening year. I conduct 500 estimations of the treatment intensity analysis and plot the distribution of the 500 placebo coefficients and compare them with the observed effect size using the real sample (red line).

Figure 6: Placebo Test with Random Monitor Locations



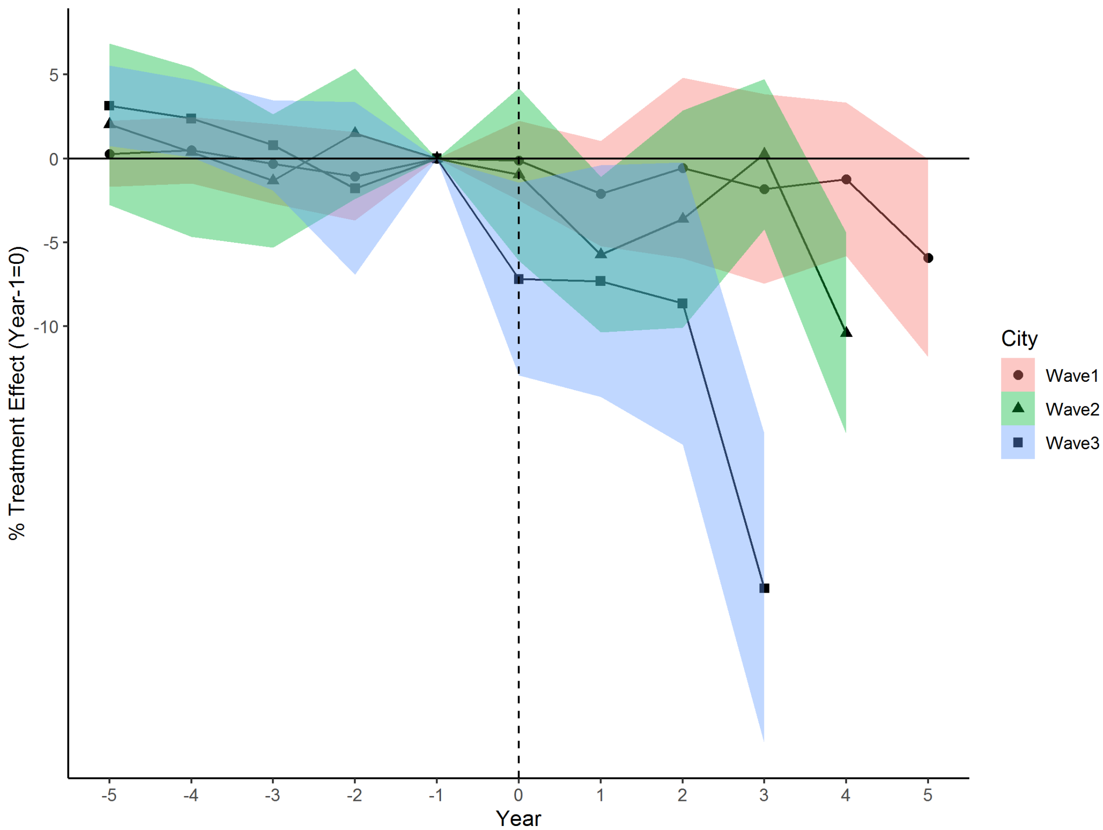
Note: This figure shows the results of a "placebo" test that conducts 500 randomly relocations of all monitors within a city and keep the opening year unchanged. I conduct 500 estimations of equation (2) and plot the distribution of the 500 placebo coefficients, and compare them with the observed effect size using the real sample (red line).

Figure 7: Event study: Change in Impact of Monitoring on Air Pollution in Monitored vs. Unmonitored Areas



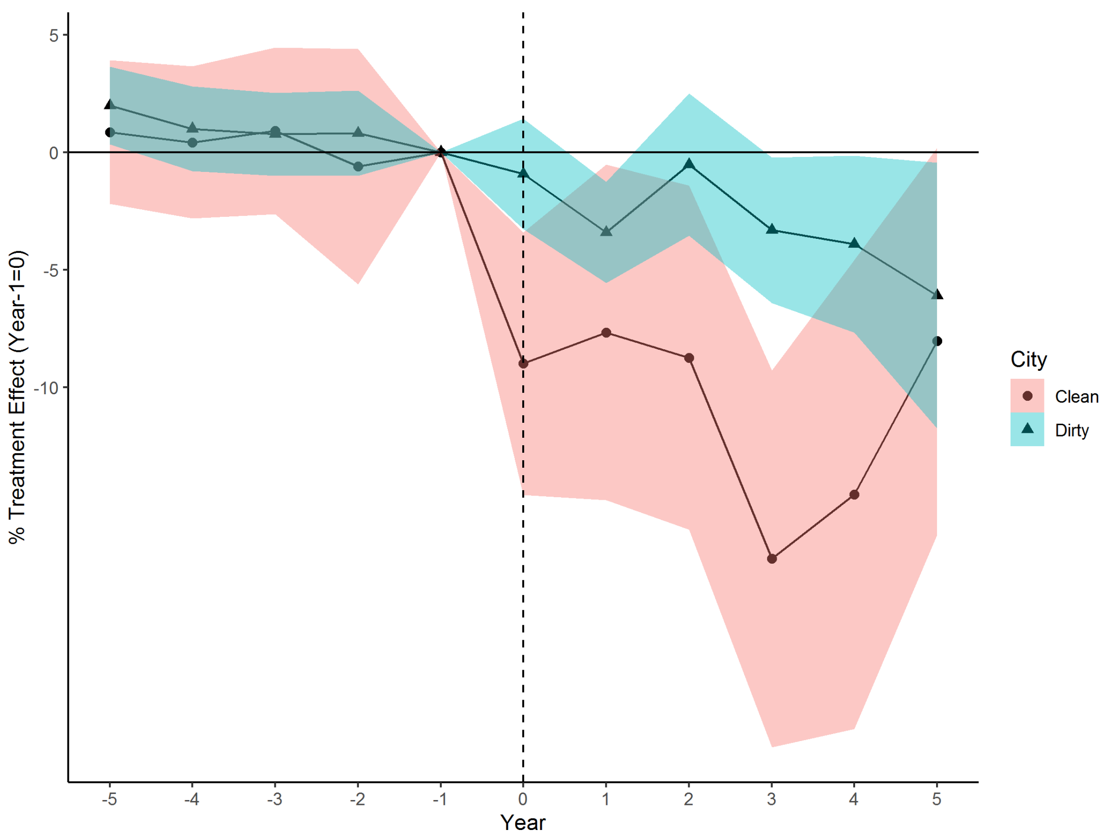
Note: This figure shows the event study results of monitor opening with treatment intensity on air pollution. (Column 1 from Table 4), where I regress the PM2.5 on interactions of treatment intensity indicator Near, and five pre-opening indicators and six post-opening indicators, controlling for cell fixed effects, and wave by year fixed effects. The year before monitoring is the base interval. Each estimate represents the difference in PM2.5 between monitored areas (cells within 3km of monitors) and unmonitored areas (cells outside 3km) at a given period. Standard errors are clustered at city level.

Figure 8: Heterogeneity Analysis by Waves: Change in Impact of Monitoring on Air pollution in Monitored vs. Unmonitored Areas



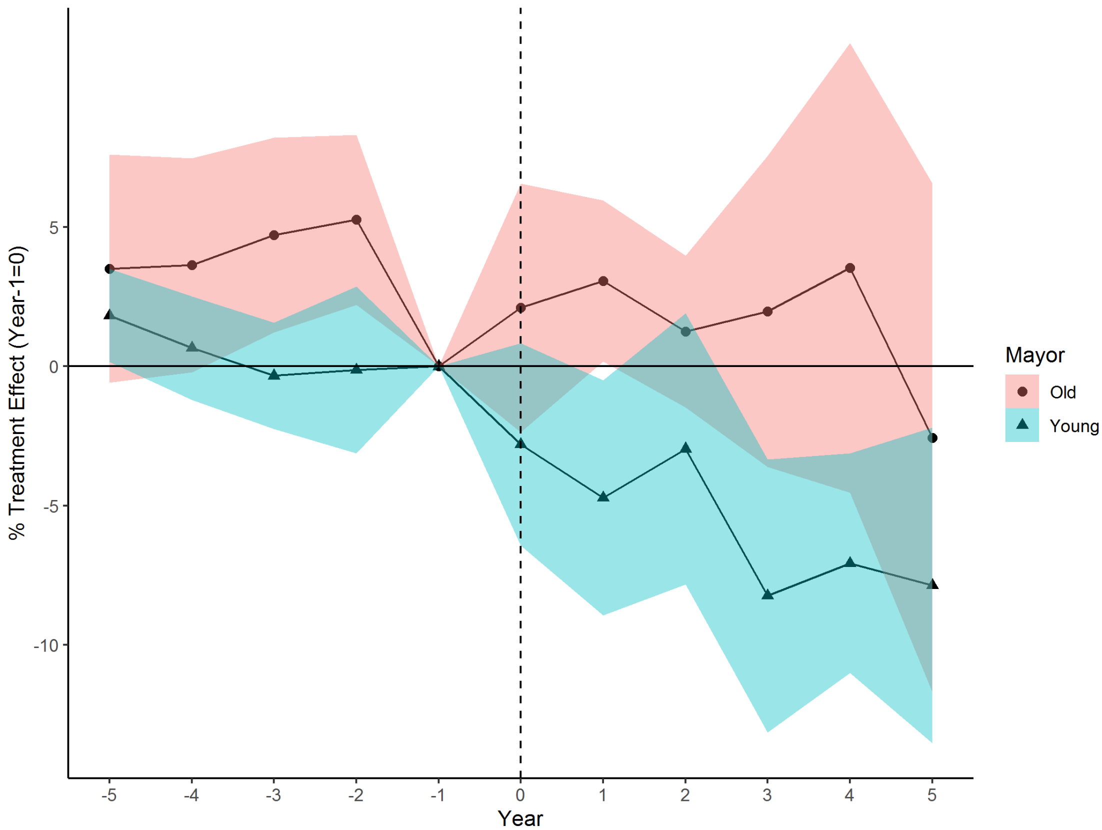
Note: This figure shows the event study results of monitor opening with treatment intensity on air pollution for three subsamples divided by roll-out waves. Using each city group, I regress the PM2.5 on interactions of treatment intensity indicator Near, and pre-opening and post-opening indicators, controlling for cell and year fixed effects. The year before monitoring is the base interval. Each estimate represents the difference in PM2.5 between monitored areas (cells within 3km of monitors) and unmonitored areas (cells outside 3km) at a given period. Standard errors are clustered at city level.

Figure 9: Heterogeneity Analysis by City Average Pollution Level: Change in Impact of Monitoring on Air Pollution in Monitored vs. Unmonitored Areas



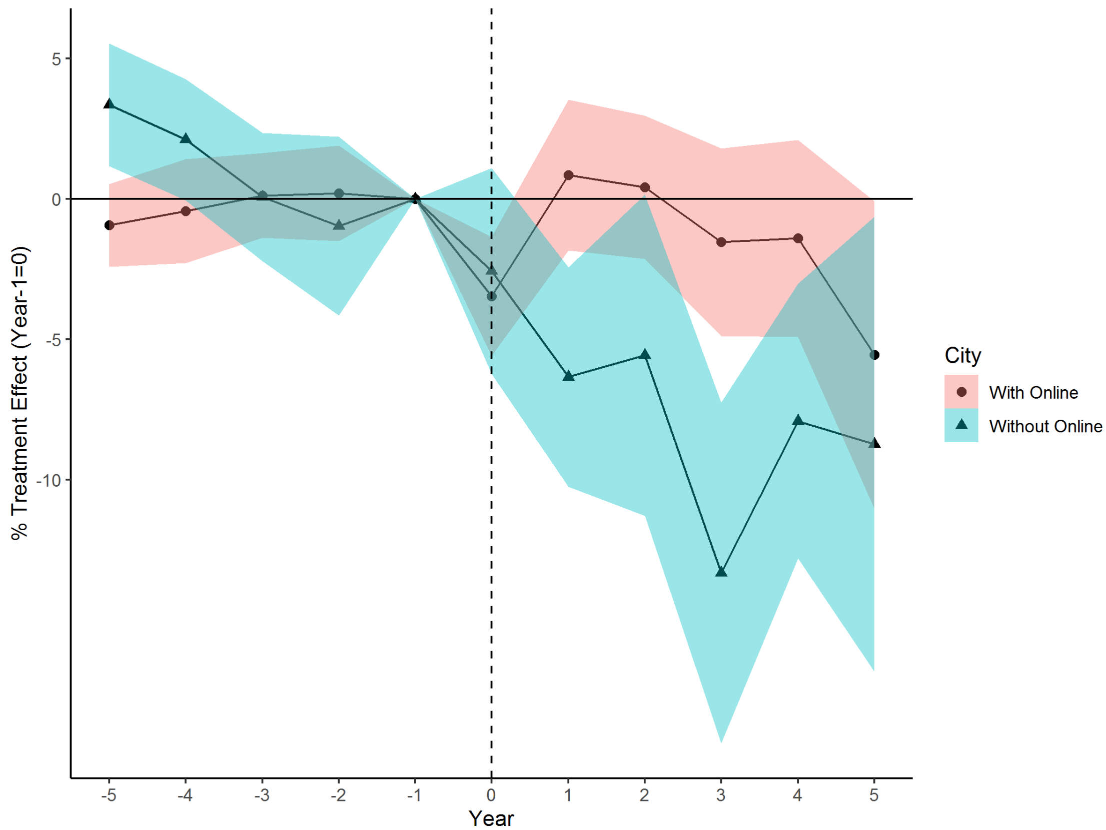
Note: This figure shows the event study results of monitor opening with treatment intensity on air pollution, for two subsamples of cities classified by comparing the cities' average PM2.5 level with national annual standard, 35 $\mu g/m^3$. Using each city group, I regress the PM2.5 on interactions of treatment intensity indicator Near, and five pre-opening indicators and six post-opening indicators, controlling for cell fixed effects, and wave by year fixed effects. The year before monitoring is the base interval. Each estimate represents the difference in PM2.5 between monitored areas (cells within 3km of monitors) and unmonitored areas (cells outside 3km) at a given period. Standard errors are clustered at city level.

Figure 10: Heterogeneity Analysis by City Mayors' Age: Change in Impact of Monitoring on Air Pollution in Monitored vs. Unmonitored Areas



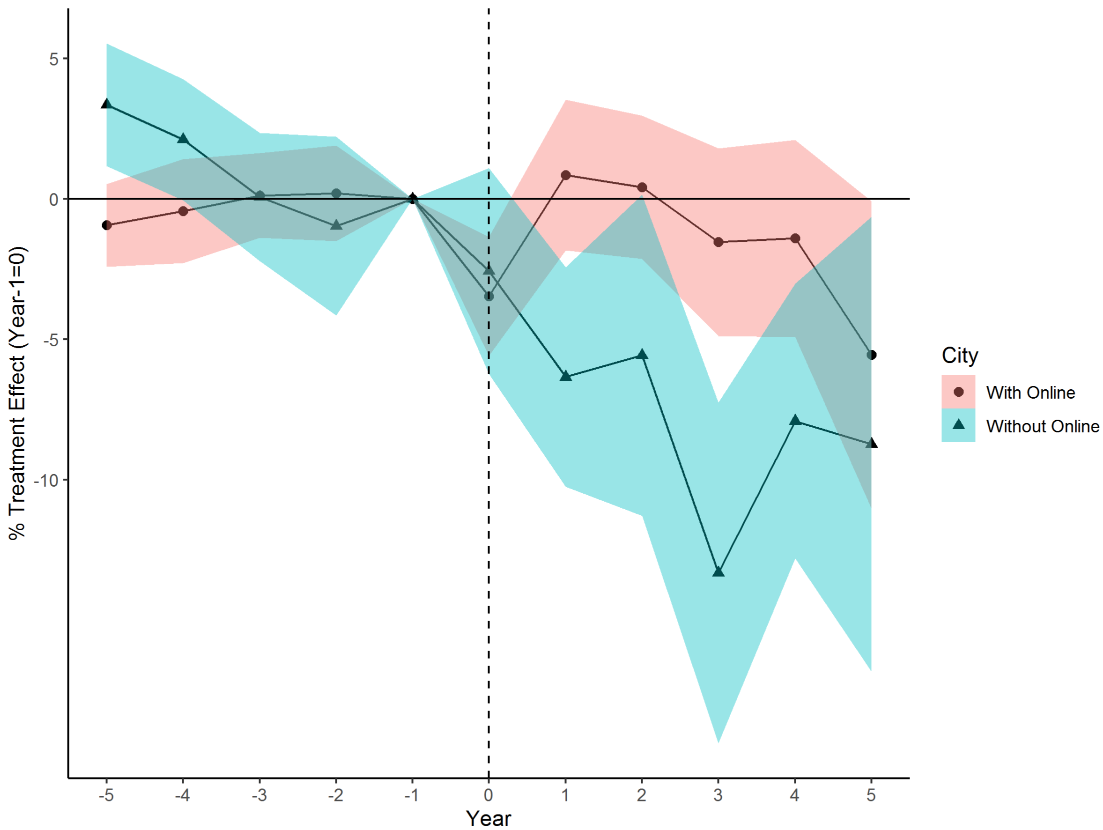
Note: This figure shows the event study results of monitor opening with treatment intensity on air pollution, for two subsamples of cities classified by city mayors' age. The cutoff point for mayor's age is 57 because this is the ceiling threshold for a mayor to get promoted. Using each city group, I regress the PM2.5 on interactions of treatment intensity indicator Near, and five pre-opening indicators and six post-opening indicators, controlling for cell fixed effects, and wave by year fixed effects. The year before monitoring is the base interval. Each estimate represents the difference in PM2.5 between monitored areas (cells within 3km of monitors) and unmonitored areas (cells outside 3km) at a given period. Standard errors are clustered at city level.

Figure 11: Heterogeneity Analysis by Province Online Disclosure: Change in Impact of Monitoring on Air Pollution in Monitored vs. Unmonitored Areas



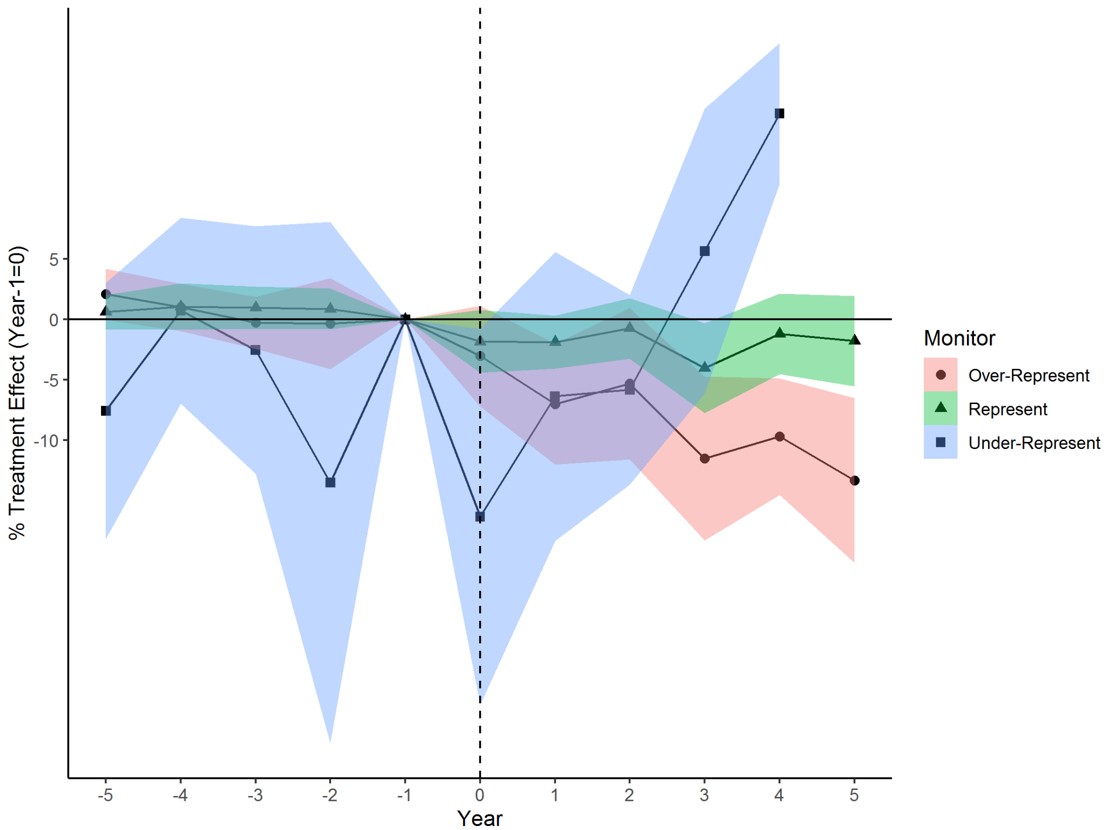
Note: This figure shows the event study results of monitor opening with treatment intensity on air pollution, for two subsamples divided by whether a province has its own online pollution disclosure platform or not. Using each city group, I regress the PM2.5 on interactions of treatment intensity indicator Near, and five pre-opening indicators and six post-opening indicators, controlling for cell fixed effects, and wave by year fixed effects. The year before monitoring is the base interval. Each estimate represents the difference in PM2.5 between monitored areas (cells within 3km of monitors) and unmonitored areas (cells outside 3km) at a given period. Standard errors are clustered at city level.

Figure 12: Heterogeneity Analysis by Province Online Disclosure: Change in Impact of Monitoring on Air Pollution in Monitored vs. Unmonitored Areas



Note: This figure shows the event study results of monitor opening with treatment intensity on air pollution, for two subsamples divided by whether a province has its own online pollution disclosure platform or not. Using each city group, I regress the PM2.5 on interactions of treatment intensity indicator Near, and five pre-opening indicators and six post-opening indicators, controlling for cell fixed effects, and wave by year fixed effects. The year before monitoring is the base interval. Each estimate represents the difference in PM2.5 between monitored areas (cells within 3km of monitors) and unmonitored areas (cells outside 3km) at a given period. Standard errors are clustered at city level.

Figure 13: Heterogeneity Analysis by Monitor Representativeness: the change in impact of monitoring on air pollution in monitored vs. unmonitored areas



Note: This figure shows the event study results of monitor opening with treatment intensity on air pollution, for three subsamples divided by monitors spatial representativeness at the years of opening. Representation errors are defined as the difference between population-weighted city average PM at monitored cells and at all cells. Over-represent cities have representation errors greater than 10%. Well-represent cities have error between -10% to 10%. Under-represent cities are with errors less than -10%. Using each city group, I regress the PM2.5 on interactions of treatment intensity indicator Near, and five pre-opening indicators and six post-opening indicators, controlling for cell fixed effects, and wave by year fixed effects. The year before monitoring is the base interval. Each estimate represents the difference in PM2.5 between monitored areas (cells within 3km of monitors) and unmonitored areas (cells outside 3km) at a given period. Standard errors are clustered at city level.

Table 1: Summar Statistics: Satellite-based Air Pollution (PM2.5, $\mu g/m^3$)

	Wave 1		Wave 2		Wave 3	
	(1)	(2)	(3)	(4)	(5)	(6)
	city_avg	city_monavg	city_avg	city_monavg	city_avg	city_monavg
2009	54.25	59.26	49.23	56.48	40.02	48.10
	(17.24)	(18.35)	(17.23)	(17.21)	(19.98)	(21.31)
2010	53.10	57.98	50.23	57.65	41.34	49.22
	(17.44)	(18.46)	(19.93)	(19.88)	(22.90)	(23.84)
2011	50.51	55.35	47.15	54.34	38.15	45.99
	(17.54)	(18.66)	(18.26)	(18.57)	(20.40)	(21.47)
2012	46.91	51.45	44.97	52.08	36.40	44.13
	(16.20)	(17.56)	(18.35)	(18.73)	(19.51)	(20.78)
2013	54.66	59.73	51.55	58.92	41.30	49.34
	(20.81)	(22.16)	(21.38)	(21.57)	(22.56)	(23.74)
2014	55.31	60.29	50.58	57.97	41.83	50.48
	(18.73)	(19.81)	(19.14)	(18.89)	(22.85)	(24.03)
2015	51.67	56.54	47.92	54.45	37.85	44.87
	(18.51)	(19.68)	(18.25)	(18.37)	(19.63)	(20.25)
2016	46.48	51.33	43.37	51.23	34.32	42.41
	(17.71)	(18.59)	(17.33)	(17.83)	(18.65)	(20.53)
2017	52.48	56.34	47.98	54.73	40.57	47.45
	(15.51)	(15.81)	(15.77)	(15.84)	(19.01)	(20.14)

Notes: The underlying observations are at the city level. Standard deviations are in parentheses. Column (1), (3), (5) show population-weighted PM2.5, column (2), (4), (6) show population-weighted PM2.5 level at monitored cells are average of post-monitoring period. Population data is in 2015.

Table 2: Summar Statistics: Other Variables

	(1)	(2)	(3)
Variable	Wave1	Wave2	Wave3
City Pollution, GDP, Population			
Population Weighted PM2.5	51.71	48.11	39.09
	(17.95)	(18.55)	(20.76)
Population Weighted PM2.5 at Monitored Cells	55.87	55.48	46.55
	(19.27)	(18.67)	(21.60)
GDP Per Capita	63944	48187	30641
	(30064)	(31222)	(16396)
GDP in 3rd Industry	236.72	67.06	31.62
	(300.22)	(76.67)	(21.17)
GDP in 2nd Industry	204.54	94.08	44.45
	(173.30)	(89.73)	(30.89)
GDP in 1st Industry	20.52	16.20	14.79
	(15.71)	(10.53)	(9.88)
Population in 2015	4857550	2716721	1940899
	(3499542)	(1531311)	(1341832)
Leader's Characteristics			
Age	51.88	50.09	50.00
	(4.66)	(3.65)	(3.57)
Young (Age<57)	.930	.995	.984
	(.256)	(.070)	(.127)
Master	.497	.572	.540
	(.501)	(.495)	(.499)
PhD	.269	.218	.158
	(.444)	(.414)	(.364)
Bachelor	.2104121	.193	.261
	(.408)	(.395)	(.440)
Number of Cities	74	98	176

Notes: The underlying observations are at the city level. Standard deviations are in parentheses. Population-weighted PM2.5 are measured by 2009-2017 average, PM2.5 level at monitored cells are average of post-monitoring period. GDP data is from 2001-2017 for 281 cities. Leader's characteristics data ranges from 2009-2015.

Table 3: Baseline Difference in Differences Estimation Results

	Dependent variable: $\ln(PM2.5_{it})$					
	(1)	(2)	(3)	(4)	(5)	(6)
Open	0.188***	0.106***	0.048**	0.040**	0.049**	0.041**
	(0.038)	(0.026)	(0.022)	(0.020)	(0.022)	(0.020)
(0-3km)*Open					-0.100***	-0.062***
					(0.024)	(0.014)
Controls	No	Cell FE	Cell FE Year FE	Cell FE Year FE Wave×T	Cell FE Year FE	Cell FE Year FE Wave×T
Observations	84,349,384	84,349,384	84,349,384	84,349,384	83,293,774	83,293,774
R^2	0.009	0.958	0.965	0.966	0.965	0.966

Note: Column (1)-(6) show DID estimation results with different fixed effects. The first three columns represent baseline DID results, where *Open* is the treatment indicator that equals one if a cell is in a city that has joined the new monitoring program. Column (4)-(6) show DID estimation results with treatment intensity defined by distances to monitors. Cells within 3km to the monitor are in the monitored group and cells outside 3km are unmonitored cells. Standard errors are clustered at city level. Significance: *p<0.1; **p<0.05; ***p<0.01

Table 4: Baseline DID Estimation Results, Alternative Unmonitored Areas

	Dependent variable: $\ln(PM2.5_{it})$					
Unmonitored Area:	>3km	>30km	>50km	>15km	>30km	>50km
	(1)	(2)	(3)	(4)	(5)	(6)
Open	0.041**	0.043*	0.043*	0.044**	0.046**	0.049**
	(0.020)	(0.022)	(0.025)	(0.021)	(0.022)	(0.024)
(0-3km)*Open	-0.062***	-0.068***	-0.076***	-0.066***	-0.070***	-0.079***
	(0.014)	(0.016)	(0.017)	(0.015)	(0.016)	(0.018)
(3-6km)*Open				-0.065***	-0.069***	-0.078***
				(0.014)	(0.015)	(0.017)
(6-9km)*Open				-0.066***	-0.071***	-0.079***
				(0.014)	(0.015)	(0.017)
(9-12km)*Open				-0.066***	-0.071***	-0.080***
				(0.014)	(0.015)	(0.017)
(12-15km)*Open				-0.067***	-0.072***	-0.081***
				(0.014)	(0.015)	(0.017)
Observations	83,293,774	74,496,330	65,280,883	83,293,774	77,485,464	68,270,017
R^2	0.966	0.966	0.966	0.966	0.966	0.966

Note: Column (1)-(6) show DID estimation results. *Open* is the treatment indicator that equals one if a cell is in a city that has joined the new monitoring program. The first three columns use cells within 3km to the monitor as the monitored group and compare different unmonitored groups: cells outside 3km, 30km or 50km of the monitors. Column (4)-(6) add four more distance bins to the monitored group. All columns include cell fixed effects, year fixed effects and a wave specific time trend. Standard errors are clustered at city level. Significance: *p<0.1; **p<0.05; ***p<0.01

Table 5: Difference in Differences with Alternative Treatment Intensity Bins Estimation Results

	Dependent variable: $\ln(PM2.5_{it})$					
Unmonitored Area:	>3km	>30km	>50km	>15km	>30km	>50km
	(1)	(2)	(3)	(4)	(5)	(6)
(0-3km)*Open	-0.065***	-0.072***	-0.079***	-0.069***	-0.074***	-0.083***
	(0.013)	(0.015)	(0.016)	(0.014)	(0.015)	(0.017)
(3-6km)*Open				-0.068***	-0.073***	-0.082***
				(0.014)	(0.015)	(0.016)
(6-9km)*Open				-0.069***	-0.074***	-0.083***
				(0.014)	(0.014)	(0.016)
(9-12km)*Open				-0.069***	-0.074***	-0.084***
				(0.014)	(0.015)	(0.016)
(12-15km)*Open				-0.070***	-0.075***	-0.084***
				(0.014)	(0.015)	(0.016)
Observations	83,293,774	74,496,330	65,280,883	83,293,774	77,485,464	68,270,017
R^2	0.967	0.966	0.966	0.967	0.967	0.967

Note: Column (1)-(6) show DID estimation results with treatment intensity defined by distances to monitors. *Open* is the treatment indicator that equals one if a cell is in a city that has joined the new monitoring program. The first three columns use cells within 3km to the monitor as the monitored group and compare different unmonitored groups: cells outside 3km, 30km or 50km of the monitors. Column (4)-(6) add four more distance bins to the monitored group. All columns include both the cell FE and Wave×Year FE. Standard errors are clustered at city level. Significance: *p<0.1; **p<0.05; ***p<0.01

Table 6: Event Study with Alternative Treatment Intensity Groups Estimation Results

	Dependent variable: $\ln(PM2.5_{it})$					
Monitored Area:	$\leq$3km			$\leq$10km		
Unmonitored Area:	>3km	>30km	>50km	>10km	>30km	>50km
	(1)	(2)	(3)	(4)	(5)	(6)
Near*(y-5)	0.018**	0.021**	0.025**	0.022**	0.024**	0.028***
	(0.009)	(0.009)	(0.010)	(0.009)	(0.009)	(0.010)
Near*(y-4)	0.012	0.013	0.016	0.014	0.015	0.017
	(0.009)	(0.010)	(0.011)	(0.009)	(0.010)	(0.011)
Near*(y-3)	-0.002	-0.003	-0.007	-0.001	-0.003	-0.006
	(0.009)	(0.010)	(0.012)	(0.009)	(0.010)	(0.011)
Near*(y-2)	-0.006	-0.005	-0.008	-0.010	-0.008	-0.012
	(0.012)	(0.013)	(0.015)	(0.012)	(0.013)	(0.015)
Near*(y0)	-0.030**	-0.030**	-0.033*	-0.033**	-0.033**	-0.037**
	(0.014)	(0.015)	(0.017)	(0.014)	(0.015)	(0.017)
Near*(y+1)	-0.050***	-0.054***	-0.061***	-0.049***	-0.052***	-0.059***
	(0.016)	(0.017)	(0.019)	(0.016)	(0.017)	(0.019)
Near*(y+2)	-0.044**	-0.044*	-0.051*	-0.049**	-0.049**	-0.056**
	(0.021)	(0.024)	(0.028)	(0.021)	(0.023)	(0.027)
Near*(y+3)	-0.101***	-0.108***	-0.121***	-0.105***	-0.111***	-0.126***
	(0.022)	(0.025)	(0.029)	(0.022)	(0.025)	(0.028)
Near*(y+4)	-0.068***	-0.080***	-0.095***	-0.074***	-0.084***	-0.100***
	(0.018)	(0.022)	(0.026)	(0.017)	(0.020)	(0.024)
Near*(y+5)	-0.080***	-0.098***	-0.118**	-0.080***	-0.095***	-0.116***
	(0.026)	(0.034)	(0.046)	(0.026)	(0.033)	(0.045)
Observations	87,843,991	77,211,912	67,374,686	87,843,991	78,853,329	69,016,103
R^2	0.966	0.966	0.967	0.966	0.967	0.967

Note: Column (1)-(6) show event study results with different treatment intensity groups. *Near* is the monitored area indicator which equals one for cells within 3km from monitors in column (1)-(3), and 10km from monitors in column (4)-(6). y-5, y-4,...,y+5 represent each year within the 5-year time window around monitor openings. For each monitored group, the three columns compare different unmonitored groups: cells outside 3km (10km), 30km or 50km of the monitors. All columns include both the cell FE and Wave×Year FE. Standard errors are clustered at city level. Significance: *p<0.1; **p<0.05; ***p<0.01

Table 7: Heterogeneity Analysis: Difference in Differences with Treatment Intensity Estimation Results

	Dependent variable: $\ln(PM2.5_{it})$					
Unmonitored Area:	>3km	>30km	>50km	>15km	>30km	>50km
	(1)	(2)	(3)	(4)	(5)	(6)
(0-3km)*Open	-0.073***	-0.079***	-0.087***	-0.077***	-0.082***	-0.091***
	(0.015)	(0.017)	(0.019)	(0.016)	(0.017)	(0.019)
(3-6km)*Open				-0.071***	-0.076***	-0.085***
				(0.015)	(0.016)	(0.017)
(6-9km)*Open				-0.069***	-0.074***	-0.083***
				(0.014)	(0.014)	(0.016)
(9-12km)*Open				-0.070***	-0.075***	-0.085***
				(0.013)	(0.014)	(0.016)
(12-15km)*Open				-0.072***	-0.077***	-0.086***
				(0.013)	(0.014)	(0.016)
(0-3km)*Open*Dirtier	0.010	0.010	0.010	0.010	0.010	0.010
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
(3-6km)*Open*Dirtier				0.004	0.005	0.005
				(0.010)	(0.010)	(0.010)
(6-9km)*Open*Dirtier				0.0001	0.0002	0.0003
				(0.012)	(0.012)	(0.012)
(9-12km)*Open*Dirtier				0.002	0.002	0.002
				(0.012)	(0.012)	(0.012)
(12-15km)*Open*Dirtier				0.004	0.004	0.004
				(0.012)	(0.012)	(0.012)
Observations	83,293,774	74,496,330	65,280,883	83,293,774	77,485,464	68,270,017
R^2	0.967	0.966	0.966	0.967	0.967	0.967

Note: Column (1)-(6) show DID estimation results with treatment intensity defined by distances to monitors. *Open* is the treatment indicator that equals one if a cell is in a city that has joined the new monitoring program. *Dirtier* is a dummy variable indicating if the pollution in a cell is above the average city PM2.5. The first three columns use cells within 3km to the monitor as the monitored group and compare different unmonitored groups: cells outside 3km, 30km or 50km of the monitors. Column (4)-(6) add four more distance bins to the monitored group. All columns include both the cell FE and Wave×Year FE. Standard errors are clustered at city level. Significance: *p<0.1; **p<0.05; ***p<0.01

Table 8: Robustness Check Using Raw AOD Data

	Dependent variable: AOD			
Unmonitored Areas:	>3km	>12km	>50km	>50km
	(1)	(2)	(3)	(4)
(0-3km)*Open	-0.021***	-0.022***	-0.033***	-0.034***
	(0.005)	(0.005)	(0.006)	(0.006)
(3-6km)*Open		-0.025***		-0.037***
		(0.005)		(0.006)
(6-9km)*Open		-0.030***		-0.042***
		(0.005)		(0.006)
(9-12km)*Open		-0.034***		-0.046***
		(0.005)		(0.006)
Observations	10,136,285	10,136,285	6,992,163	7,330,347
R^2	0.876	0.876	0.849	0.859

Note: Column (1)-(4) show DID estimation results with treatment intensity defined by distances to monitors. *Open* is the treatment indicator that equals one if a cell is in a city that has joined the new monitoring program. The dependent variable is the annual AOD at 3km by 3km grid cells. Column (1) & (3) use cells within 3km to the monitor as the monitored group and compare different unmonitored groups: cells outside 3km, or 50km of the monitors. Column (2) & (4) add three distance bins to the monitored group and compare two unmonitored groups. All columns include both the cell FE and Wave×Year FE. Standard errors are clustered at city level. Significance: *p<0.1; **p<0.05; ***p<0.01

Table 9: Hereogeneous Analysis: Clean vs. Dirty Cities by Roll-out Waves

	Dependent variable: $\ln(PM2.5_{it})$			
	All	Wave1	Wave2	Wave3
	(1)	(2)	(3)	(4)
(0-3km)*Open	-0.062***	-0.015	-0.052***	-0.133***
	(0.013)	(0.015)	(0.018)	(0.035)
(0-3km)*Open*1(Clean City)	-0.030	-0.085***	0.003	-0.031
	(0.025)	(0.024)	(0.056)	(0.034)
(0-3km)*Open*1(Clean City)*Compliance	-0.004**	-0.008**	-0.009	-0.006***
	(0.002)	(0.003)	(0.012)	(0.002)
Observations	86,844,613	9,856,110	16,784,304	60,204,199
R^2	0.967	0.954	0.954	0.963

Note: Column (1)-(4) show DID estimation results with heterogeneity in treatment effect, separated by roll-out waves. Open is the treatment indicator that equals one if a cell is in a city that has joined the new monitoring program. $\mathbf{1}(CleanCity)$ is a dummy variable indicating cells inside a city with average pollution level (based on monitored cells at the opening years) below the national standard, 35 μ/g^3. *Compliance* represents the closeness to the national standard. All columns use cells within 3km to the monitor as the monitored group and cells outside 3km as unmonitored groups. Column (2)-(4) show the estimation using subsamples of cities in three waves. All columns include both the cell FE and year FE (Wave×Year FE for column (1)). Standard errors are clustered at city level. Significance: *p<0.1; **p<0.05; ***p<0.01

Table 10: Hereogeneous Analysis: GDP Growth Pressure by Roll-out Waves

	Dependent variable: $\ln(PM2.5_{it})$			
	All	Wave1	Wave2	Wave3
	(1)	(2)	(3)	(4)
(0-3km)*Open	-0.064**	-0.017	-0.045**	-0.139*
	(0.025)	(0.013)	(0.022)	(0.072)
(0-3km)*Open*1(Economic Decline)	0.079***	0.031	0.092***	0.116***
	(0.021)	(0.047)	(0.027)	(0.029)
Observations	41,532,584	7,983,534	10,647,527	22,901,523
R^2	0.945	0.956	0.933	0.940

Note: Column (1)-(4) show DID estimation results with heterogeneity in treatment effect, separated by roll-out waves. Open is the treatment indicator that equals one if a cell is in a city that has joined the new monitoring program. $\mathbf{1}(EconomicDecline)$ is a dummy variable indicating cells inside a city that experienced an economic recession in the previous year (decreased GDP). All columns use cells within 3km to the monitor as the monitored group and cells outside 3km as unmonitored groups. Column (2)-(4) show the estimation using subsamples of cities in three waves. All columns include both the cell FE and Wave×Year FE. Standard errors are clustered at city level. Significance: *p<0.1; **p<0.05; ***p<0.01

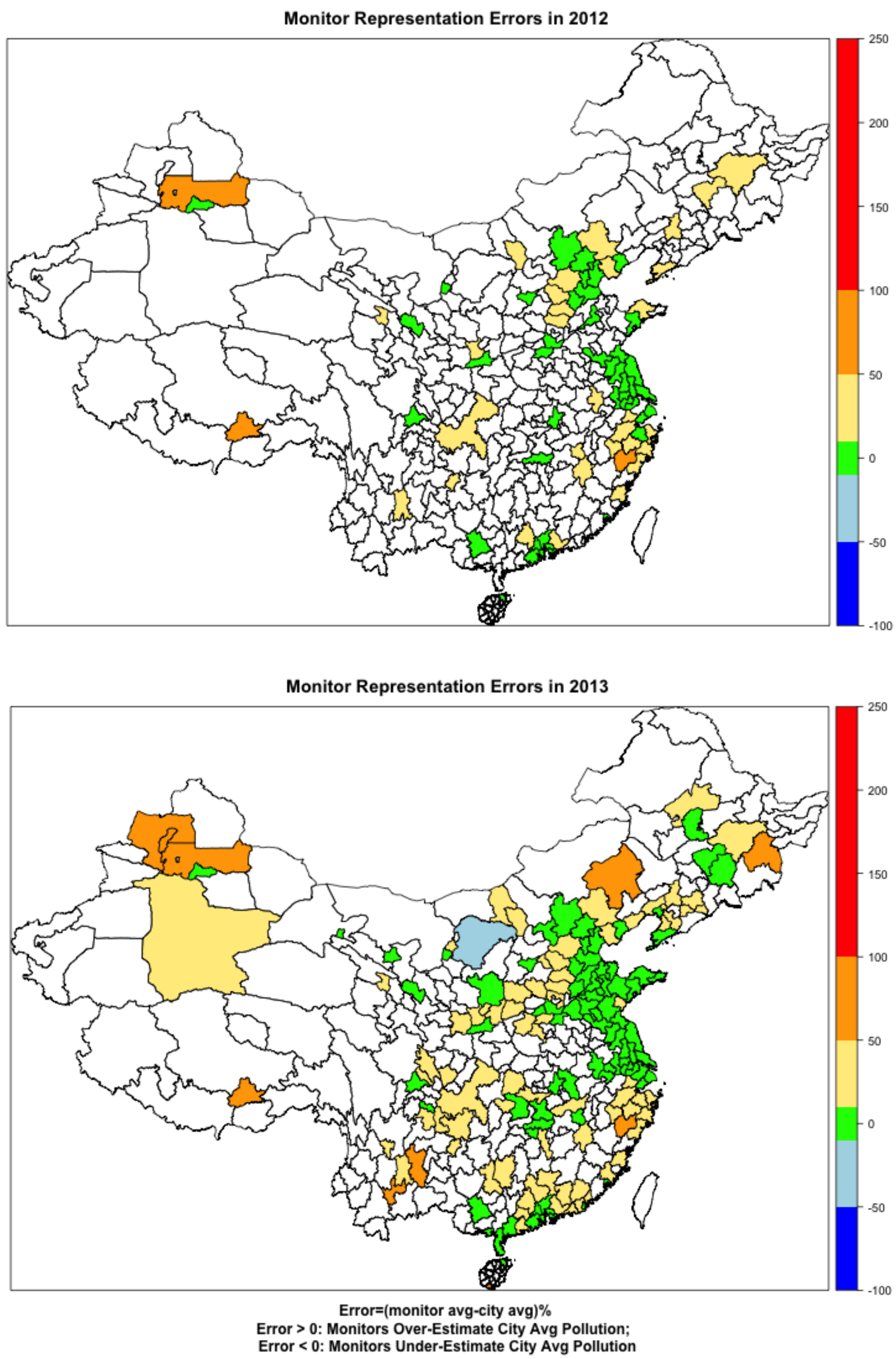
Table 11: Impact of Monitoring on Air Pollution with respect to Distances from Monitors

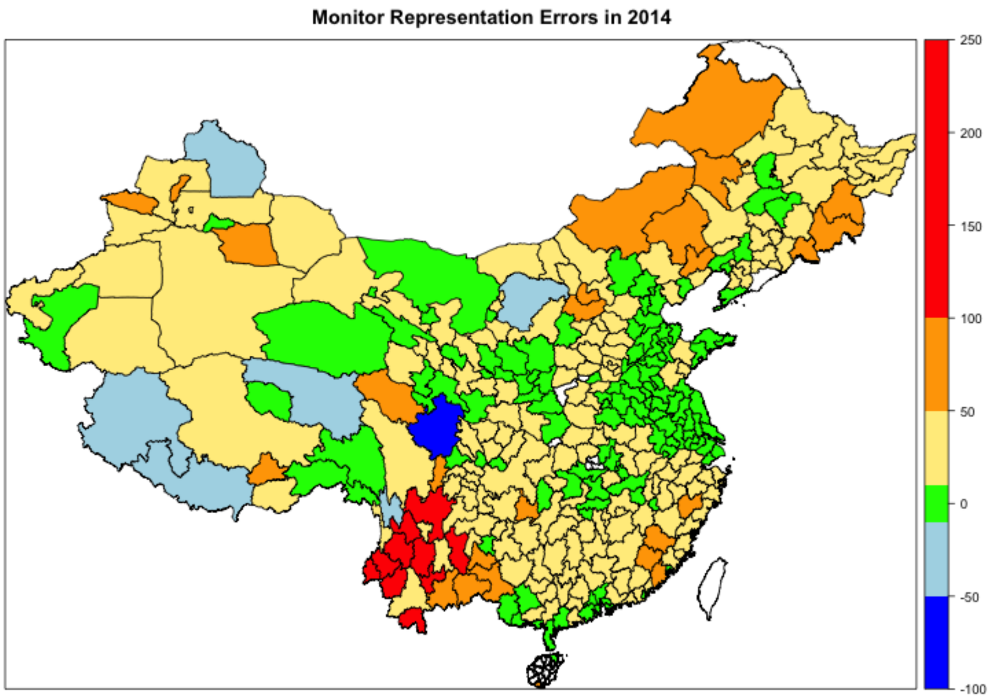
	Dependent variable: ln(PM2.5)		
	(1)	Population (Million)	(2)
Open	0.151***	Outside 300km	1.914
	(0.048)		
(0-10km)*Open	-0.172***	0-10km	261.386
	(0.041)		
(10-20km)*Open	-0.172***	10-20km	108.383
	(0.042)		
(20-30km)*Open	-0.173***	20-30km	97.286
	(0.042)		
(30-40km)*Open	-0.170***	30-40km	88.179
	(0.043)		
(40-50km)*Open	-0.165***	40-50km	80.283
	(0.043)		
(50-60km)*Open	-0.160***	50-60km	70.712
	(0.044)		
(60-70km)*Open	-0.153***	60-70km	55.526
	(0.045)		
(70-80km)*Open	-0.145***	70-80km	43.907
	(0.046)		
(80-90km)*Open	-0.135***	80-90km	32.314
	(0.047)		
(90-100km)*Open	-0.126***	90-100km	24.369
	(0.048)		
(100-150km)*Open	-0.092*	100-150km	51.003
	(0.053)		
(150-200km)*Open	-0.052	150-200km	11.966
	(0.058)		
(200-300km)*Open	-0.051	200-300km	8.113
	(0.050)		
Observations	83,293,774		
R^2	0.967		

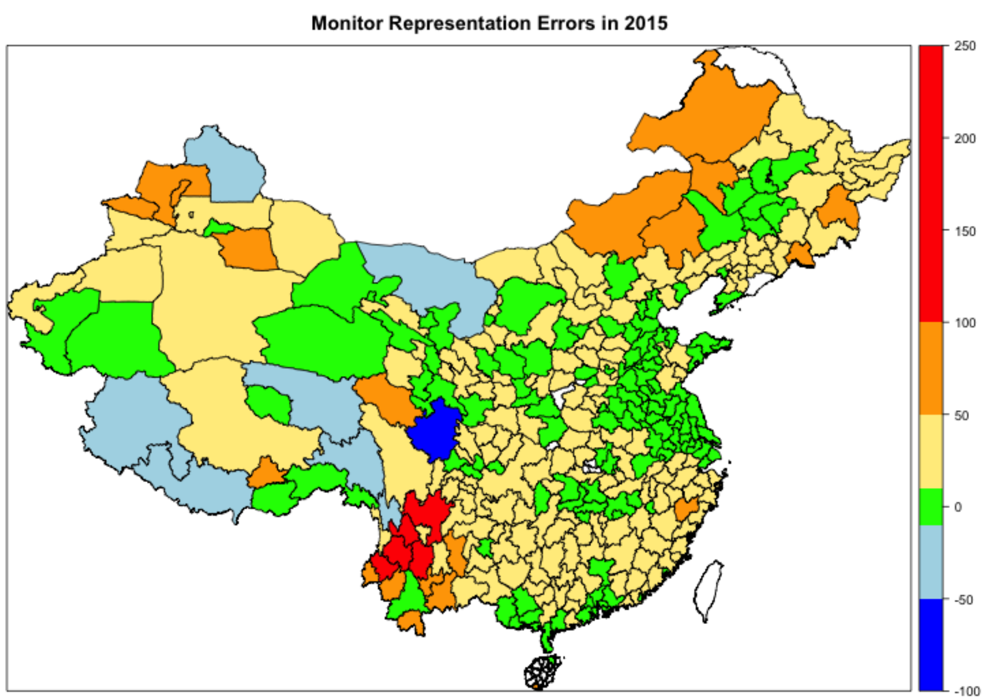
Note: This table shows DID estimation results with treatment intensity bins. Open is the treatment indicator that equals one if a cell is in a city that has joined the new monitoring program. The coefficient estimates of Open represents the impact of monitoring on air pollution in the base group, which includes cells outside of 300km of monitors. The interactions represent the effect in each treatment intensity group. Column (2) shows the total population in each distance bin using 2015 population data. Controls include both the cell FE and Wave×Year FE. Standard errors are clustered at city level. Significance: *p<0.1; **p<0.05; ***p<0.01

A Appendix

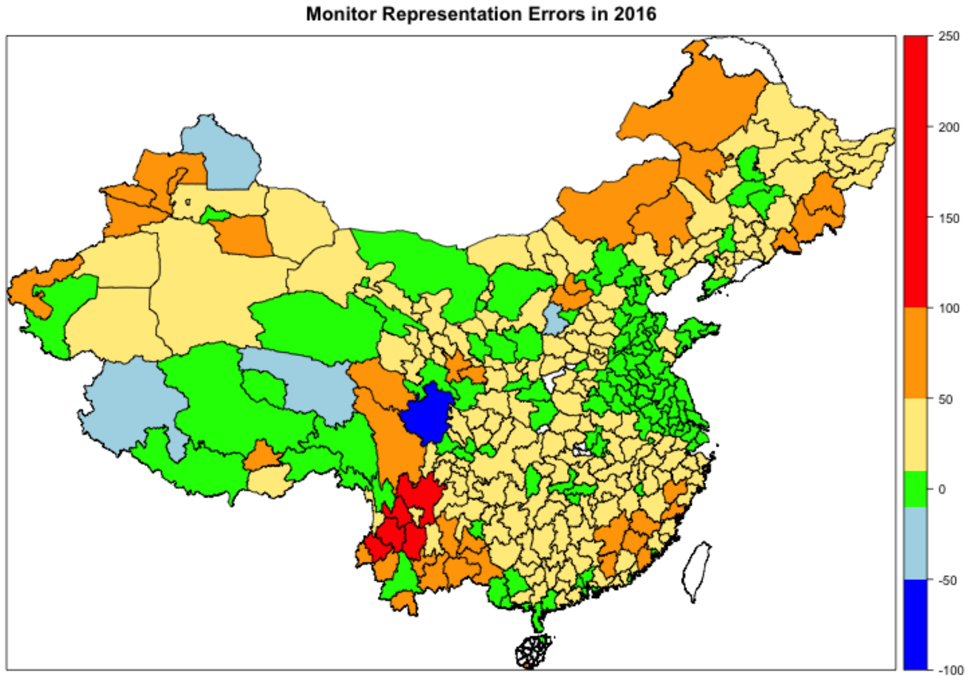
Figure A1: Monitor Representation Errors by Year: all cells vs. monitored cells

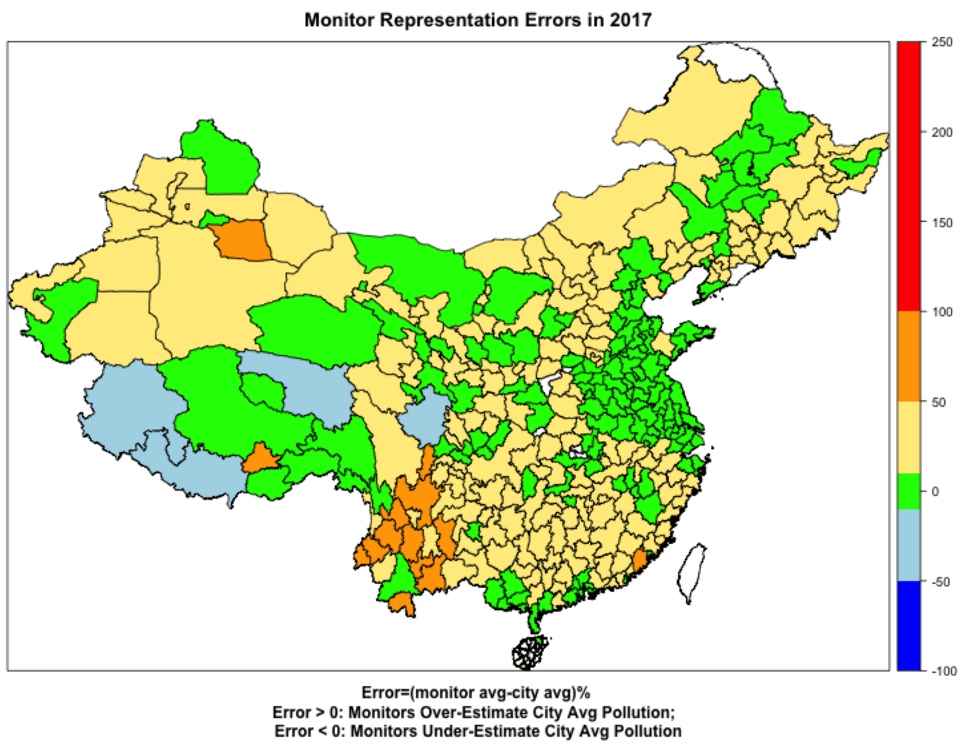






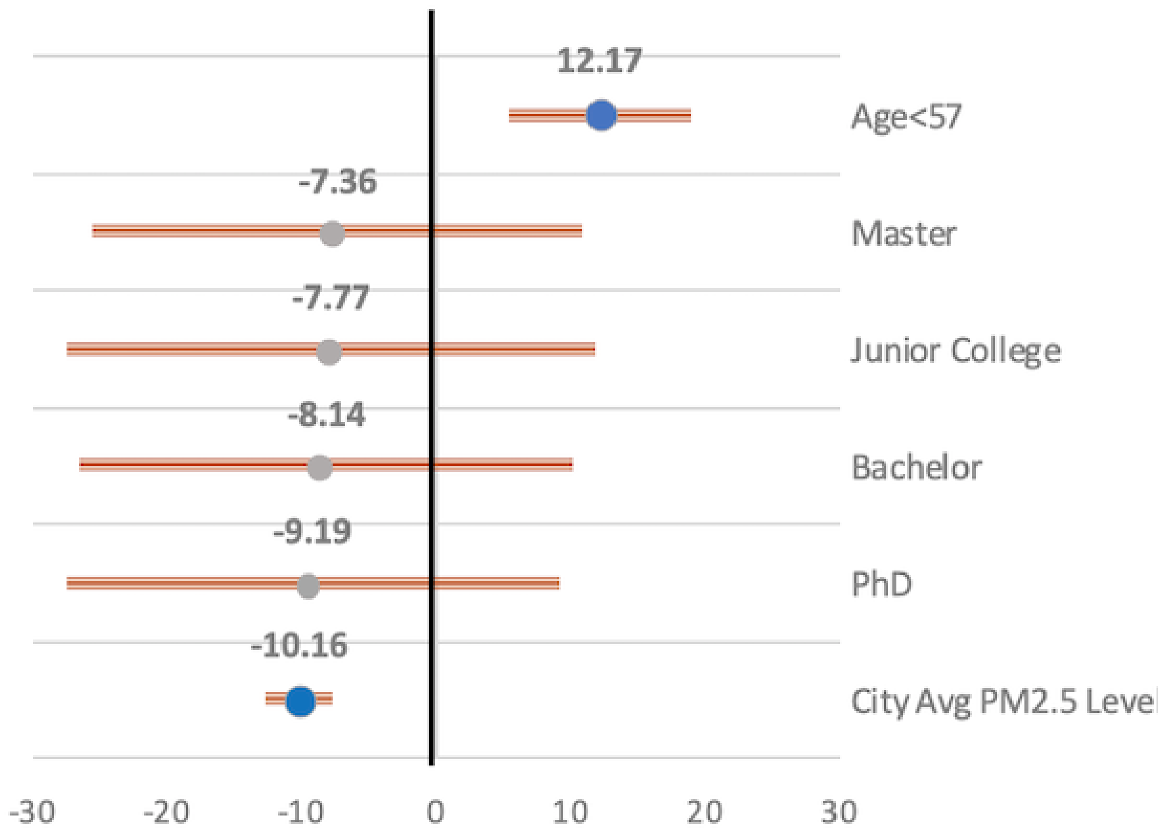
Error=(monitor avg-city avg)%
Error > 0: Monitors Over-Estimate City Avg Pollution;
Error < 0: Monitors Under-Estimate City Avg Pollution





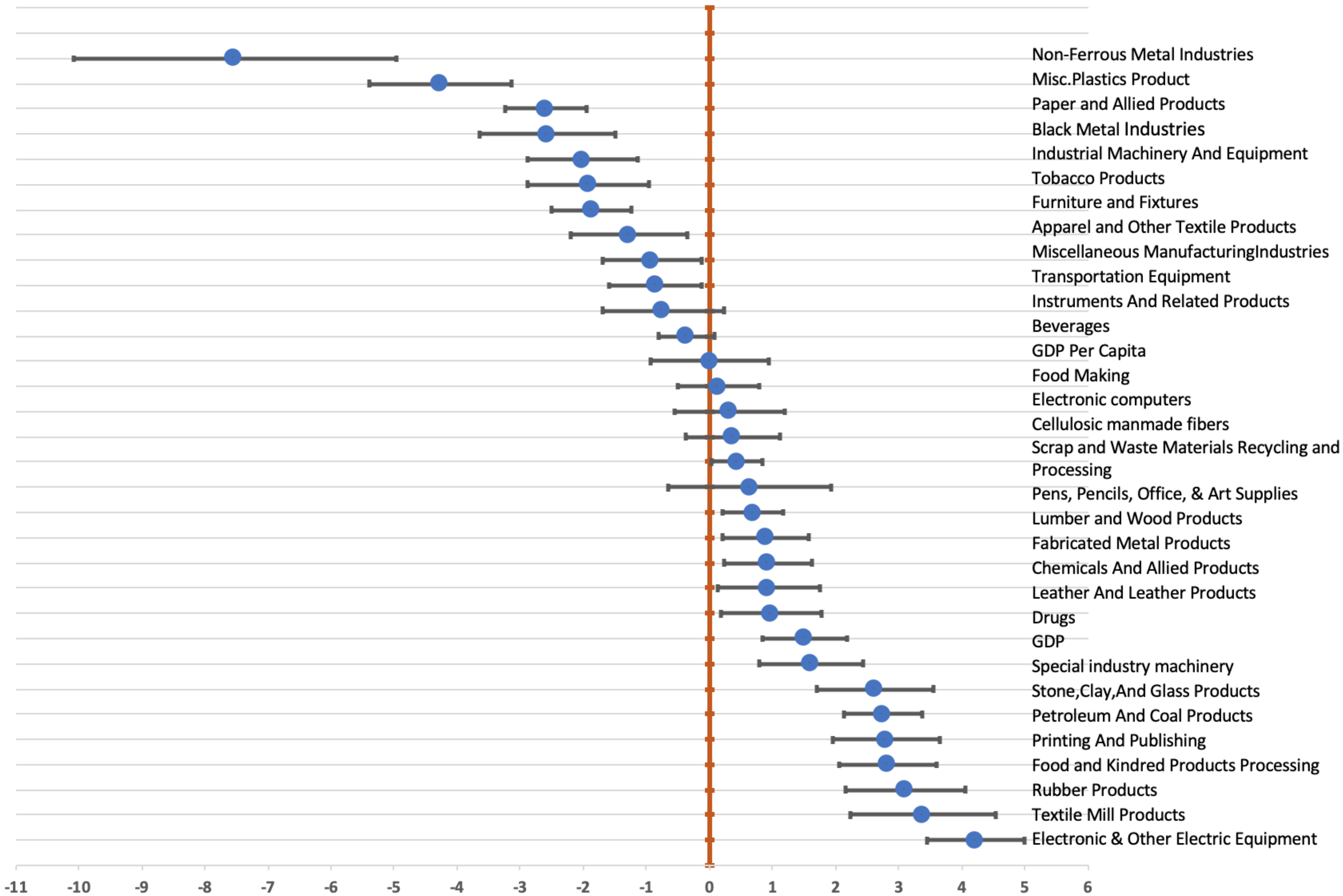
Note: The figures show the monitors representation errors from 2012 to 2017. The representation error is defined as the percentage difference between city average pollution level calculated based on only monitored cells and city average pollution based on all cells. All the pollution levels are weighted by the grid level population in 2015. Cities in green means the monitors well-represent city average PM2.5, with representation errors in [-10% , 10%]. Cities in warm colors (error $>$ 10%) means the monitors over-represent city average pollution, and those in cool colors (error $<$ -10%) means the monitors under-present city average pollution level. The map is based on raw data and presented at city level.

Figure A2: Correlates of Monitor Representation Errors



Notes: This graph reports coefficient estimates with 95% Confidence intervals from a single panel regression of measurement errors on city characteristics. Year and Province Fixed Effects are included.

Figure A3: Correlates of Monitor Representation Errors



Notes: This graph reports coefficient estimates with 95% Confidence intervals from a single panel regression of measurement errors on industrial concentrations (based on 2011 data). Year and Province Fixed Effects are included.