Can Technology Solve the Principal-Agent Problem? Evidence from Air Pollution Monitoring in China

MICHAEL GREENSTONE, GUOJUN HE, RUIXUE JIA, TONG LIU

(October 2019)

This paper examines the consequences of automation of air pollution monitoring as part of the Chinese central government's "war on pollution". Exploiting 654 regression discontinuity designs based on city-level variation in the exact day that monitoring was automated, we find that *reported* airborne particulate matter (i.e., PM_{10}) concentrations increased by around 40% just after the new monitors were operating, which is further corroborated by a difference-in-differences design. The improved quality of data after automation also provides an opportunity to measure data manipulation before automation. We find that the city-level measures of manipulation are positively correlated with corruption prevalence. Further, the accurate information post-automation induces more avoidance behaviors, as measured by online searches for air filters and face masks. Our study illustrates the extent of the principal (i.e., central government) – agent (i.e., local officials) problem in pollution monitoring and shows that technology can mitigate this problem by reducing the scope for hidden actions.

Keywords: Technology, Automation, Monitoring and Surveillance, Moral Hazard, Data Quality

---

Michael Greenstone, University of Chicago and NBER mgreenstone@uchicago.edu

Guojun He, The Hong Kong University of Science and Technology gjhe@me.com

Ruixue Jia, University of California San Diego and CIFAR rxjia@ucsd.edu

Tong Liu, The Hong Kong University of Science and Technology tliuaj@connect.ust.hk

Acknowledgment: We thank Christoper Knittel, Liguo Lin, Alberto Salvo, Bing Zhang, Junjie Zhang, and seminar participants at NBER Summer Institute, Peking University, Nanjing University, MIT, and CIFAR, for constructive comments.

I. Introduction

Social scientists have long recognized the principal-agent problem inherent in the delegation of authority by governments to bureaucratic officials (e.g., Mitnick, 1980; Wilson, 1989; Williamson, 1996; Aghion and Tirole, 1997; and a large literature in public choice). In economics, there exists a rich theoretical literature outlining complicated contracts that align the principal's and agent's incentives (e.g., Laffont and Tirole, 1993; Benabou and Tirole, 2006). An alternative solution and potentially more powerful solution is to find a technology that can greatly reduce the agent's scope for hidden actions.

The execution of environmental policy in China provides an appealing setting to explore these issues. An important underlying feature in China's political system is that local officials are given high-powered incentives to achieve certain economic and social targets, which link their performance in these targets to their promotion as in a career concerns model (Holmström, 1999). While such an incentive system can be effective in achieving targets, it creates incentives to manipulate data.[1] The case of air pollution data is an especially poignant example of this dilemma: in recent years, reducing air pollution became an important target of the central government (i.e. the principal), yet the power of collecting air pollution information is designated to local officials (i.e. the agents). Due to the high volatility in pollution variation and the high cost to verify local information, local officials have strong incentives to manipulate or even fabricate air pollution data before reporting them to the central government. The data quality problem imposes significant costs, because it raises questions about the credibility of linking pollution measures to key outcomes such as life expectancy (Chen et al. 2013; Ebenstein et al. 2017) and it complicates

1 For instance, in an influential study, Li and Zhou (2005) document a positive correlation between economic growth and promotion probability of provincial leaders, which pioneers a growing empirical literature on the consequences of career concerns. Recently, Chen et al. (2019) provide a comprehensive evaluation on the GDP data in China and document misreporting as a result of local officials' career concerns.

government efforts to undertake international agreements to reduce emissions as it raises questions about the trustworthiness of China's reported efforts to reduce pollution.[2]

This paper examines the introduction of automatic pollution monitoring as part of this "war on pollution". The aims of the automation of monitoring were to provide reliable measurements of pollution to identify local officials' success at achieving their targets and where more stringent policy is necessary, as well as to close any gaps between reported concentrations and true concentrations. In this context, automatic monitoring enables real-time sharing of data with the central government and the public and improvements in quality assurance checks through the use of cross-validation statistical checks. This reform was motivated by questions about the validity of the pre-automation data and some high profile incidents of manipulation.[3] This change greatly increases the costs for local governments to influence or manipulate the data and provide a compelling context to investigate the efficacy of technology to limit the hidden actions of local officials.

The analysis exploits several appealing features of the setting. We collect the exact date that automatic monitoring was implemented in 123 different cities with 654 monitoring stations, which provides station-specific regression discontinuity designs to test for manipulation. Moreover, the implementation date varies across cities allowing for difference-in-differences designs that provide a longer-run test for manipulation. Further, satellite data offer a non-manipulated albeit noisy measure of air pollution concentrations that we correlate with the reported measures before and after automation.

There are three key findings. First, there is striking evidence of the underreporting of air pollution concentrations before automation. The station-level regression discontinuity designs

[2] For example, it was reported that China was reluctant to allow other countries to verify its carbon emission data until the Paris Agreement where China signed on to an agreement that outlines a single transparent verification system for all countries. https://www.pri.org/stories/2017-11-08/china-really-stepping-world-s-new-climate-leader

[3] From our disucssion with regulators, we learn that typical ways of mapiluation include selected reporting and intentional misreporting. Public media often covers more colorful ways such as spraying water in front of a monitor, which may be less common in routine reporting.

based on city-level variation on the exact day that monitoring was automated indicate that *reported* PM_{10} concentrations increased by 35 $\mu g/m^3$ or 41% (relative to 2012 mean) on average, just after monitoring was automated. There is evidence of manipulation for other air pollutants but the impacts are smaller. This is not surprising since other pollutants are less important determinants of the overall pollution index that local officials are held accountable for. Notably, there is not any evidence of a discontinuity in satellite-based Aerosol Optical Depth (AOD) readings, which is used by atmospheric scientists to approximate surface particulate matter concentrations, around the automation date, suggesting that actual air quality did not deteriorate right after implementation. Moreover, the increase in *reported* air pollution concentrations post-automation is also evident in difference-in-differences designs that exploit the variation in the timing of automation across cities, indicating that the underreporting was pervasive in the pre-automation period.

Second, several measures of the quality of *reported* air pollution concentrations indicate a substantial improvement post-automation. For example, the monitor reports of PM_{10} concentrations exhibit greater variability and correlation with AOD data (e.g., Yap and Hashim, 2012). The reliable AOD readings and a rich set of weather variables also give us an opportunity to correct PM_{10} concentrations before automation, using a neural network approach recently developed (see Mullainathan and Spiess (2017) for discussion of this approach in economics). Specifically, we train an artificial neural network to create a data set of corrected PM_{10} concentrations before automation. The neural network performs well and is able to explain 82% of PM_{10} variation post automation. We provide the corrected PM_{10} data as supplementary materials for academic research and other purposes in the future.

Third, by estimating city-specific RD, we find a wide variation in the magnitude of discontinuity across cities. An important appeal of this exercise is that it is rare to have empirical estimates of government misconduct, yet the estimation of city-level RD designs for increases in PM_{10} concentrations right after automation provides a compelling quantitative metric. Using this metric, we define manipulation of a city in several different ways by considering the significance and size

of our RD estimates and test whether several political economy factors are correlated with the variation in manipulation across cities. One specific finding is that in cities where corruption is more prevalent, which is measured by the share of indicted officials during China's anti-corruption campaign, the reported air pollution data are less reliable, suggesting that data manipulation is a possible metric for corruption.

This paper contributes to several strands of literature. First, the Chinese government has been adopting new monitoring and surveillance technologies in many sectors, but little is known about the consequences of these technologies. While an extensive literature exists on the impacts of technology adoption on economic development, only a few of them have investigated how information and monitoring technology affects public sector governance and efficiency (e.g., Orphanides 2001; Duflo, Hanna and Ryan, 2012; Muralidharan, Niehaus, and Sukhtankar, 2016).[4] Our study provides one example in environmental regulation and the implications, but are likely to matter in other areas of monitoring and regulation.

Second, our study contributes to a growing literature on environmental monitoring and regulation (e.g., Duflo et al., 2013, 2018; Shimshack, 2014; Greenstone and Hanna, 2014; Browne et al., 2019). We find that technology can play an important role in improving data quality, which matters for researchers, regulators, as well as citizens. In a concurrent study, Barwick et al. (2019) focus on the aspect of information sharing with the public in the new air quality monitoring system in China and documents a relationship between information disclosure and people's avoidance behaviors. Our study complements theirs in that we reveal that the system also significantly improved the air pollution data quality, without which disclosure would be less effective in shaping people's behaviors.

Third, by uncovering the pre-automation manipulation, our study is also related to a nascent line of literature that assesses the reliability of data from China. In particular, Chen et al. (2012) and

[4] Orphanides (2001) investigated monetary policy rules based on real-time data. Duflo, Hannan and Ryan (2012) examined how monitoring (by cameras) and financial incentives affect teachers' attendance in India. Muralidharan, Niehaus, and Sukhtankar (2016) studied the impact of biometric smartcards on state capacity.

Ghanem and Zhang (2014) document the unsmoothness in the distribution of air quality data and reveal that the data are manipulated at a critical threshold (Air Pollution Index=100). One challenge for these studies is that it is unclear whether data manipulation is local and only exists around the threshold. Our findings suggest that this concern is more general and closely related to city characteristics. Researchers should thus be cautious about China's pre-automation air pollution data, particularly in less developed and more polluted cities. As mentioned above, we provide corrected PM_{10} data as a byproduct of this paper.

II. Automating the Air Quality Monitoring System

A. Policy Motivation

As China has experienced rapid economic growth over the last several decades, the demand for better air quality and better data on air pollution has increased. As recently as the early part of this century, China only provides readings of the opaque Air Pollution Index (API), rather than individual readings on PM_{10}, SO_2, and NO_2. A sea change in air pollution reporting was set off when in 2008 the U.S. Embassy in Beijing, and later Consulates in four large Chinese cities (i.e., Shanghai, Guangzhou, Chengdu, and Shenyang) started tweeting hourly fine particulate matter ($PM_{2.5}$) concentrations readings. These readings were more detailed and typically higher than official Chinese statistics, which led to public doubts about the official readings and elevated concerns about air quality. Further, the Beijing Olympic Games in 2008 and celebrity posts of air quality information and measurement on Weibo (Chinese Twitter) raised public concerns over air pollution and information in China (He et al., 2016; Ito and Zhang, 2016).

To meet the public need and ensure public trust, the Chinese government revised the air quality standards in 2012 and later launched the "war on pollution". The Air Quality Index (AQI) was established to replace API with a stricter standard on PM_{10}. Three more pollutants, including $PM_{2.5}$, O_3, and CO, were added in the determination of the AQI and were required to be monitored.

Meanwhile, a new nationwide monitoring network was built to link readings at monitoring sites directly to the central government and the public in real time, making it the largest network of its kind in a developing context.

B. What Does Automation Do?

The automation of the national air quality monitoring network consists of purchasing new monitoring equipment and establishing a new real-time reporting system, which was expected to cost over 2 billion RMB (International Finance News, 2011). Most of the funding was spent on purchasing and installing new equipment to monitor $PM_{2.5}$, CO, and O_3, whose information was not available in the past. Importantly, however, the equipment measuring PM_{10} is unchanged, assuring that differences in PM_{10}, if any, are not due to changes in equipment. Instead, the existing equipment was integrated into the new monitoring system. The primary feature of the new approach to monitoring is real-time reporting, which enables online validition and higher-standard requirements on measurement. We summarize the main changes below.

Before automation, the local environmental bureaus collected data and then submitted them to the central authority without validation. This created possibilities for local governments to manipulate the air quality data by, for example, excluding readings from very polluted hours and days, or simply reporting a lower number than was accurate. In the new monitoring system, opportunities for selective reporting are greatly mitigated as air quality data are integrated in real time and released to the public on an hourly basis by the Ministry of Ecology and Environment, provincial and municipal environmental bureaus simultaneously. The introduction of the Internet of Things and the improvement in surveillance technology allow remote control of data measurement and transmission, as well as quality assurance and quality control.

With the integregated system, the centralized system can conduct online validation of data in real time. For example, inconsistencies across different monitoring sites that are geographically close trigger alerts automatically. In addition, with the availability of real-time data, a higher

standard applies to how to measure pollutants. Specifically, the minimum requirement for calculating daily PM_{10} increased from 12 hours per day to 20 hours per day, from 5 days to 27 days for monthly PM_{10}, and from 60 days to 324 days for annual PM_{10}.

C. Implementation across Cities

The automated monitoring system was introduced into different cities in three waves, as planned by the central government (Appendix Figure S1). In the first wave, 74 key polluting cities (with 496 stations), mostly around Beijing, Yangtze River Delta, and Pearl River Delta, were required to finish the upgrade by January 1st, 2013. In the second wave, another 116 cities (with 449 stations) were ordered to join the central network by January 1st, 2014. Importantly for our analysis, the actual dates that automation was implemented vary greatly even across cities in the same wave. The third wave further required the remaining 177 prefectural cities to build 552 stations by November 2014. Since the stations in the third wave were newly built, no official air quality (daily) data were available before the automation.

This study focuses on 123 cities (with their 654 monitoring stations) – 60 cities in the first wave and 63 cities in the second wave – with pollutant data available both before and after automation. Since the automation policy is implemented at the city level, different stations within a city were automated on the same day, which gives us more power in conducting RD analysis separately for each city.

III. Data and Descriptive Patterns

A. Data

Station-Daily Data on Pollutants

The core of our analysis is based on daily air pollution data from 654 monitoring stations between 2011 and 2016, for all 123 cities involved in Wave 1 and Wave 2 of automated monitoring.

The station-level data are published by the Ministry of Ecology and Environment and local environmental departments and were continuously collected by us. We geocoded the location of each monitoring station using official reports and Google API and matched location-specific information with station-level pollution data.

Station-Daily Data on Weather

As weather conditions are important determinants of local air pollution levels, we collect meteorological data from 403 weather stations. We match each pollution station with the closest weather station available. Daily average temperature, precipitation, humidity, and wind speed are merged with the daily pollution monitor readings.

Station-Monthly Data on AOD

AOD data were obtained from two NASA satellites, TERRA, and AQUA with Moderate Resolution Imaging Spectroradiometer (MODIS). AOD measures the total vertical distribution of particles and gases within a grid according to the light extinction coefficient. It is a measure of 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 $PM_{2.5}$, which is a subset of PM_{10}.[5] Meteorological conditions can alter AOD significantly. The state-of-the-art remote sensing techniques find better correlations between AOD and ground-level PM with coarser spatial and temporal resolutions by month or year (see Hoff and Christopher (2009) for an overview).

Corruption Prevalence and Other City-Leader Characteristics

To investigate which city and political leader characteristics are correlated with data manipulation, we collect a rich set of variables on the cities and their political leaders. We use the information on city-level socioeconomic variables from the official websites of local governments

[5] PM_{10} is particulate matter 10 micrometers or less in diameter, while $PM_{2.5}$ is particulate matter with a diameter of 2.5 micrometers or less. For context, a human hair is about 100-200 micrometers in width.

and the National Bureau of Statistics like GDP per capita. Moreover, we collect detailed biographical data of city leaders (both the Party secretaries and mayors).

We are particularly interested in whether data manipulation is related to corruption. While it is usually difficult to measure corruption, China's largest-scale anti-corruption campaign since 2013 offers an option. As of 2016, more than 100,000 people have been indicted for corruption and over 120 high-ranking officials have been cracked down. Scholars in the China field have used the number of indicted officials (normalized by the number of civil servants) as a measure of corruption prevalence of a city (e.g., Nie et al. 2018). We borrow this measure in 2016 to correlate with our measures on reveal manipualation. Note that our exercises allows for possible selection concerns in the anti-corruption campaign. If those indicted were not more corrupt than the others, we expect no systematic correlation between this corruption prevalence measure and our manipulaton measures.

Policy Implementation Dates across Cities

We collect news reports on the automation date for each city. The deadline for automation dates was assigned by the central government, however, cities were allowed to implement the policy before the deadline. In practice, 50% of the Wave 1 cities implemented it before the deadline and the comparable figure for the Wave 2 cities is 14% (see Appendix Figure S2). One concern is that the local governments might strategically choose the automation dates to hide pre-automation manipulation by for example choosing a time of year when pollution concentrations decline for seasonal reasons. The empirical analysis focuses on a regression discontinuity design that compares the days right before and after automation and this mitigates concerns about the selective choice of the automation date because it is difficult to predict day changes in pollution far in advance. Further, we also use the subgroup of cities that automate on their deadline date as a robustness check under the presumption that this was not a strategic choice to hide manipulation.

B. Descriptive Patterns

Before presenting the empirical results, we describe two sets of empirical patterns as motivational evidence. First, even in the yearly data, we observe discontinuity in reported PM_{10} concentrations pre and post automation. Second, we present data from a few stations to illustrate the variation in reported PM_{10} levels across stations/cities.

Patterns in the Yearly Data

In the yearly data between 2011 and 2016, there is a downward trend in AOD data during the entire sample period, suggesting an overall improvement in air quality in these cities (plotted in Appendix Figure S3). In comparison, the official *reported* PM_{10} concentrations significantly increased in 2013 and 2014, during which the central government automated the air quality monitoring system. For cities in the first wave, for example, *reported* annual PM_{10} levels increased by more than 30 $\mu g/m^3$ from 2012 to 2013, which was about the same magnitude as the total improvement in PM_{10} reduction in the following four years (see Appendix Table S1 for the summary statistics of key variables).

City-Level Cases

This subsection takes an admittedly selective examination of the reported time series from four stations as a means of highlighting the high geographic and temporal variation of the data and qualitatively previewing the finding of extensive manipulation in some locations before automation. For instance, in the monitoring station in the development zone of Shijiazhuang city (the upper left panel of Appendix Figure S4), the reported PM_{10} concentrations jumped from roughly 100 $\mu g/m^3$ to a range of 200 $\mu g/m^3$ to 800 $\mu g/m^3$ immediately after the automation; it seems implausible that changes in weather conditions are so sharp as to cause this increase in concentrations. In the monitoring station installed at Tower II of Tiantai Villa in Zhuzhou city (the upper right panel of Appendix Figure 3), the average PM_{10} concentrations were around 11

$\mu g/m^3$ pre-automation with quite small variations over time. After the automation, in sharp contrast, the PM_{10} levels became several times higher with wider day-to-day and seasonal variation.

These are the time series from just two monitoring sites and indeed not all cities exhibit the same pattern of sharp changes after automation. In the case of Gucheng station of Beijing and Beihai station of Guangxi (the lower panels of Appendix Figure 3), the PM_{10} levels did not change much after the automation and, at least based on visual inspection, seasonal and day-to-day variation seems roughly unchanged. The next section reports on more systematic efforts to assess the degree of manipulation based on the 654 monitors from the 123 cities with PM_{10} data before and after automation.

IV. Empirical Results

A. Assessing Changes in PM_{10} Data

Levels of PM_{10}

We use a regression discontinuity (RD) design based on the dates of air quality monitoring automation to detect air quality data manipulation using the following specification:

$$P_{i,c,t} = \beta_1 I\left(t \geq Auto_{i,c,t}\right) + \beta_2 \mathrm{f}\left(t - Auto_{i,c,t}\right) + \beta_3 \mathrm{I} * \mathrm{f}\left(t - Auto_{i,c,t}\right) + u_{i,c,t} \quad (1)$$

where $P_{i,c,t}$ indicates the pollution levels reported by station *i* of city *c* at time *t* (daily/ monthly). $I(t \geq Auto_{i,c,t})$ is an indicator variable that equals one if station *i* at time *t* is automated. $t - Auto_{i,c,t}$ represents the number of days from the automation and is our running variable. The specification includes a function $\mathrm{f}\left(t - Auto_{i,c,t}\right)$ and allows its effect to differ pre and post automation, which is the basis of the "control function" style approach of the RD design. We assess the sensitivity of the results to several functional forms for f, using both local linear non-parametric and parametric methods. In the parametric case, we report results based on allowing the number of days to interact with a linear term only and a different version with both linear and

square terms. $u_{i,c,t}$ is the error term. Since a city can have multiple stations, we cluster our standard errors at the city level.

The parameter of interest is β_1, which provides an estimate of whether there is a discontinuity in air pollution levels immediately post automation after flexible adjustment for the days before/after automation and the covariates detailed in equation (1).

In the simplest form, we do not include any control variables in the regression, as the dates of automation are arguably exogenous and adding controls should not affect the RD estimate. A natural extension of equation (1) is to include more covariates in the model to check the validity of this assumption. To do so, we can add station fixed effects, month fixed effects, and weather conditions in the equation:

$$P_{i,c,t} = \beta_1 I\left(t \geq Auto_{i,c,t}\right) + \beta_2 \mathrm{f}\left(t - Auto_{i,c,t}\right) + \beta_3 \mathrm{I} * \mathrm{f}\left(t - Auto_{i,c,t}\right) + \beta_4 \mathrm{W}_{\mathrm{i,c,t}} + \alpha_i + month_t + u_{i,c,t} \quad (2)$$

The station-specific effects, α_i, account for time-invariant confounders that are specific to each station. Month fixed effects, $month_t$, can be used to control for seasonality, and weather conditions, $\mathrm{W}_{\mathrm{i,c,t}}$, are often important determinants for pollution concentrations. We include temperature, precipitation, relative humidity, and wind speed in the regression.

To include covariates in the non-parametric RD, we "residualize" the dependent variable — subtract from P_{it} a prediction of P_{it} based on the available covariates —and then conduct and RD analysis on the residuals. This procedure provides a consistent estimate of the same RD parameter of interest (Lee and Lemieux, 2010).[6]

If unobserved determinants of $P_{i,c,t}$ are uncorrelated with the automation indicator, the estimated β_1 is the average causal effect of automation across the regression discontinuity designs

[6] Alternatively, we also include covariates in the non-parametric RD analysis using the methodology developed by Calonico et al. (2018) and obtain similar results.

available from each of the 654 monitors. As we document in Section IV B., the estimates of β_1 vary across location and indeed that section aims to uncover some of the sources of this variability.

Before turning to the estimation of equations (1) and (2), we visualize the patterns in the data in the two years surrounding the automation date. In Figure 1(A), the X-axis indicates the number of days before and after automation. The Y-axis indicates the daily PM_{10} concentrations after adjustment for monitoring station fixed effects, month fixed effects and meteorological conditions (the unadjusted data are plotted in Appendix Figure S5). The residualized concentrations are estimated from running an OLS regression in which the dependent variable is daily *reported* PM_{10} and the explanatory variables are station fixed effects, month fixed effects, and weather controls. It is apparent that there is a discontinuous increase in reported PM_{10} concentrations immediately after automation.

We present the statistical estimates of the discontinuity from equations (1) and (2) in Panel A of Table 1. Columns (1) and (2) report the results from the local linear non-parametric method with and without covariates (i.e. $\mathbf{W}_{i,c,t}, \alpha_i, month_t$). The bandwidths are 109 days and 145 days in the two columns, respectively, which are the optimal bandwidths according to the Calonico et al. (2014) method, and they both use a triangle weighting function. Columns (3)-(4) report the results from the parametric method: column (3) controls for the first-order polynomial and interaction terms and column (4) adds the second-order polynomial and interaction terms.

The entries confirm that there is a striking increase in PM_{10} concentrations immediately after automation. The estimated discontinuity ranges from roughly 31 $\mu g/m^3$ to 35 $\mu g/m^3$, which is a 36%-41% increase, relative to the pre-automation mean (the annual PM_{10} concentration in 2012 is 85.1 $\mu g/m^3$). The choice of the RD method has little effect on the estimated discontinuities and we will emphasize the local linear approach with adjustment for covariates going forward.

The remaining columns analyze subsets of the monitors. In columns (5) and (6) report estimated discontinuous increases in PM_{10} of 28 $\mu g/m^3$ or 33% for Wave-1 cities and 65 $\mu g/m^3$ or 76% for Wave-2 cities, as compared to annual PM_{10} concentration in 2012. In column (6), we restrict the

sample to the 84 cities that implemented the policy upon the deadline (drawn from both waves) so the possibility that the automation date was chosen strategically may be less of a concern. As shown in column (6), we find an increase of 57 $\mu g/m^3$ or 67% for this subgroup, which is higher than the overall estimate and therefore consistent with the possibility that cities chose their automation dates to obscure the extent of pre-automation manipulation.

Overall, the analysis indicates that there was a discrete and substantial increase in *reported* PM_{10} concentrations immediately after automation. The magnitude of our finding is also meaningful economically. Borrowing estimates from Ebenstein et al. (2017), for example, a 35 $\mu g/m^3$ difference in PM_{10} concentration implies a loss in life expectancy by 2.24 years for an average person living in China.

Variability in PM_{10}

As another measure of data quality, we examine the variability of PM_{10} under the presumption that manipulated measures are likely to exhibit less variability than true realizations. We fit equations (1) and (2) by replacing the outcome variable by the monthly standard deviation of the PM_{10} levels. The monthly standard deviation of PM_{10} by is calculated by $SD = \sqrt[2]{\sum_i^n (P_{it} - \overline{P})^2 / (n-1)}$, where P_{it} is daily PM_{10} reading at station *i* on day *t*, $\overline{P}$ is the monthly average, and n is the number of days in a month.

The graphical presentation is illustrated by Panel B of Figure 1. Like the levels of PM_{10}, we find that automation also significantly increased the variability of the reported PM_{10} concentrations. Panel B of Table 1 reports the corresponding estimates. As shown in columns (2) to (4), the magnitude is large: when weather and seasonality are controlled, the standard deviation of PM_{10} increased by around 36% after automation (the mean standard deviation before automation is 39.5). This finding adds more evidence on quality change post automation.

No Similar Changes in AOD Data

As discussed, AOD measures how much direct sunlight is prevented from reaching the ground by aerosol particles and is commonly used as a proxy for particulate air pollution. Our AOD data are at the city-month level, so we fit a version of equations (1) and (2), with the time being measured in months instead of days, to investigate whether there is a discrete change in AOD after automation. For reference, the mean of AOD in this sample before automation is 0.56.

The graphical and statistical analyses fail to find a discontinuity in AOD levels after automation (Panel C of Figure 1 and Panel C of Table 1). In fact, after adjustment for seasonality and weather, the estimated RD coefficient is precisely zero (column (2)). This result confirms that this measure of true air quality did not deteriorate after the automation.

As a basis of comparison, we calculate monthly city-level mean PM_{10} concentrations by averaging across all monitor readings and fit a monthly version of equations (1) and (2). The resulting estimates (Panel C of Table 1) reveal a discrete jump in monthly PM_{10} concentrations that is strikingly similar to the estimated discontinuity from the daily data. The graphical version of the monthly RD estimates for PM_{10} is reported in Panel D of Figure 1.

Alternative Identification Strategy: Difference-in-Differences

A key feature of the RD approach is that it offers a demanding test of the effect of automation immediately after its implementation. We complement the main analysis by estimating a set of fixed effects models that provide less strict tests but offer the potential to estimate the effect of automation on PM_{10} levels in the medium run. Specifically, we conduct "event-study" type analyses and compare reported PM_{10} levels one year before and one year after automation using the following equation:

$$P_{i,c,t} = \gamma_\tau \sum_{\tau=-4}^{\tau=+3} Auto_{i,c,\tau} + \beta W_{i.c.t} + \alpha_i + month_t + \epsilon_{i,c,t} \quad (3)$$

where $P_{i,c,t}$ indicates the pollution levels reported by station *i* of city *c* on day *t*, $W_{i.c.t}$ are weather variables similar to our RD setup above. α_i is the station fixed effect and $month_t$ is the month or year-by-month fixed effects (we will show results for both specifications).

Here, for each station, $Auto_{i,c,\tau}$ indicate different time periods before and after the automation, and we set the pollution readings 1-2 months before the automation date as the reference group ($\tau = -1$). Then, $\tau \in$ {-4, -3, -2, 0, 1, 2, 3} respectively refers to 7-12 months before automation, 5-6 months before automation, 3-4 months before automation, 1-2 months post automation, 3-4 months post automation, 5-6 months post automation, and 7-12 months post automation.

The coefficients of $\gamma_0, \gamma_1, \gamma_2$, and γ_3 allow us to examine whether the automation increases PM_{10} readings in the short and medium runs (relative to the PM_{10} readings 1-2 months just before automation). The coefficients of γ_{-4} , γ_{-3} and γ_{-2} further tell us if the reported PM_{10} readings months ago were comparable to the baseline readings (PM_{10} readings 1-2 months before automation).

Note that the automation dates varied across stations, and more and more stations were automated over time, the estimated coefficients from the above equation using the entire sample will not be exactly the difference-in-differences (DiD) estimates. This is because the control group for each of the "after" coefficients differs over time. Several recent methodical papers emphasize that in such settings, the estimated coefficients can be difficult to interpret and indeed measure some weighted average of the treatment effect (e.g., de Chaisemartin and D'HaultfŒuille, 2017; Goodman-Bacon, 2018; Imai and Kim, 2018).

To deal with this issue, we restrict the sample to cities that automated their monitoring stations only at the deadline of their respective and use data from January 1st 2012 to December 31st 2013. Thus, the "treatment" monitors are from Wave 1 cities where automation occurred on January 1st 2013 and "control" monitors are from Wave 2 cities where automation never occurred during this two-year period. With this set-up, the Wave 2 (deadline) cities were never treated and could serve as the clean "control" group for the Wave 1 (deadline) cities.

The results are reported in Panel A of Table 2. In column (1), we control for station and month fixed effects, and in columns (2), we further control for the station and year-by-month fixed effects. We omit 1–2 months before automation from the regression, so the estimated coefficients are the

difference in PM_{10} readings between the treated and control group at different periods relative to the difference 1–2 months before automation. While the difference becomes significantly positive only after the automation, the pre-treatment estimates (7–12 months, 5-6 months, and 3-4 months before automation) are negative and occasionally statistically significant, indicating that cities in these two waves are different in trends.

To take the different trends into consideration, we match each monitoring station in the Wave 1 (deadline) city to the nearest monitoring station in the Wave 2 (deadline) cities, expecting that geographically-close pairs of monitors should have similar pre-trends in their reported PM_{10} levels. In columns (3) and (4) of Table 2, we report the estimates using this matched sample. Here, we observe that the "3–4 months before", "5–6 months before", and "7–12 months before" coefficients become small in magnitude and statistically insignificant using the matched sample, implying that matching by distance alleviates the pre-trend concern.

It may appear that the estimated increase in the reported PM_{10} concentrations became smaller as time passed. This is in fact driven by the underlying levels of pollution: for instance 1-2 months post automation happen to be winter and are a more polluting season. To take the level effect into consideration, we use the logarithm of PM_{10} as the outcome variable in column (4). Here, we find that the relative PM_{10} difference between the two groups of cities remained largely the same over time (24%, 32%, 29%, and 24% respectively for "1–2 months after" coefficient, "3–4 months after" coefficient, "5–6 months after" coefficient, and "7+ months after" coefficient). We thus conclude that the improvement in air pollution data post automation is not a "local" impact that just applies to several days and that cities did not find a new way to manipulate the data after automation at least during our study period.

Correlation between PM_{10} and AOD

As a further test of whether the PM_{10} data quality improved post-automation, we examine the correlation between PM_{10} and the satellite AOD data, treating the latter as an unbiased measure.

The observation is at the city-month level and we standardized both the PM_{10} and AOD data for this analysis. For standardization, we subtract the mean from an observation and divide it by its standard deviation. The standardization changes the units to standard deviations and makes it easy to compare the size of the coefficients on a comparable scale.

Panel B of Table 2 summarizes the findings. The dependent variable is the standardized AOD levels of city c at month t and the independent variables include standardized PM_{10} levels, the interaction term between the PM_{10} level and a "Post" dummy that equals one if the city automated the monitoring system and zero otherwise.

In column (1), we present the correlations between PM_{10} and AOD conditional only on weather controls. We find that the correlation became stronger after automation, suggesting an improvement in PM_{10} data. In columns (2) and (3), we further include different time fixed effects to control for seasonality. Again, we find that the correlation between PM_{10} and AOD became significantly stronger after automation and the slope increased by over 90% post automation.

Column (4) includes city fixed effects, so this test relies on within city variation in the AOD-PM_{10} relationship over time and is therefore more demanding. The R-Squared statistic increases dramatically, but the AOD-PM_{10} relationship is significantly attenuated both before and after automation. This statistical pattern is consistent with Fowlie et al. (2019), which also found that the high correlations that are typically reported between satellite-derived air pollution data and monitoring station data tend to weaken when moving from cross-sectional to panel variation. So although the results in the other columns reveal a strengthened post automation correlation between AOD and PM_{10}, the limited variation in AOD within location over time provides an important caveat to these conclusions. It is also apparent that future research on the relationship between AOD and PM_{10} would be valuable.

Correcting the Pre-Automation PM_{10}

In light of the results in columns (1) – (3) of Table 2 Panel B, we attempt to correct the pre-automation PM_{10} data by exploiting the relationship between PM_{10}, AOD and weather conditions. To increase our predictive power, we use an artificial neural network (ANN) to train the post-automation data set, assuming that the post-automation data on PM_{10}, AOD, and weather conditions are reliable.

Specifically, we implement a backpropagation algorithm for the multi-layered neural network (Doherr, 2018). The neural networks are capable of performing input-output mapping of data without a priori knowledge of distribution patterns (see Mullainathan and Spiess (2017) for discussion of their applications in economics). Our inputs in the algorithm include AOD, weather conditions, and a rich set of dummies indicating location and month. We use two layers with 20 nodes each, and train the model using post-automation data with 1,000 iterations.

The trained neural network can explain 82% of the variation in PM_{10} in the post-automation sample. As a basis of comparison, this model outperforms polynomial regression models; we obtain an R-squared of roughly 0.7 when we regress PM_{10} on polynomial functions of AOD and weather conditions, conditional on city and month fixed effects. We thus use the trained network to predict PM_{10} concentrations for each month in each "data-manipulating" city (detected by city-specific RD regressions).

The correction shifts the distribution of the pre-automation PM_{10} data to the right (Appendix Figure S7). The mean of PM_{10} in this corrected distribution is 21.8 $\mu g/m^3$ or 25% higher than the mean of the reported pre-automation distribution. These corrected PM_{10} data are provided as an online appendix and can be used for academic or other research.

Results for Other Pollutants

We examine the other two pollutants, SO_2, and NO_2 that were monitored throughout our sample. The manipulation of SO_2 and NO_2 is likely less rewarding because the central government uses the API to measure local environmental performance and PM_{10} determines the level of the

API for more than 90% of the time. We also find a statistically significant post-automation discontinuity of 3.0 or 6.0% in NO_2 concentrations and a positive but statistically insignificant effect on reported SO_2 concentrations (see Appendix Table S2). Together, the PM_{10}, NO_2, and SO_2 results are consistent with the hypothesis that manipulation of pollution readings was governed by local officials' incentives.

Additional Checks

We conduct three additional checks in the Appendix, which together lend additional credibility to our findings. First, short-term variations in air quality are often driven by changes in weather conditions. It is thus instructive to examine whether there exist similar discontinuities in the meteorological measures right before and after the automation. This is not the case in our data. We find that all the weather variables (temperature, precipitation, relative humidity, and wind speed) are continuously distributed across the threshold (Appendix Figure S8 and in Table S3), suggesting that the dramatic changes in the air pollution levels across the switching dates were not driven by weather conditions.

Second, a non-manipulation based explanation for the low pre-automation PM_{10} levels is that local governments might have only sampled during regular work hours in this period. As the new system required local governments to monitor air pollution for more hours each day (from no less than 12 hours before automation to no less than 20 hours after automation), the changes in PM_{10} levels might be mechanically driven by having sampled more polluted hours post-automation.

However, the data do not support this explanation. The percentage of missing values between daytime hours and nighttime hours were similar pre-automation. Additionally, we calculate daily daytime and nighttime PM_{10} levels using hourly data before and after the automation and then fit equations (1) and (2) separately: we find qualitatively similar post-automation discontinuities in the daytime and nighttime (Appendix Table S4).

Third, we check the sensitivity of the RD estimates using alternative kernel weighting and higher-order global polynomial functions (Appendix Table S5). For the local linear RD, using a different kernel function yields similar estimates. For the global polynomial RD, the estimates remain quantitatively similar when we apply a third or fourth polynomial to the daily data.

B. Understanding the Variation in Data Quality

We have so far focused on average changes. Next, we turn to city-specific RD estimates and try to explore the heterogeneity in data quality across different cities. We estimate equation (2) city by city using the local linear approach. Just as above, the unit of observation is a monitor by day. This is a very flexible approach as it allows the effects of the days before and after to vary by city and returns city-specific estimates of whether there is a discontinuous change in reported PM_{10} concentrations post automation.

While our analysis above has enough power for the average estimate across cities, we would like to be cautious about statistical power when examining the estimate for each city. In particular, before upgrading, some cities suspended data reporting when they installed and tested the new equipment for $PM_{2.5}$, CO, and O_3 monitoring, 49 of which did not report PM_{10} readings for more than 2 months. Our earlier results are robust to including or excluding the 49 cities. However, the city-level estimate for this group is affected by the missing values.[7] Thus, we start with the 74 cities without this concern and include the 49 cities in a robustness check.

Figure 2 plots the estimated coefficients and their 95% confidence intervals. We find the estimated RD coefficient is positive for more than 70% of these cities. Among them, 33 cities' RD estimates are statistically significantly positive at the 5% level. The average of all the RD coefficients is 28.9 $\mu g/m^3$. If we weight the RD coefficient by the inverse of the standard error, the weighted average is 24.7 $\mu g/m^3$, as denoted by the red horizontal line. It is apparent that there

[7] Within the 49 cities, we do not have enough data to get an RD estimate for 25 of them. For the other 24 cities, the estimates are noisily estimated.

is substantial variation in these estimates effects of automation. Perhaps, the most noteworthy finding is that there are 12 cities with the estimated discontinuities greater than 75 $\mu g/m^3$ and 11 of them would be judged statistically significant at the 1% level. We have also identified some big Chinese cities (such as Shijiazhuang and Xi'an) as examples of how severe the under-reporting could be before the automation.

Not surprisingly, at the city level, the RD estimate could be affected by other shocks in a specific city at a specific time, which generates noises when using a city's RD estimate to measure its manipulation. Such noises can explain why we observe statistically negative RD estimates for 11 cities. To take such noises into consideration, we employ four definitions of data manipulation among the 74 cities in Figure 2: (i) an indicator variable that equals 1 for the 33 cities where there is a positive and statistically significant RD estimate at the 5% level; (ii) an RD estimated to be statistically significant at 10% level (where 41 cities qualify as data manipulators with this criterion); (iii) an RD estimate larger than 10 $\mu g/m^3$ and is statistically significant at the 10% level; and (iv) the point estimate of the estimated discontinuity.

In addition, we further include the cities that did not report data for more than 2 months before automation in the sample as a robustness check. Although we do not have a clear-cut RD estimate for each city, we can compare the average readings before and after automation and draw some inferences. Here, we define a city as a data-manipulating one if its post-automation 6-month average PM_{10} concentration is 35 $\mu g/m^3$ higher than the pre-automation average PM_{10} concentration for the same months.[8]

We plot the spatial variation in manipulation by city (following definition (iv)) in Appendix Figure S9, which shows a wide variation across regions. A natural question arises: why were some

[8] For example, the data in city A were available from January to October 2013, became missing from November to December 2013, and became available again starting from January 2014 when the system was automated. In such case, the city-level non-parametric RD estimate can become very noisy and sometimes does no converge (as there is no observation close to the threshold for a pro-longed period of time). We then compare the PM_{10} levels between the January–June 2013 and January–June 2014: if the reported average increased by 35 $\mu g/^3$, we treat city A as a data-manipulating city. We choose 35 $\mu g/ m^3$ as the threshold because this reflects the average discontinuity in Table 1. Using 20, 30, or 40 $\mu g/m^3$ as alterative threshold would yield qualitatively similar results.

cities more likely to manipulate the air pollution data than other cities? Our quantitative city-specific estimations of manipulation provide a rare opportunity to investigate the political economy of hidden actions, which we will link to city-level factors.

Is Manipulation Correlated with City-Leader Characteristics?

The variation in manipulation revealed in Figure 2 is striking and here we aim to examine which characteristics of cities and leaders are associated with manipulation. We do not attempt to identify a causal impact but to provide cross-sectional correlations between city-leader characteristics and manipulation. One specific factor we are interested in is the corruption prevalence in a city, because it is useful to know whether data manipulation is a metric for corruption.

We start with GDP per capita as a sanity check. As GDP per capita likely captures both the demand and supply of transparency and hence mitigates the hidden action problem of local officials, we expect to see a negative correlation between GDP per capita and our manipulation measure. This is indeed the case in column (1) of Table 3, where we use definition (i) for manipulation.

We add two city-level variables in column (2). The first is the pollution level in a city, which is a proxy for the necessity of manipulation. Since the pre-automation pollution data are likely to be manipulated, we use the mean PM_{10} levels measured in 2013 for Wave 1 cities and in 2014 for Wave 2 cities (i.e., when the data are more reliable). As shown, necessity matters: pollution level is positively correlated with manipulation. The second is the corruption prevalence measure discussed above. We find that the corruption index is positively associated with manipulation likelihood; so in cities where more officials are indicted during the anti-corruption campaign, the air pollution data are more likely to be manipulated. Compared to column (1), the goodness of fit of the regression in column (2) significantly increased (R-square increases from 0.06 to 0.22), implying that city characteristics can explain a sizable variation in manipulation.

In column (3), we further include a set of personal characteristics of the Party secretary of a city in the regression: whether the Party secretary is local (born in the same province), whether he or she worked for the Youth League (an elite career track) before, his or her age, where he or she has engineering or science background (and hence might know more about the environment). Including the same variables for mayors (Appendix Table S6) does not vary these results. Among these characteristics, the only significant variable is whether the party secretary has an engineering or science degree. The change in R-squared from that in column (2) is not large (from 0.22 to 0.30) either. Thus, in these cross-sectional correlations, city characteristics appear to have more explanatory power than leader characteristics.

These patterns are similar when using the definitions (ii) and (iii) for manipulation. In column (4), if the RD estimate is positive and statistically significant at the 10% level, a city is treated as a data-manipulating one. In column (5), we further require the RD coefficient to be larger than 10 $\mu g/m^3$ to be considered as a data-manipulating city. As shown, the results remain very similar.

In columns (6) and (7), manipulation comes from definition (iv), i.e., the estimated RD coefficient of each city. Column (6) reports the unweighted regression and column (7) reports the weighted regression, where we weight the regression by the inverse of the standard errors of the RD coefficients. In essence, we put a larger weight on cities with more accurate estimate in column (7). As before, we find that GDP per capita is negatively associated with the RD estimate, and PM_{10} level is positively associated with the RD estimate. The coefficient of the corruption measure becomes weaker and that of the engineering background becomes stronger. In either case, the signs remain the same.

Finally, we can conduct a full-sample analysis. As discussed above, for cities where data manipulation cannot be detected by the RD design (due to missing data before automation), we define the manipulation dummy by the difference in the average pre- and post- PM_{10} readings. In column (8), we find the results remain the same.

These findings suggest that data manipulation is a metric for corruption. They also provide some suggestions for where we should be more cautious when using the pre-automation data. First, the data of PM_{10} seems to be less reliable than other pollutants, so one should be particularly careful when using the PM_{10} and API data. Second, perhaps not surprising in hindsight, data quality in less developed and more polluting cities are more of a concern. It would be useful to divide the data based on the development stage as well as post-automation levels to ensure that the analysis is not driven by manipulated data. Finally, as already done in many existing studies, it is helpful to use satellite-derived air pollution data, especially for the pre-automation periods.

C. Any Consequences of Better Information for Citizens?

One may wonder whether there is any real behavioral consequence of providing better air pollution information. We test this possibility by examining whether more people search for "anti-haze face masks" and "air filters" immediately after the automation. We focus on people's online searching behaviors because these measures can be immediately taken when people see high levels of pollution and they often mark the beginning of a series of consumptions of these products in the long run.

We use Baidu's search indices for "anti-haze face masks" and "air filters" in this analysis. Baidu is the biggest search engine in China and provides search indices for specific keywords that are analogous to Google Trends. The index measures how many people searched for a specific keyword in a city during a specific period of time, with a higher value indicating more search activities.

We apply the non-parametric RD estimator to the data and examine whether people immediately searched more for "anti-haze face masks" and "air filters" after automation on their computers.[9] As the actual air pollution levels did not change right before and after the automation, the

[9] Baidu report search indices from PC terminals and from mobile terminals. Here we focus on the indices from PCs as the data from mobile devices were not available before May 2013.

discontinuity in people's avoidance behaviors, as measured by the search indices, could be entirely attributed to the new automatic reporting system. Appendix Table S7 summarizes our findings.[10] We find that people's monthly online search for "face masks" immediately doubled (columns (1) and (2)) after automation. When splitting the sample based on whether on each city's data manipulating status (according to the first definition), we further find that the discontinuity is even greater for cities that under-reported the air pollution levels before automation. For the data-manipulating cities, the Baidu search index for "anti-haze face masks" increased by more than 200% immediately after automation. For "air filters", we find similar pattern: the Baidu search index for "air filters" increased by roughly 25% after automation (column (2)) and this effect is twice larger for the data-manipulating cities (column (8)). This analysis illustrates the critical role of accurate information in shaping people's avoidance behaviors and implies that the under-reporting of the air pollution levels before automation could lead to significant welfare loss (as people cannot protect themselves at the optimal level).

V. Conclusion

In modern societies, governments delegate authority to bureaucratic officials, which makes the principal-agent problem inherent in government organizations. The case of pollution data quality in China shows that high-powered incentives in the public sector can be a double-edged sword: when local officials obtain a strong incentive to perform better, they also have incentives to manipulate data.

The advancement of information technology and the adoption of real-time monitoring offers a possible tool to address this downside. We show that automating the monitoring system significantly improves data quality. The improvement of data quality is an important underlying factor to explain China's success in "war on pollution" in recent years – it is difficult to imagine

[10] Corresponding RD plots are presented in Appendix Figure S10.

an effective policy without reliable information. In addition, we show that the more reliable information post automation induces more people to take avoidance behaviors against pollution, which may imply potentially significant welfare gains. That said, new monitoring and surveillance technologies are likely to have other important implications for governance, about which we do not know much. We consider our study as one example to illustrate the positive side of technology and governance and hope it opens avenues for future research.

REFERENCES

Aghion, P. and Tirole, J., 1997. Formal and real authority in organizations. *Journal of Political Economy*, *105*(1), pp.1-29.

Barwick, Panle Jia, Shanjun Li, Liguo Lin, and Eric Zou, 2019. The value of pollution information: Evidence from China's air quality disclosure. Working paper.

Bénabou, R. and Tirole, J., 2006. Incentives and prosocial behavior. *American Economic Review*, *96*(5), pp.1652-1678.

Browne, O., Gazze, L., Greenstone, M., and Rostapshova O., 2019. Enforcement and Deterrence with Certain Detection: An Experiment in Water Conservation Policy. Working Paper.

Calonico, S., Cattaneo, M.D., Farrell, M.H. and Titiunik, R., 2018. Regression discontinuity designs using covariates. *Review of Economics and Statistics*, pp.1-10.

Calonico, S., Cattaneo, M.D. and Titiunik, R., 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6), pp.2295-2326.Chen, W., Chen, X., Hsieh, C.T. and Song, Z.M., 2019. A forensic examination of China's national accounts. Working paper.

Chen, Y., Jin, G.Z., Kumar, N. and Shi, G., 2012. Gaming in air pollution data? Lessons from China. *The BE Journal of Economic Analysis & Policy, 12*(3).

de Chaisemartin, C., & D'HaultfŒuille, X. 2017. Fuzzy differences-in-differences. *The Review of Economic Studies*, 85(2), 999-1028.

Doherr, T., 2018. BRAIN: Stata module to provide neural network.

Duflo, E., Greenstone, M., Pande, R. and Ryan, N., 2013. Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India. *The Quarterly Journal of Economics, 128*(4), pp.1499-1545.

Duflo, E., Hanna, R. and Ryan, S.P., 2012. Incentives work: Getting teachers to come to school. *American Economic Review*, 102(4), pp.1241-78.

Duflo, E., Greenstone, M., Pande, R. and Ryan, N., 2018. The value of regulatory discretion: Estimates from environmental inspections in India. *Econometrica*, *86*(6), pp.2123-2160.

Fowlie, M., Rubin, E., and Walker, R. 2019. Bringing Satellite-Based Air Quality Estimates Down to Earth. *AEA Papers and Proceedings*, (Vol. 109, pp. 283-88).

Ghanem, D. and Zhang, J., 2014. 'Effortless Perfection:' Do Chinese cities manipulate air pollution data?. *Journal of Environmental Economics and Management, 68*(2), pp.203-225.

Goodman-Bacon, A. 2018. Difference-in-differences with variation in treatment timing. *National Bureau of Economic Research Working Paper*, No. w25018.

Greenstone, M. and Hanna, R., 2014. Environmental regulations, air and water pollution, and

infant mortality in India. *American Economic Review*, 104(10), pp.3038-72.

Hoff, R.M. and Christopher, S.A., 2009. Remote sensing of particulate pollution from space: have we reached the promised land?. *Journal of the Air & Waste Management Association*, *59*(6), pp.645-675.

Holmström, B. 1999. Managerial incentive problems: A dynamic perspective. *Review of Economic Studies*, 66, 169–182. (Originally published in 1982 in Essays in Honor of Professor Lars Wahlbeck.)

He, G., Fan, M. and Zhou, M., 2016. The effect of air pollution on mortality in China: Evidence from the 2008 Beijing Olympic Games. *Journal of Environmental Economics and Management*, *79*, pp.18-39.

Imai, K., and Kim, I. S. 2018. On the use of two-way fixed effects regression models for causal inference with panel data. Working Paper.

International Finance News. 2011. http://paper.people.com.cn/gjjrb/html/2011-12/23/content_ 982293.htm.

Lee, D.S. and Lemieux, T., 2010. Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2), pp.281-355.

Ito K. and Zhang. S., 2018. Willingness to Pay for Clean Air: Evidence from Air Purifier Markets in China. *Woking Paper*.

Laffont, J.J. and Tirole, J., 1993. *A theory of incentives in procurement and regulation*. MIT press.

Li, H. and Zhou, L.A., 2005. Political turnover and economic performance: the incentive role of personnel control in China. *Journal of Public Economics*, *89*(9-10), pp.1743-1762.

Mitnick, B.M., 1980. *The political economy of regulation: Creating, designing, and removing regulatory forms*. New York: Columbia University Press.

Mullainathan, S. and Spiess, J., 2017. Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, *31*(2), pp.87-106.

Muralidharan, K., Niehaus, P. and Sukhtankar, S., 2016. Building state capacity: Evidence from biometric smartcards in India. *American Economic Review*, *106*(10), pp.2895-2929.

Nie, H., Han, D., Ma, L. and Zhang, N., 2018. Government-business relations in Chinese cities (2017). http://www.niehuihua.com/a/chuban/487.html.

Orphanides, A., 2001. Monetary policy rules based on real-time data. *American Economic Review*, *91*(4), pp.964-985.

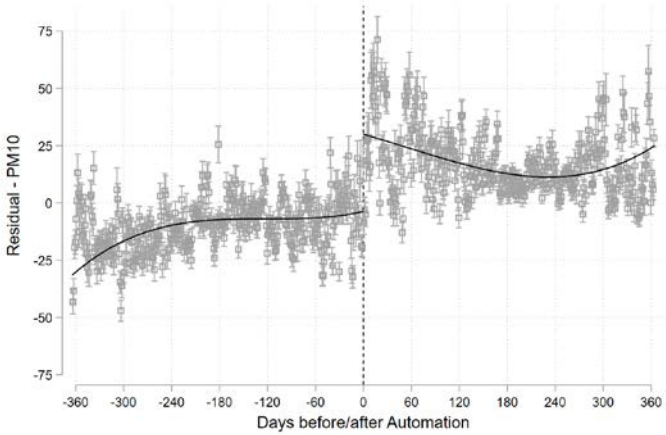
Shimshack, J.P., 2014. The economics of environmental monitoring and enforcement. *Annual Review of Resource Economics*, 6(1), pp.339-360.

Williamson, O.E., 1996. *The mechanisms of governance*. Oxford University Press.

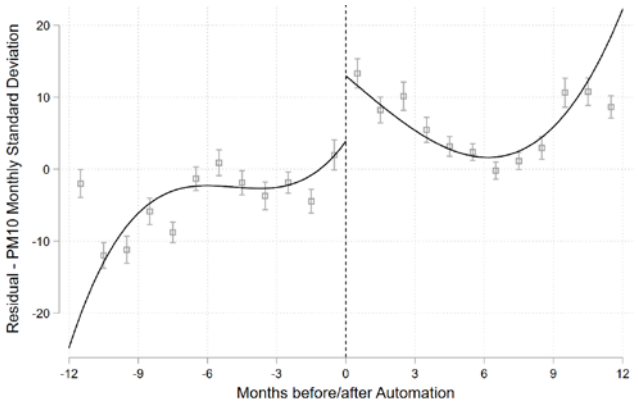
Wilson, J.Q., 1989. *Bureaucracy: What government agencies do and why they do it*. New York: Basic.

Yap, X.Q. and Hashim, M., 2012. A robust calibration approach for PM 10 prediction from MODIS aerosol optical depth. *Atmospheric Chemistry & Physics Discussions*, 12(12).

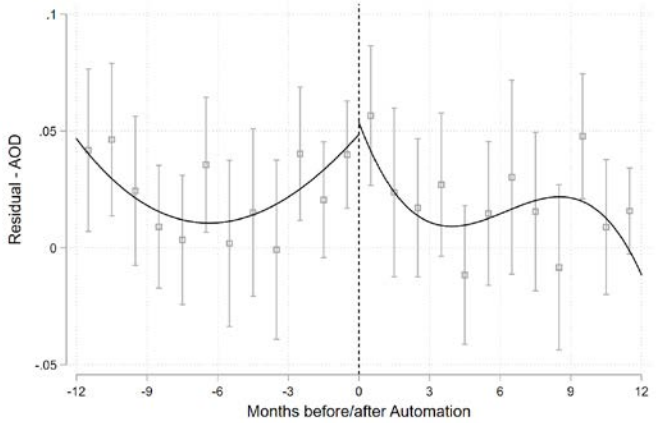
Figure 1. RD Plots for PM_{10} and AOD



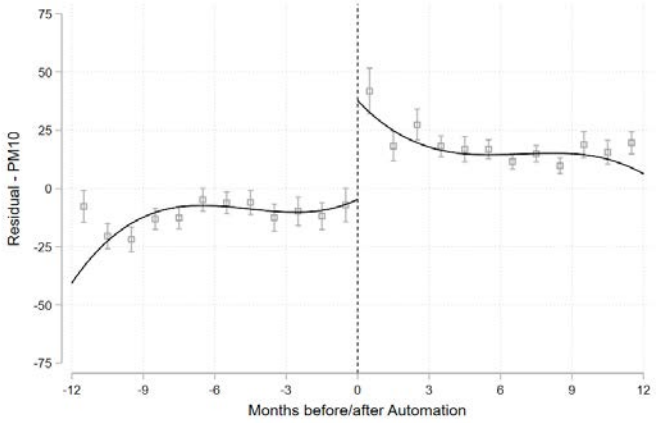
(A). Daily Residual PM_{10}



(B). PM_{10} Variability



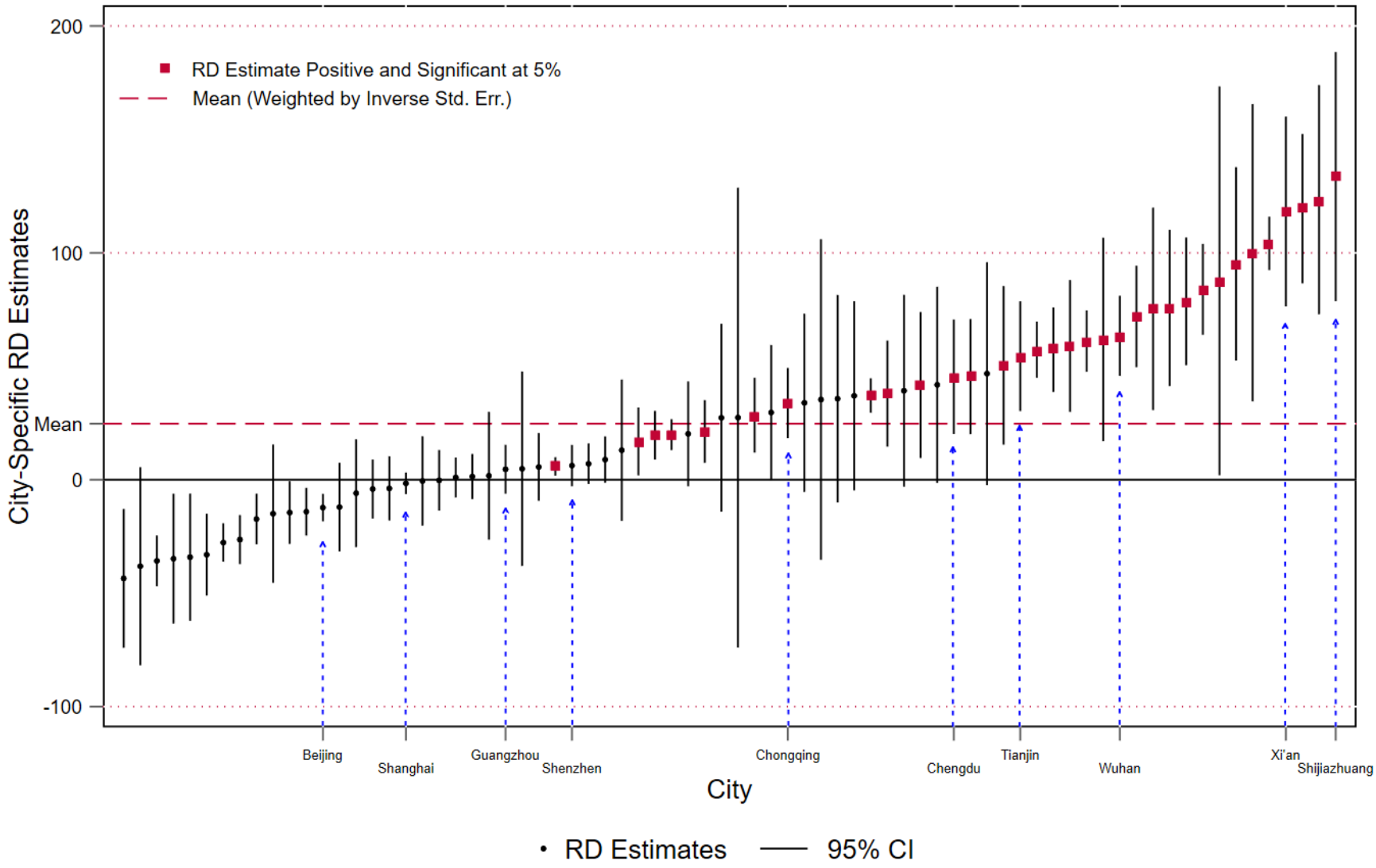
(C). Monthly Residual AOD



(D). Monthly Residual PM_{10}

Notes: Panel (A) shows the increase in PM_{10} immediately after automation using daily data, Panels (B) and (C) show similar changes in PM_{10} standard deviation but no significant change in monthly AOD data. Panel (D) shows the increase in monthly PM_{10} data. Location (station or city) fixed effects, month fixed effects, and weather conditions are absorbed before plotting these discontinuities.

Figure 2. Manipulation Status in Chinese Cities



Notes: The weighted average RD estimate is denoted by the red line, with the weights being the standard errors of the RD coefficients.

Table 1. Automating Air Quality Monitoring System and PM10 Concentrations

	All				Wave 1	Wave 2	Deadline
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
RD Method	LLR		Parametric		LLR		LLR
Panel A. RD Estimates for Daily PM10 Concentrations							
PM10 (Daily)	34.7***	34.9***	32.7***	31.0***	27.6***	64.7***	57.1***
	(10.7)	(5.8)	(4.1)	(4.3)	(9.8)	(9.9)	(8.6)
Panel B. RD Estimates for Variability in Daily PM10 Concentrations							
SD in PM10 (Monthly)	11.6**	14.2***	10.8***	13.3***	14.3***	23.1***	20.2***
	(4.9)	(2.6)	(1.6)	(1.6)	(4.0)	(3.8)	(3.5)
Panel C. RD Estimates for AOD and PM10 (City-Month Level)							
AOD (Monthly)	0.08**	0.00	0.00	0.01	0.04	-0.03	0.01
	(0.04)	(0.02)	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)
PM10 (Monthly)	29.3***	43.9***	30.4***	31.3***	32.8***	70.6***	68.5***
	(10.0)	(5.1)	(2.6)	(3.5)	(7.7)	(8.8)	(9.0)
Location (Station or City) FE		Y	Y	Y	Y	Y	Y
Month FE		Y	Y	Y	Y	Y	Y
Weather Controls		Y	Y	Y	Y	Y	Y
Linear Interaction			Y	Y			
Quadratic Interaction				Y			
Kernel Function	Triangle	Triangle			Triangle	Triangle	Triangle
Daily Data Obs.	1,049,325	1,049,325	1,049,325	1,049,325	628,487	420,838	584,618
Monthly Data Obs.	8,688	8,688	8,688	8,688	4,349	4,339	4,873

Notes: Each cell in the table represents a separate RD estimate. Weather controls include temperature, relative humidity, precipitation and wind speed. Monitoring fixed effects are only applicable to Panels A and B, and city fixed effects are applicable or Panel C. Standard errors clustered at the city level are reported below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

Table 2. Automation and Changes in PM10 Levels

	(1)	(2)	(3)	(4)
Panel A. Changes in PM10 Concentrations After Automation				
	PM_{10}	PM_{10}	PM_{10}	Log (PM_{10})
7-12 Months before	-8.6*	-17.2**	-10.8	-0.13*
	(4.7)	(6.7)	(9.7)	(0.07)
5-6 Months before	6.8	-19.2**	-2.2	0.02
	(6.0)	(9.3)	(12.1)	(0.11)
3-4 Months before	-6.4	-12.0*	-5.2	-0.03
	(5.6)	(6.9)	(9.2)	(0.09)
1-2 Months after	60.3***	31.4***	45.6***	0.24***
	(11.0)	(11.1)	(16.3)	(0.09)
3-4 Months after	45.0***	33.6***	32.5**	0.32**
	(7.8)	(8.8)	(14.2)	(0.12)
5-6 Months after	28.1***	22.2***	29.0**	0.29**
	(6.7)	(8.0)	(13.7)	(0.12)
7-12 Months after	40.0***	9.8	15.8	0.24*
	(6.1)	(8.8)	(14.0)	(0.15)
Sample	DDL	DDL	Matching	Matching
Station FE	Y	Y	Y	Y
Month FE	Y			
Year-Month FE		Y	Y	Y
Weather Controls	Y	Y	Y	Y
R-Squared	0.34	0.35	0.34	0.38
Obs.	176,426	176,426	186,499	186,469

Panel B. Association between AOD (Standardized) and PM10 (Standardized)				
	AOD	AOD	AOD	AOD
PM10	0.202***	0.187***	0.195***	0.034
(per 1 SD)	(0.069)	(0.067)	(0.070)	(0.025)
PM10*Post	0.197***	0.196***	0.185**	0.007
	(0.070)	(0.068)	(0.071)	(0.028)
Weather Controls	Y	Y	Y	Y
Month FE		Y		
Year-Month FE			Y	Y
City FE				Y
R-Squared	0.28	0.35	0.41	0.71
Obs.	3,740	3,740	3,740	3,740

Notes: In columns (1) and (2) of Panel A, there are 242 Wave-1 (deadline) stations and 123 Wave-2 deadline stations. In columns (3) and (4) of Panel A, we match each Wave-1 (deadline) station with its nearest Wave-2 deadline station (with replacement). Standard errors clustered at the city level are reported below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

Table 3. Correlations b/w Data Quality and City/Leader Characteristics

Defining Manipulation	(1)	(2) (i) p<0.05, RD>0	(3)	(4) (ii) p <0.1, RD>0	(5) (iii) p <0.1, RD≥10	(6) (iv) RD Estimate Raw	(7) (iv) RD Estimate Weighted	(8) (ii) +diff. ≥35
ln(GDP per capita)	-0.13**	-0.09	-0.09	-0.10*	-0.12**	-7.30**	-8.13***	-0.02
(per 1 SD)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(3.59)	(2.98)	(0.04)
PM10		0.21***	0.22***	0.21***	0.24***	25.53***	24.00***	0.21***
(per 1 SD)		(0.04)	(0.05)	(0.05)	(0.05)	(3.36)	(3.58)	(0.04)
Corruption Index		0.11**	0.11**	0.13**	0.12**	1.45	4.32	0.09**
(per 1 SD)		(0.05)	(0.05)	(0.06)	(0.06)	(4.73)	(4.18)	(0.04)
Local PS			-0.07	-0.09	-0.16	-9.39	-9.17	-0.04
			(0.11)	(0.11)	(0.10)	(7.81)	(6.94)	(0.09)
Youth League PS			0.11	0.11	0.06	0.14	9.14	0.14
			(0.11)	(0.11)	(0.12)	(9.00)	(7.55)	(0.09)
Age PS			0.01	0.01	0.00	-0.48	-0.09	0.01
			(0.01)	(0.01)	(0.01)	(0.78)	(0.56)	(0.01)
Engineer PS			-0.23**	-0.11	-0.10	-15.40*	-17.95**	-0.04
			(0.10)	(0.12)	(0.11)	(7.98)	(7.24)	(0.09)
Observations	74	74	74	74	74	74	74	123
R-squared	0.06	0.22	0.30	0.27	0.28	0.46	0.45	0.20

Notes: The dependent variables are a dummy variable indicating manipulation in columns (1)-(5) and (8), and are the regression discontinuity (RD) estimates in columns (6)-(7). In correspondence to Figure 2, columns (1)-(3) define manipulation as cities with RD estimates that are significant at 5% or above. Columns (4)-(8) define manipulation as cities with RD estimates that are significant at 10% or above. Column (8) includes the whole sample. Robust standard errors in parentheses. * significant at 10% ** significant at 5% *** significant at 1%.

Table S1. Summary Statistics

	Mean and Std. Dev.					
	2011	2012	2013	2014	2015	2016
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Pollution and AOD						
PM_{10}	87.3	85.1	112.0	106.4	94.0	87.7
($\mu g/m^3$)	(64.0)	(60.7)	(86.4)	(69.7)	(65.0)	(64.2)
AOD	0.600	0.560	0.560	0.546	0.514	0.462
	(0.276)	(0.282)	(0.271)	(0.286)	(0.262)	(0.247)
SO_2	16.0	14.6	15.3	13.3	10.6	8.9
(ppb)	(16.6)	(15.2)	(17.2)	(14.5)	(12.6)	(10.7)
NO_2	19.6	20.0	22.8	21.4	20.0	19.7
(ppb)	(13.8)	(14.4)	(14.6)	(12.2)	(11.7)	(11.4)
Panel B: Weather						
Temperature	14.59	14.68	15.37	15.50	15.60	15.41
(°C)	(11.15)	(11.50)	(11.23)	(10.58)	(10.44)	(10.97)
Precipitation	2.41	3.49	3.42	3.30	3.69	4.05
(mm)	(7.36)	(10.15)	(10.95)	(10.35)	(11.41)	(12.07)
Relative Humidity	63.82	65.47	64.41	64.87	67.20	67.23
(%)	(18.09)	(19.05)	(18.66)	(19.06)	(19.11)	(19.20)
Wind Speed	2.20	2.65	2.69	2.59	2.69	2.77
(m/s)	(1.00)	(1.45)	(1.52)	(1.41)	(1.42)	(1.45)

Notes: Daily air quality data are collected from China's air quality monitoring platform. Weather data are collected from local meteorological stations. AOD data are collected from MODIS.

Table S2. Automation and Reported Concentrations of SO_2 and NO_2

	All			Wave 1	Wave 2	Deadline
	(1)	(2)	(3)	(4)	(5)	(6)
RD Estimates for Pollutant Concentrations						
SO_2	3.28	1.28	1.55	3.25	-0.70	2.40
	(3.29)	(2.30)	(2.08)	(2.97)	(2.30)	(3.04)
NO_2	1.99	3.38***	2.98***	3.48***	2.99**	4.68***
	(1.34)	(0.89)	(0.87)	(1.11)	(1.37)	(1.28)
Station FE	Y	Y	Y	Y	Y	Y
Month FE		Y	Y	Y	Y	Y
Weather Controls			Y	Y	Y	Y
Kernel Function	Triangle	Triangle	Triangle	Triangle	Triangle	Triangle
Obs.	1,106,783	1,106,783	1,106,783	670,092	436,691	616,845

Notes: Each cell in the table represents a separate RD estimate. Weather controls include temperature, relative humidity, precipitation and wind speed. Standard errors clustered at the city level are reported below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

Table S3. Changes in Weather Conditions after Automation

	(1)	(2)	(3)
Temperature	0.92	0.95	0.88
	(0.60)	(0.60)	(0.60)
Relative Humidity	1.27	1.50	1.00
	(1.21)	(1.27)	(1.29)
Precipitation	-0.02	-0.07	-0.06
	(0.22)	(0.22)	(0.24)
Wind Speed	0.00	0.03	-0.04
	(0.06)	(0.06)	(0.06)
Obs.	225,667	225,667	225,667
Sample (Months)	All	All	All
Kernel Function	Triangle	Epanech.	Uniform

Notes: Each cell in the table represents a separate regression. City fixed effects and day fixed effects are always controlled. Standard errors clustered at the city level are reported below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

Table S4. Automation and Reported PM_{10} at Day and Night

	All			Wave 1	Wave 2	Deadline
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. RD Estimates for Daytime and Nighttime						
Daytime (6am-6pm)	39.0***	34.7***	33.5***	29.2***	55.0***	56.4***
	(8.9)	(6.5)	(6.9)	(10.4)	(8.6)	(9.2)
Nighttime	42.3***	39.6***	37.6***	33.2***	64.2***	60.2***
	(8.9)	(6.6)	(5.9)	(10.5)	(9.3)	(9.0)
Panel B. RD Estimates for On-Duty and Off-Duty Time						
On Duty (9am-6pm)	38.8***	35.0***	33.5***	29.1***	57.1***	55.4***
	(9.0)	(6.6)	(6.9)	(10.4)	(8.5)	(9.1)
Off Duty	42.2***	38.7***	32.6***	33.1***	61.5***	60.7***
	(8.9)	(6.3)	(5.1)	(10.6)	(9.2)	(9.2)
Station FE	Y	Y	Y	Y	Y	Y
Month FE		Y	Y	Y	Y	Y
Weather Controls			Y	Y	Y	Y
Kernel Function	Triangle	Triangle	Triangle	Triangle	Triangle	Triangle
Obs.	1,159,032	1,159,032	1,159,032	699,938	459,094	643,329

Notes: Each cell in the table represents a separate RD estimate. Weather controls include temperature, relative humidity, precipitation and wind speed. Standard errors clustered at the city level are reported below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

Table S5. RD Estimates Using Alternative Kernel Weightings and Higher Order Polynomials

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	LLR			Global Polynomial			
Panel A. Station-Day RD							
PM10 (Daily)	34.9***	36.0***	35.7***	32.7***	31.0***	26.3***	31.3***
	(5.8)	(6.4)	(6.6)	(4.1)	(4.3)	(4.6)	(5.3)
SD in PM10 (Monthly)	14.2***	14.5***	14.8***	10.8***	13.3***	10.0***	11.1***
	(2.6)	(2.7)	(2.5)	(1.6)	(1.6)	(1.8)	(1.9)
Panel B. City-Month RD							
PM10 (Monthly)	43.9***	43.5***	43.5***	30.4***	31.3***	3.9	-2.5
	(5.1)	(5.2)	(5.2)	(2.6)	(3.5)	(2.8)	(3.1)
AOD (Monthly)	0.00	0.00	-0.01	0.00	0.01	-0.03	-0.04*
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Location (Station or City) FE	Y	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y	Y
Weather Controls	Y	Y	Y	Y	Y	Y	Y
Kernel/Polynomial Function	Tri.	Epa.	Uni.	Linear	Quadratic	Cubic	Quartic
Daily Data Obs.	1,049,325	1,049,325	1,049,325	1,049,325	1,049,325	1,049,325	1,049,325
Monthly Data Obs.	8,688	8,688	8,688	8,688	8,688	8,688	8,688

Notes: Each cell in the table represents a separate RD estimate. Weather controls include temperature, relative humidity, precipitation and wind speed. For Panel A, monitoring station fixed effects are included in the regressions. For Panel B, city fixed effects are included in the regressions. Standard errors clustered at the city level are reported below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

Table S6. Correlations b/w Data Quality and City/Leader Characteristics with Mayor Controls

	Dummy: p<0.05			Dummy: p<0.10		RD Estimate		Dummy: All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	RD>0	RD>0	RD>0	RD>0	RD≥10	Raw	Weighted	RD>0
ln(GDP_per_capita)	-0.09	-0.09	-0.14**	-0.14**	-0.14**	-9.87*	-9.22**	-0.02
(per 1 SD)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(5.11)	(4.53)	(0.04)
PM10	0.21***	0.22***	0.21***	0.19***	0.23***	24.23***	23.53***	0.21***
(per 1 SD)	(0.04)	(0.05)	(0.06)	(0.06)	(0.05)	(3.67)	(3.70)	(0.04)
Corruption Index	0.11**	0.11**	0.12**	0.10	0.12*	0.94	4.79	0.09**
(per 1 SD)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(5.38)	(4.56)	(0.04)
Local PS		-0.07	-0.06	-0.05	-0.18	-9.06	-5.91	-0.04
		(0.11)	(0.12)	(0.12)	(0.12)	(8.89)	(8.10)	(0.09)
Youth League PS		0.11	0.14	0.15	0.06	1.14	9.24	0.14
		(0.11)	(0.12)	(0.12)	(0.12)	(8.82)	(7.45)	(0.09)
Age_PS		0.01	0.01	0.01	0.00	-0.38	-0.36	0.02
		(0.01)	(0.01)	(0.01)	(0.01)	(0.82)	(0.63)	(0.01)
Engineer_PS		-0.23**	-0.25**	-0.16	-0.11	-17.04**	-20.34***	-0.06
		(0.10)	(0.11)	(0.13)	(0.12)	(8.30)	(7.30)	(0.09)
Local Mayor			0.14	0.12	0.13	8.74	-6.41	0.04
			(0.12)	(0.13)	(0.13)	(10.05)	(9.88)	(0.09)
Youth League Mayor			0.06	-0.01	-0.03	-1.49	1.01	-0.09
			(0.13)	(0.13)	(0.12)	(8.97)	(6.66)	(0.10)
Age_Mayor			0.02	0.01	0.01	1.19	0.68	0.00
			(0.02)	(0.02)	(0.02)	(1.11)	(1.05)	(0.01)
Engineer_Mayor			-0.01	-0.02	0.03	-1.92	-4.62	-0.03
			(0.13)	(0.15)	(0.15)	(10.76)	(9.14)	(0.09)
Observations	74	74	74	74	74	74	74	123
R-squared	0.22	0.30	0.34	0.29	0.30	0.48	0.46	0.21

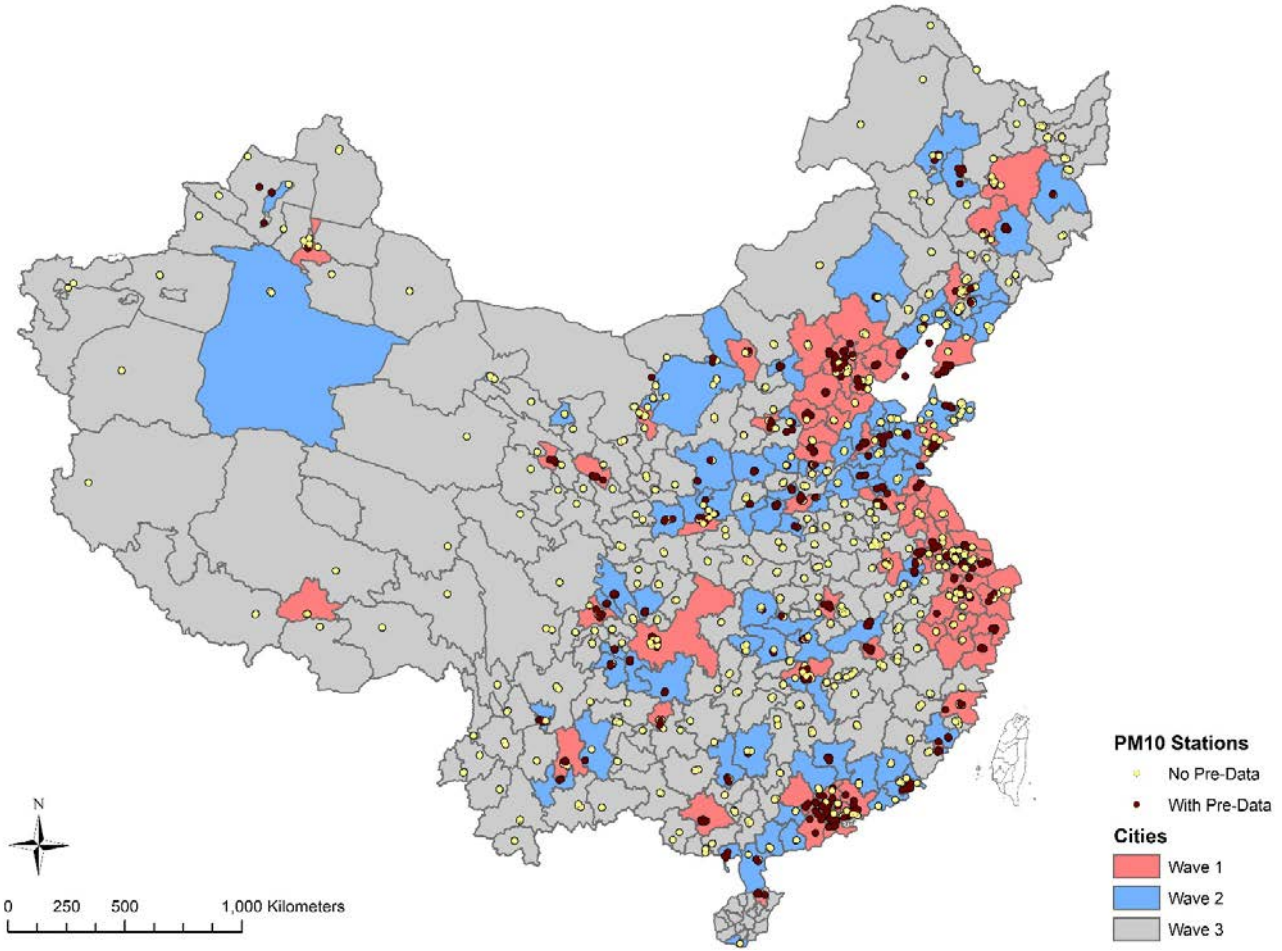
Notes: The dependent variables are a dummy variable indicating manipulation in columns (1)-(5) and (8), and are the regression discontinuity (RD) estimates in columns (6)-(7). In correspondence to Figure 2, columns (1)-(3) define manipulation as cities with RD estimates that are significant at 5% or above. Columns (4)-(8) define manipulation as cities with RD estimates that are significant at 10% or above. Column (8) includes the whole sample. Robust standard errors in parentheses. * significant at 10% ** significant at 5% *** significant at 1%.

Table S7. Regression Discontinuity Estimates for Behavioral Measures

	Log (1+ Search Index for "Anti-Haze Face Masks")				Log (1+ Search Index For "Air Filters")			
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RD Estimates	1.164***	1.545***	0.916***	2.478***	0.179*	0.241***	0.111**	0.471***
	(0.164)	(0.158)	(0.209)	(0.209)	(0.098)	(0.041)	(0.045)	(0.060)
Sample	All Cities	All Cities	Normal	Manipulated	All Cities	All Cities	Normal	Manipulated
Observations	8,791	8,791	5,433	3,358	8,719	8,719	5,361	3,358
Kernel	Tri.	Tri.	Tri.	Tri.	Tri.	Tri.	Tri.	Tri.
City FE	N	Y	Y	Y	N	Y	Y	Y
Month FE	N	Y	Y	Y	N	Y	Y	Y
Weather	N	Y	Y	Y	N	Y	Y	Y

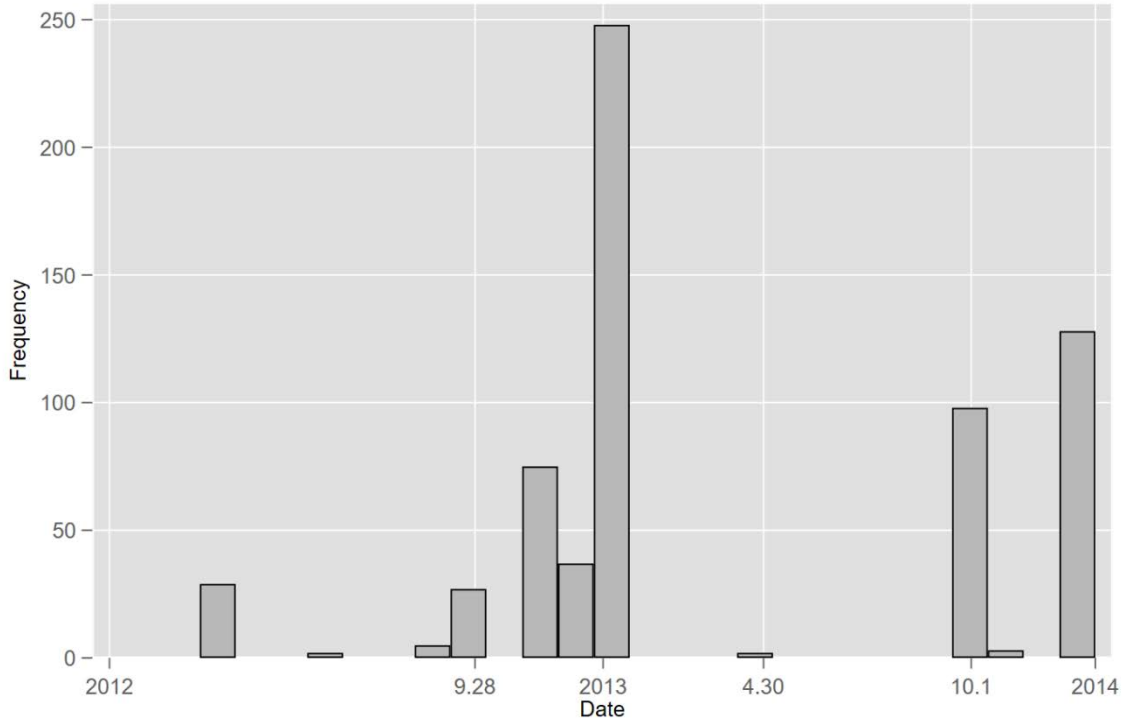
Note: Data are at the city-month level. Each cell in the table represents a separate RD estimate. Weather controls include temperature, relative humidity, precipitation and wind speed. Standard errors clustered at the city level are reported below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

Figure S1. Waves in Automation



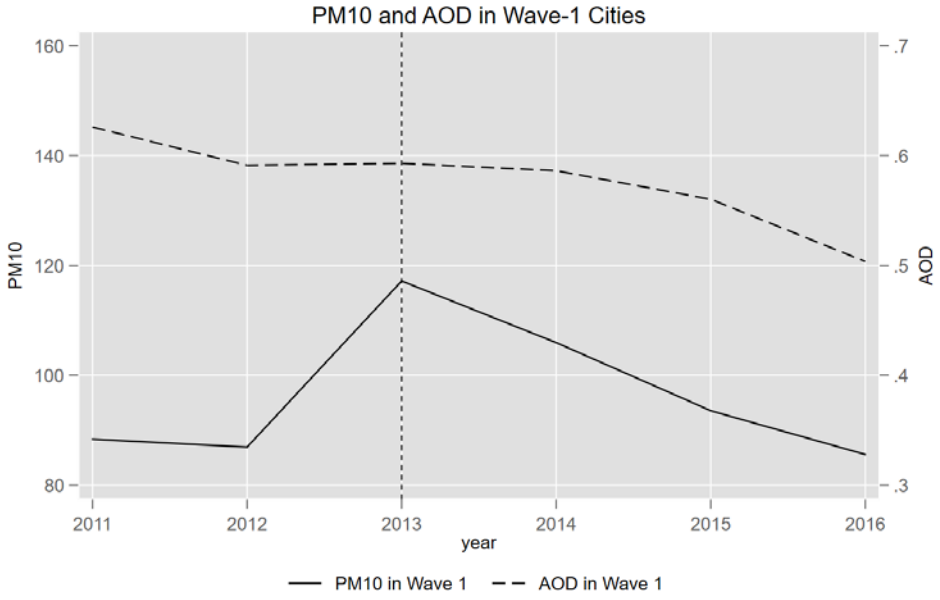
Notes: Wave 1 (red), Wave 2 (blue) and Wave 3 (gray) cities are plotted. Dark dots represent PM_{10} monitoring stations where pre-automation data are available.

Figure S2. Distribution of Automating Dates

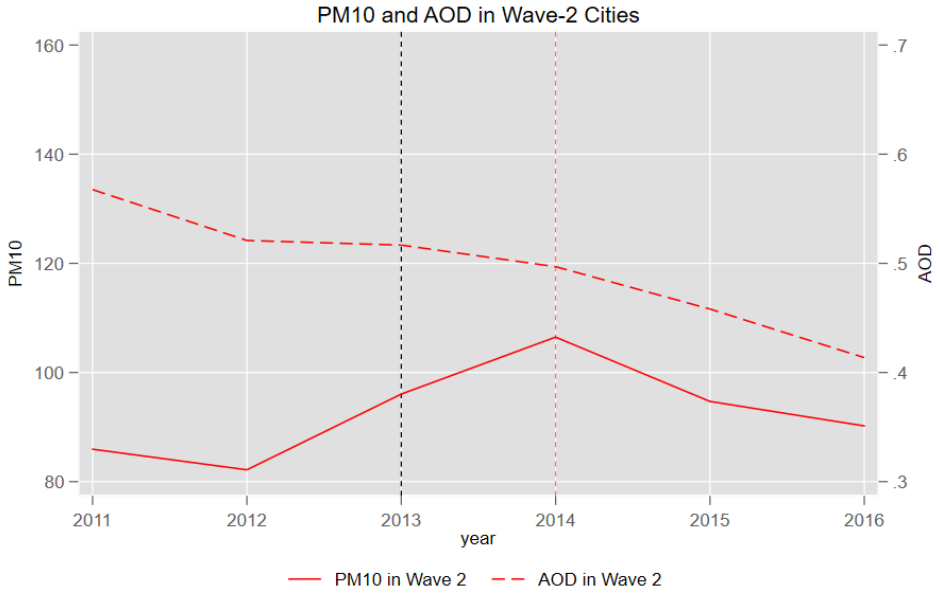


Notes: This figure summarizes the distribution of the automation dates across different cities. The majority of them automated the air quality monitoring stations on January 1st, 2013 and January 1st, 2014.

Figure S3. Annual PM10 and AOD from 2011 to 2016



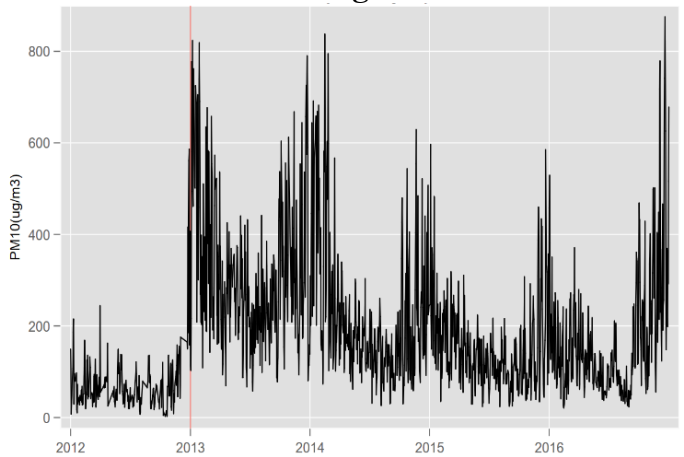
(A). Wave 1 Cities and AOD



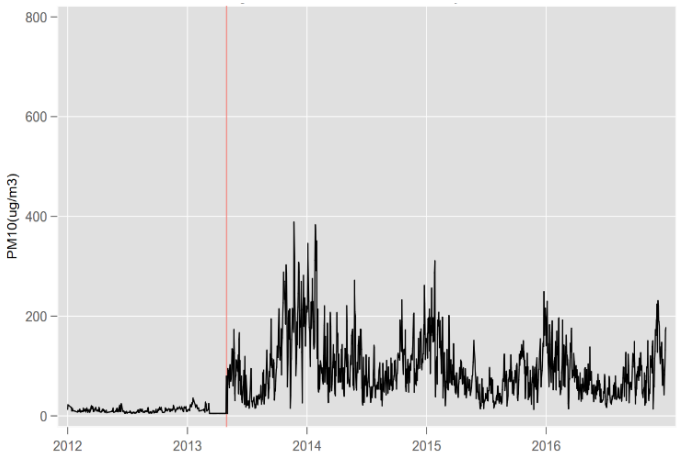
(B) Wave 2 Cities and AOD

Notes: Annual average PM_{10} concentrations ($\mu g/m^3$) in Wave 1 and Wave 2 are plotted in black and red, respectively. Corresponding AOD levels are shown in dashed lines.

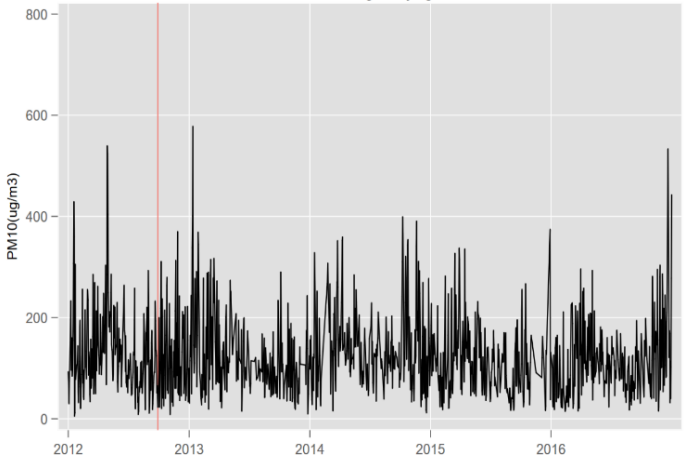
Figure S4. Times Series of PM_{10} Concentrations at Four Stations



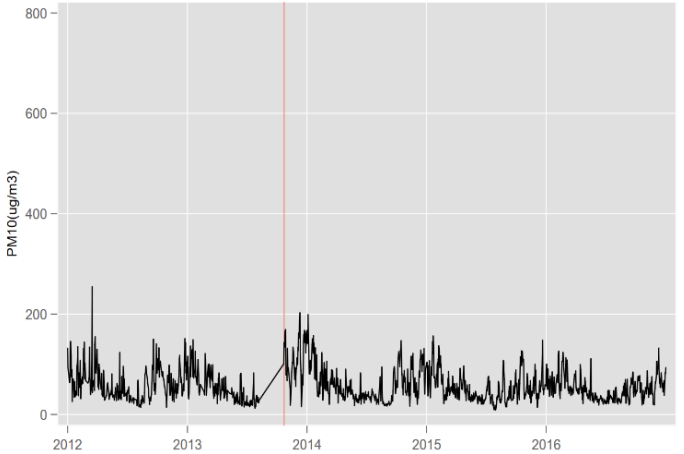
(A). Gaoxin District, Shijiazhuang City, Hebei



(B). Tower II of Tiantai Villa, Zhuzhou City, Hunan



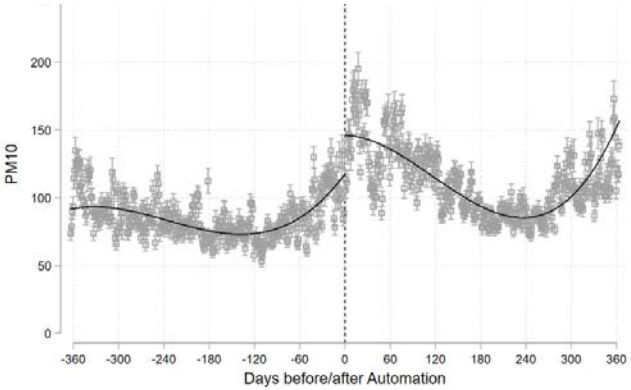
(C). Gucheng, Beijing



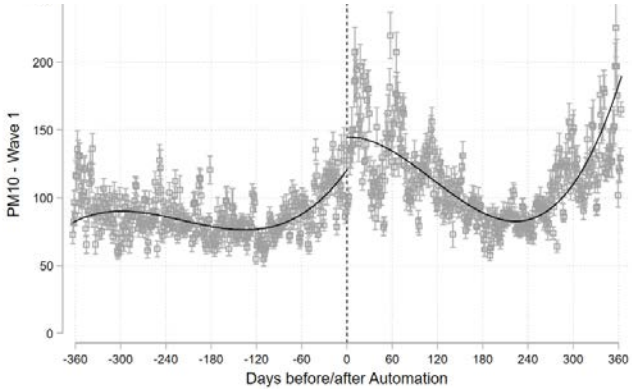
(D). Industrial Park, Beihai City, Guangxi

Notes: The time series of PM_{10} during 2012-2016 at four representative stations in the city of Shijiazhuang, Zhuzhou, Beijing, and Beihai are plotted. Automation dates are denoted in red lines.

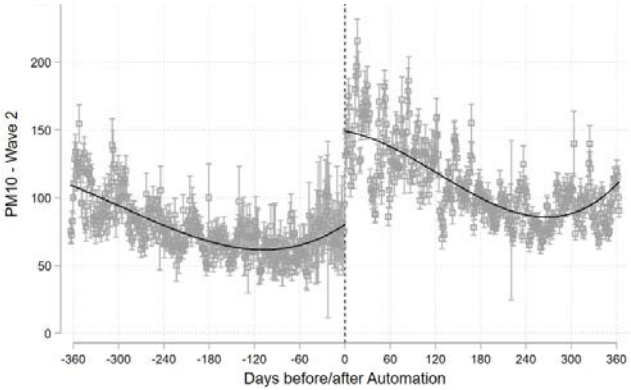
Figure S5. RD Plots Using Raw PM_{10} Data



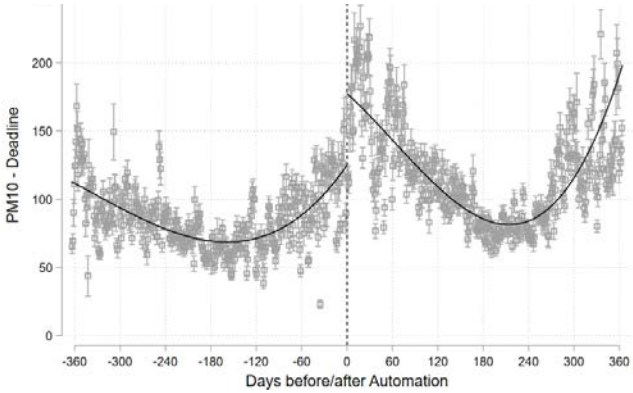
(A). Daily PM_{10}



(B). Daily PM_{10} in Wave 1



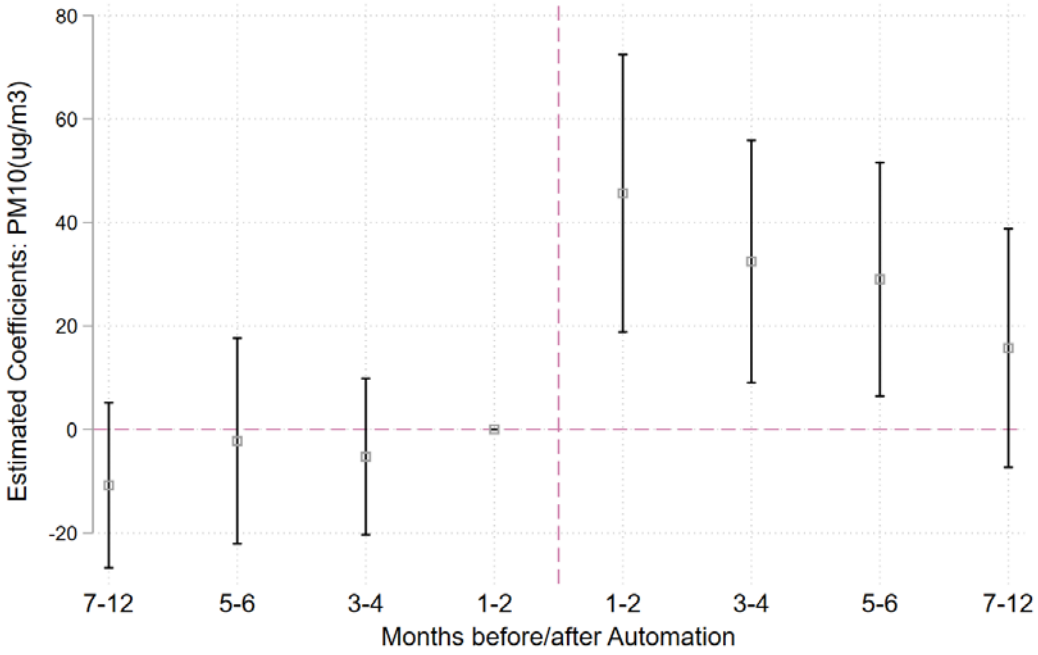
(C). Daily PM_{10} in Wave 2



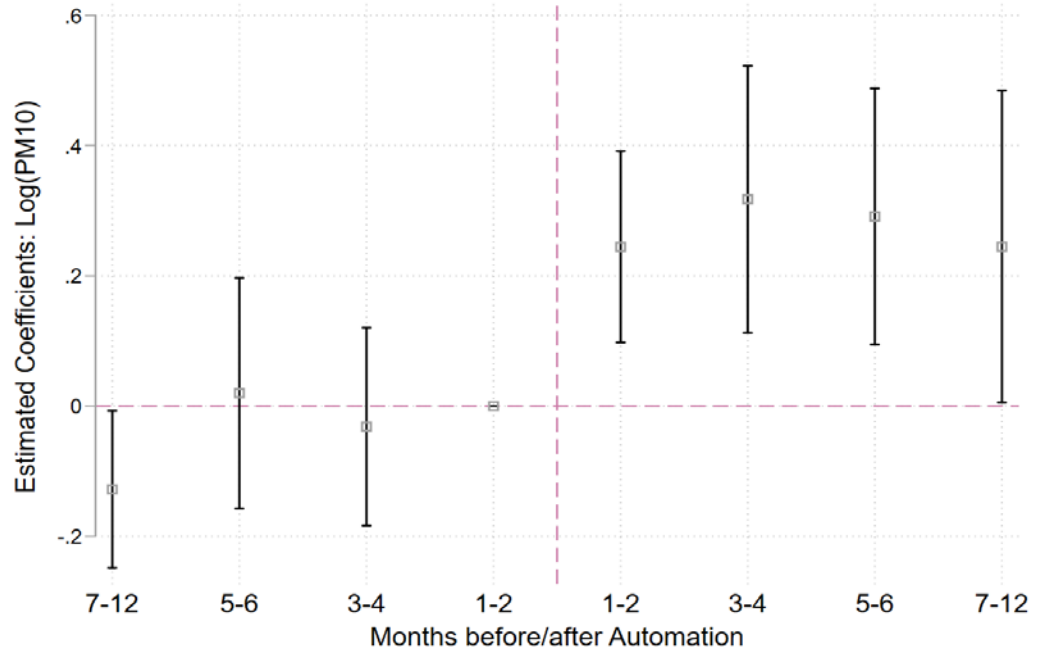
(D). Daily PM_{10} in Deadline

Notes: The discontinuities are plotted using raw PM_{10} concentrations (no controls are included).

Figure S6. Event-Study Estimates of Differences in PM_{10} before and after Automation



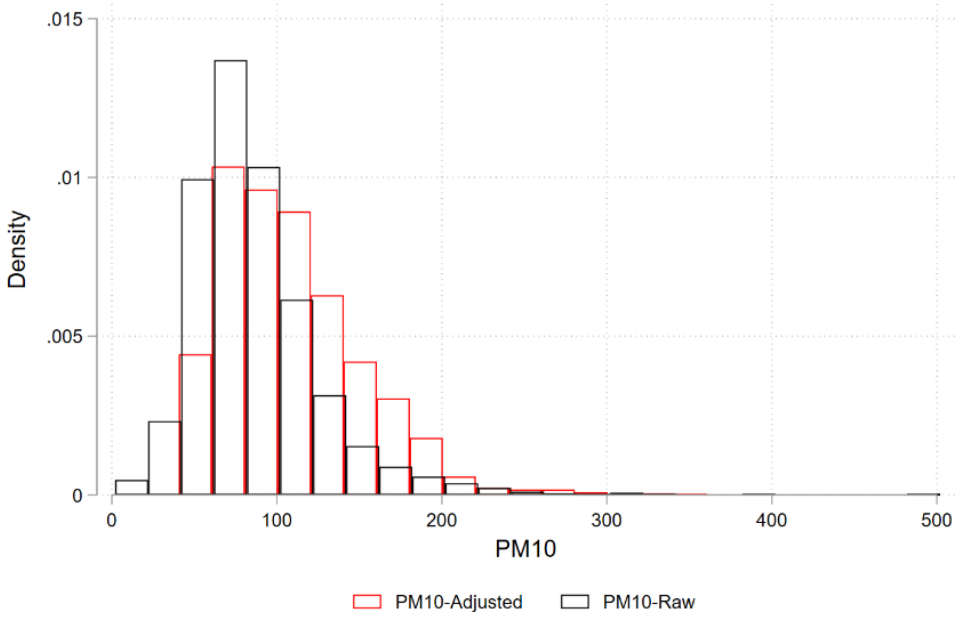
(A). PM_{10}



(B). Log (PM_{10})

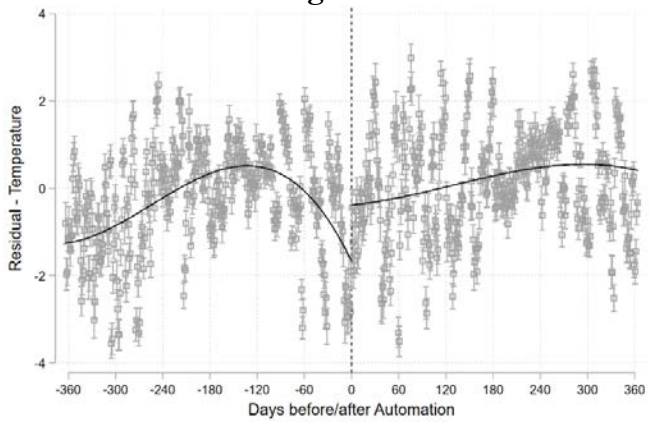
Notes: The treatment group consists of cities that automated their monitoring stations on January 1st, 2013. The control group consists of cities that automated their monitoring stations on January 1st, 2014. We keep data from January 1st 2012 to December 31st, 2013 for this estimation and uses the PM_{10} levels 1–2 months before automation as the reference group. Panel A and B show the estimated coefficients for PM10 and Log(PM10) with 90% CI, respectively.

Figure S7. Correction of Pre-Automation PM_{10} Data

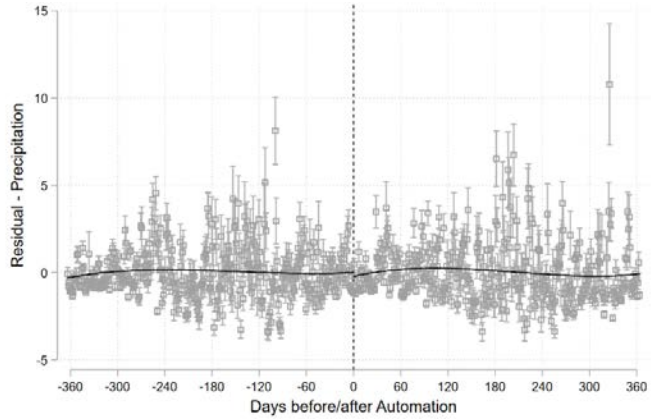


Notes: The distribution of raw PM_{10} data before automation is plotted in black, and the adjusted PM_{10} data using artificial neural networks are plotted in red.

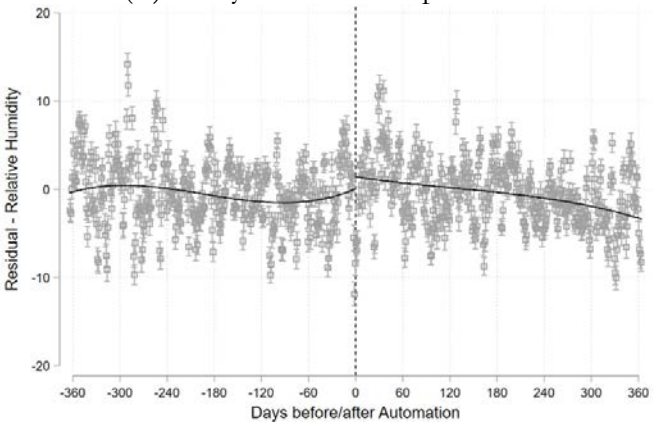
Figure S8. Weather Conditions Before and After the Automation



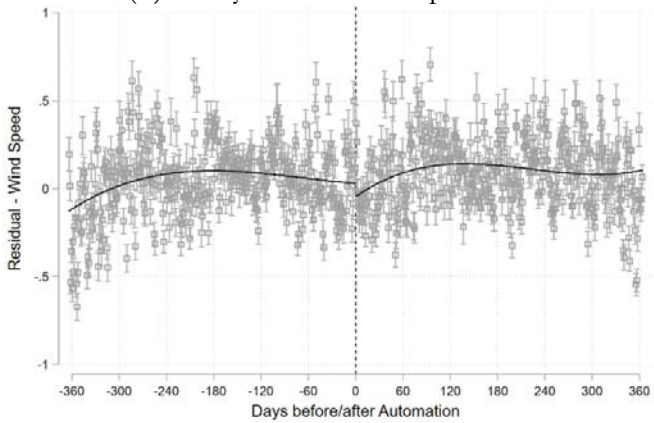
(A). Daily Residual Temperature



(B). Daily Residual Precipitation



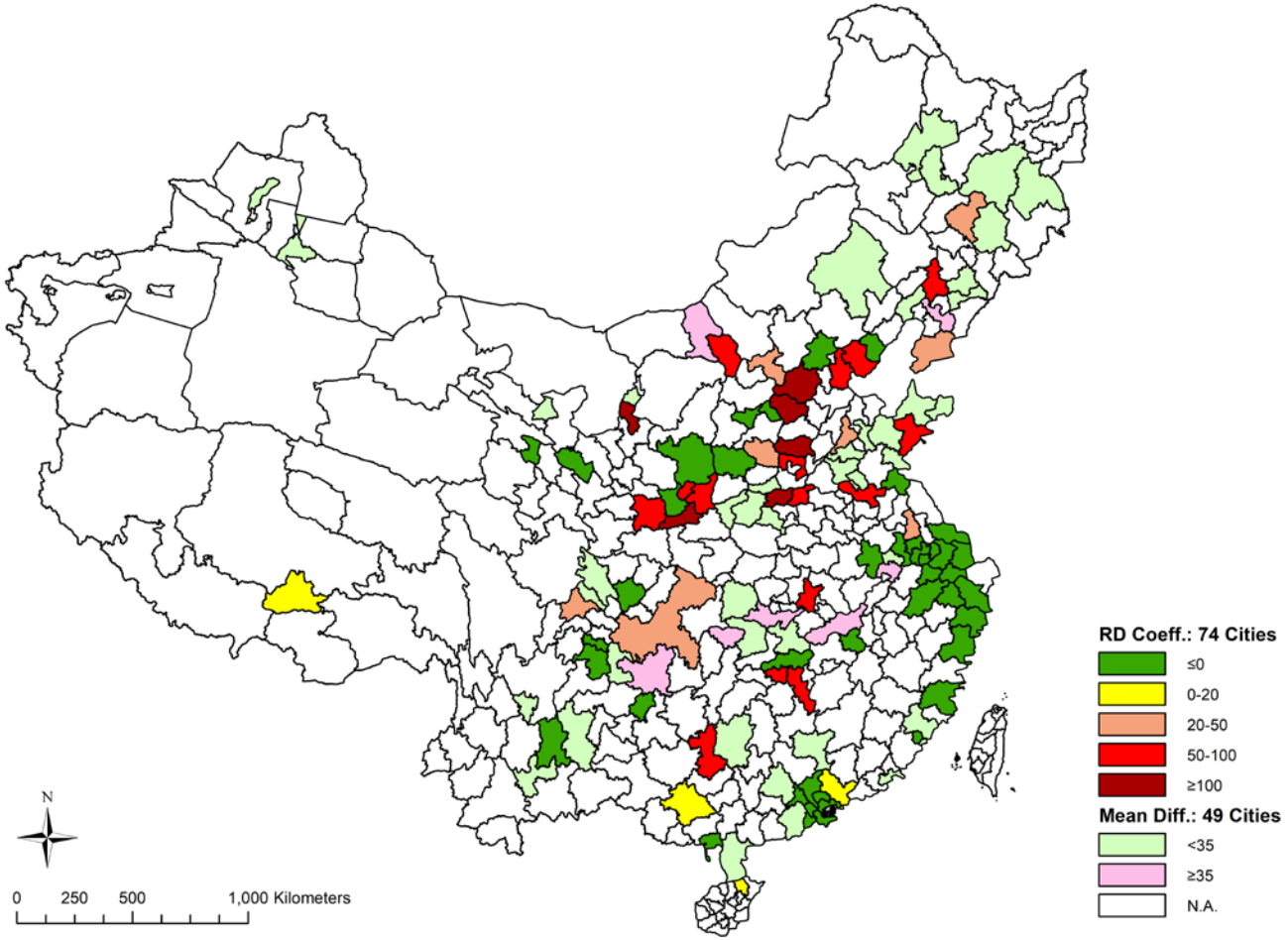
(C). Daily Residual Relative Humidity



(D). Daily Residual Wind Speed

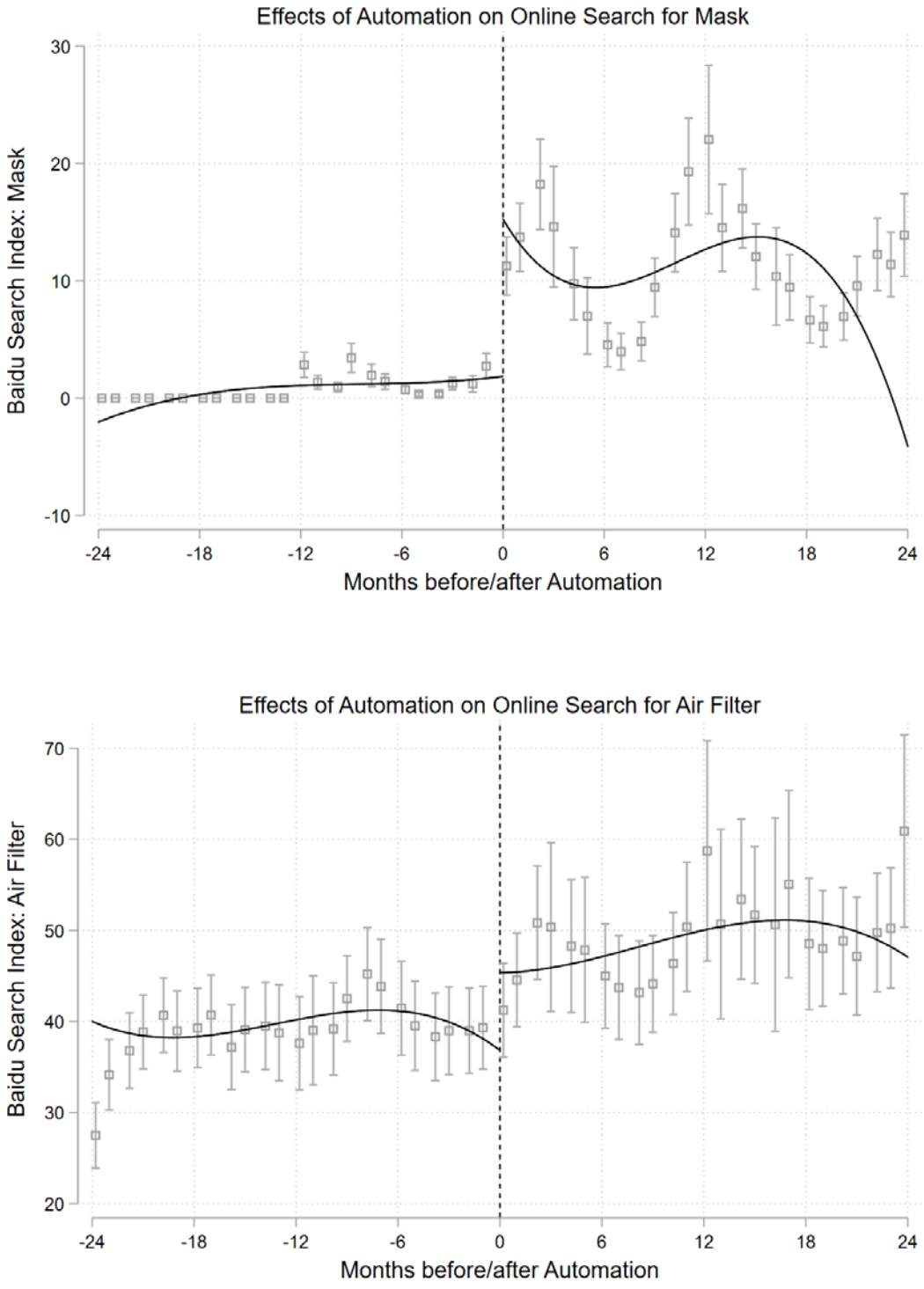
Notes: Station fixed effects and month fixed effects are absorbed before plotting the discontinuities.

Figure S9. Manipulation Status in Chinese Cities



Notes: The PM_{10} manipulation status in Chinese cities are plotted. Manipulation is confirmed if the local linear RD estimate is significantly positive at 5% for 74 cities, or if the difference in average PM_{10} one year before and after automation is no less than 35 $\mu g/m^3$ for 49 cities with missing values.

Figure S10. Online Search before and after Automation



Notes: The online search indices for anti-haze mask and air filter before and after automation (with 95% CI) are plotted in the upper and lower panel, respectively.