NBER WORKING PAPER SERIES

CAN TECHNOLOGY SOLVE THE PRINCIPAL-AGENT PROBLEM? EVIDENCE FROM CHINA'S WAR ON AIR POLLUTION

Michael Greenstone
Guojun He
Ruixue Jia
Tong Liu

Working Paper 27502
http://www.nber.org/papers/w27502

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2020

We thank Christoper Knittel, Liguo Lin, Alberto Salvo, Shaoda Wang, Bing Zhang, Junjie Zhang, and seminar participants at CIFAR, LSE, MIT, Nanjing University, NBER, and Peking University for their comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w27502.ack

Can Technology Solve the Principal-Agent Problem? Evidence from China's War on Air Pollution
Michael Greenstone, Guojun He, Ruixue Jia, and Tong Liu
NBER Working Paper No. 27502
July 2020
JEL No. Q53,Q55

ABSTRACT

We examine the introduction of automatic air pollution monitoring, which is a central feature of China's "war on pollution." Exploiting 654 regression discontinuity designs based on city-level variation in the day that monitoring was automated, we find that *reported* PM_{10} concentrations increased by 35% immediately post–automation and were sustained. City-level variation in underreporting is negatively correlated with income per capita and positively correlated with true pre-automation PM_{10} concentrations. Further, automation's introduction increased online searches for face masks and air filters, suggesting that the biased and imperfect pre-automation information imposed welfare costs by leading to suboptimal purchases of protective goods.

Michael Greenstone
University of Chicago
Department of Economics
1126 E. 59th Street
Chicago, IL 60637
and NBER
mgreenst@uchicago.edu

Ruixue Jia
School of Global Policy and Strategy
University of California at San Diego
9500 Gilman Drive #0519
La Jolla, CA 92093
and NBER
rxjia@ucsd.edu

Guojun He
Division of Social Science
HKUST
Clear Water Bay
Kowloon
Hong Kong
guojun.he@gmail.com

Tong Liu
Division of Social Science
The Hong Kong University of Science
and Technology
Clear Water Bay
Kowloon
Hong Kong
tliuaj@connect.ust.hk

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; Bénabou and Tirole, 2006). An alternative 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. A salient 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 cheat. 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 pollution information is designated to local officials (i.e. the agents). Due to the historically high cost to verify local information and the near-term benefits to high pollution in terms of economic growth, local officials have strong incentives to manipulate air pollution data before reporting them to the central government. The data quality problem is likely to impose significant costs, because it allows for inefficiently high ambient concentration of pollution, causes individuals and organizations to undertake inefficient levels of defensives investments, complicates government efforts to undertake international agreements to reduce emissions,[1] and, on the research side, raises questions about the credibility of linking pollution measures to key outcomes such as life expectancy and human capital (Ebenstein et al. 2017; Ebenstein and Greenstone 2020).

This paper examines the introduction of automatic pollution monitoring as a key part of China's extraordinarily successful "war on pollution" (Greenstone et al. 2020). The aims of monitoring automation 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.[2] In this context, automatic monitoring

---

[1] 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.

[2] From our discussion with regulators, we learn that typical ways of manipulation 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.

enables real-time sharing of data with the central government and the public and improvements in quality assurance checks through cross-validation statistical tools. This change greatly increases the costs for local governments to influence or manipulate the data and provides a compelling context to investigate the efficacy of technology to limit the hidden actions of local officials.

Our 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 (RD) designs to test for manipulation. Moreover, the implementation date varies across cities, allowing for event-study designs that provide a longer-run test for manipulation.

There are three key findings. First, there is striking evidence of the underreporting of air pollution concentrations before automation and improvement in data quality after automation. The station-level RDs 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 35% (relative to the post-automation mean of 99.5 $\mu g/m^3$) on average, just after monitoring was automated. Moreover, the increase in *reported* PM_{10} concentrations post-automation is also evident in DiD designs that exploit the variation in the timing of automation across cities, indicating that the higher recorded concentrations were a longer-run phenomenon. In comparison, there is no discontinuity in Aerosol Optical Depth (AOD) around the automation date, suggesting that satellite-based measure of air quality did not change right after implementation. We also provide corrected pre-automation PM_{10} data that are derived from the application of an artificial neural network to post-automation data on PM_{10}, AOD, and weather data, which we believe could be useful for other researchers.

Second, the estimation of city-specific RD designs produces quantitative measures of the degree of government misconduct for 74 cities. It is rare to have such measures and the variation across cities is striking: underreporting is indicated in 33 cities, because they have associated positive estimates that are statistically significant at the 5% level and 12 of these statistically significant estimates exceed 75 $\mu g/m^3$. We explore several potential explanations for this variation and the most consistent findings are that city-level underreporting is negatively correlated with city-level GDP per capita and positively correlated with the true pre-automation PM_{10} concentrations.

Third, we find that the introduction of automated monitoring likely increased investments in goods that protect individuals from air pollution. Specifically, we find a sharp increase in online searches for air filters and anti-haze face masks right after automation and this increase was sustained. Since automation was accompanied by increased efforts to inform the public about air pollution, it is not possible to isolate whether the changes in internet search behaviors were due to individuals updating their estimates of air pollution concentrations or learning about air pollution more generally or some combination of both. Regardless, it seems apparent that individuals' investments in defensive measures were below their optimum before automation, suggesting that biased and imperfect information imposed meaningful welfare costs.

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. 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 affect public sector governance and efficiency (e.g., Orphanides 2001; Duflo et al., 2012; Muralidharan et al., 2016). Our study provides one example in environmental regulation and the implications are likely to matter in other areas of monitoring and regulation as well.

Second, we contribute 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 environmental regulation, which matters for researchers, policymakers, 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 that information disclosure increased 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, our study is also related to the literature that assesses the reliability of data from China. In particular, Chen et al. (2012) and 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 is closely related to city characteristics. Researchers should be cautious about China's pre-automation air pollution data, particularly in less developed and more polluted cities.

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, 2020).

To meet the public need and gain 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 the 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. An automated nationwide monitoring network was established to collect and report pollution information.

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, the equipment and method measuring PM_{10} (as well as SO_2 and NO_2) are unchanged, assuring that differences in PM_{10} (as well as SO_2 and NO_2) if any, are not due to changes in equipment or method. 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 validation and higher-standard requirements on measurement.

Before automation, local environmental bureaus collected data and 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 sent to the central government in real time. The introduction of the Internet of Things and the improvement in surveillance technology further allow remote control of data measurement and transmission, as well as quality assurance and quality control. For example, with the new system, inconsistencies across different monitoring sites that are geographically close would trigger alerts automatically, allowing the central government to further investigate the causes. In addition, with the availability of real-time data, a higher standard also applies to how to measure pollutants. 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}.

With the new system, the concentrations of different air pollutants from more than 1,600 monitoring stations are updated on an hourly basis and are available simultaneously on the Ministry of Ecology and Environment's website, provincial and municipal environmental bureaus websites, as well as a large number of mobile apps and third-party websites.

C. Implementation across Cities

The automated monitoring system was introduced into different prefectural cities in three waves, as planned by the central government (see the map in Appendix A1). In the first wave, 74 key polluting cities (with 496 stations) were required to finish the upgrade by January 1st, 2013. In the second wave, another 116 cities (with 449 stations) were ordered to join by January 1st, 2014. The third wave further required the remaining 177 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 was implemented at the city level, different stations within a city were automated on the same day, giving us power to conduct RD analysis separately for each city.

III. Data and Summary Statistics

Station-Daily Data on Pollutants and Weather

The station-level air pollution data are published by the Ministry of Ecology and Environment and local environmental departments and were continuously collected by us between 2011 and 2016. We geocode the exact location of each monitoring station using Google Map. We also collect meteorological data from 403 weather stations, which include daily average temperature, precipitation, relative humidity, and wind speed. We match each pollution station with its closest weather station.

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 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 $PM_{2.5}$, a subset of PM_{10}.[3] 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 (Hoff and Christopher, 2009).

Automation 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 A2). One concern is that the local governments might strategically choose the automation dates to hide pre-automation underreporting by choosing a time of year when pollution concentrations decline for seasonal reasons. To address this, we also report results using the subgroup of cities that automate on their deadlines, under the presumption that this was not a strategic choice to hide manipulation.

Behavioral Responses

We measure individuals' behavioral responses through online searches. We focus on Baidu's search indices for "anti-haze face masks" and "air filters." Baidu is the biggest search engine in China and provides search indices for specific keywords that are analogous to Google Trends. The search indices are available from both PC and mobile terminals. We focus on the indices from PCs as the mobile data were not available before May 2013.

The Baidu index measures city-level search volume for a specific keyword during a specific period. Although high-frequency purchase data were unavailable prior to automation, as shown in Appendix Table D, the Baidu search indices are highly correlated with actual sales data post automation. We thus believe that it is reasonable to assume that online searches predict purchases. In fact, the internet search advertising business model is built on this idea.

Descriptive Patterns

The summary statistics are presented in Appendix A3. Even in the yearly data, we see that the reported PM_{10} concentrations significantly increased in 2012 and 2013. In comparison, there is a downward trend in AOD during the entire sample period, suggesting an overall improvement in air quality. In Appendix A4, we provide additional descriptive evidence that the reported PM_{10} readings could be systematically different before and after automation in the yearly data. We also present four case studies illustrating PM_{10} readings at the monitor level (Appendix A5).

[3] 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.

IV. Evidence on the Improvement of Air Quality Data

A. Short Rum Changes in PM_{10}: Evidence from RD Designs

We use an RD design based on the exact dates of air quality monitoring automation to detect air quality data manipulation:

$$P_{i,c,t} = \beta_1 I(t \geq Auto_{i,c,t}) + \beta_2 \mathrm{f}(t - Auto_{i,c,t}) + \beta_3 \mathrm{I} * \mathrm{f}(t - Auto_{i,c,t}) + \beta_4 \mathrm{W}_{i,c,t} + \alpha_i + month_t + u_{i,c,t} \quad (1)$$

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 the running variable. The specification includes a function $\mathrm{f}(t - Auto_{i,c,t})$ and allows its effect to differ pre and post automation, which is the basis of the "control function" style approach of the RD design. 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. Weather conditions, $\mathrm{W}_{i,c,t}$, include temperature, precipitation, relative humidity, and wind speed. $u_{i,c,t}$ 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. The discontinuity can be estimated by both parametric and non-parametric methods. We emphasize the results from the non-parametric method and use the parametric method as a robustness check.

In the simplest form, we do not include any fixed effects or control variables in the regression, as the dates of automation are arguably exogenous. To include covariates in the non-parametric RD, we first "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).[4]

We start by visualizing the patterns in the data. In Figure 1(A), the X-axis indicates the number of days before and after automation. The Y-axis indicates the reported daily PM_{10} concentrations after adjustment for monitoring station fixed effects, month fixed effects and meteorological conditions (the unadjusted data are plotted in Appendix B1). The residualized concentrations are estimated from running an OLS regression in which the dependent variable is daily reported PM_{10}

[4] Alternatively, we include covariates in the non-parametric RD analysis using the methodology developed by Calonico et al. (2019) and obtain similar results.

and the explanatory variables are station fixed effects, month fixed effects, and weather controls. We observe a striking increase in PM_{10} immediately after automation.

We present RD estimates from equation (1) in Panel A of Table 1. Columns (1) and (2) report the results from the local linear RD with and without covariates (i.e. $W_{i,c,t}, \alpha_i, month_t$). The bandwidths are 109 days and 263 days for PM_{10} in the two columns, respectively, which are the optimally selected according to the Calonico et al. (2014)'s method. Both use a triangle kernel weighting function. The estimated discontinuity is around 35 μg/m^3, which is a 35% increase, relative to the overall post-automation mean (99.5 μg/m^3). The magnitudes of the estimates are also economically meaningful. Based on Ebenstein et al. (2017), for example, a permanent 35 μg/m^3 difference in PM_{10} concentration implies a loss in life expectancy by 2.24 years for an average person living in China.

The remaining columns analyze subsets of the monitors. In columns (3) and (4), the PM_{10} levels increased 28 μg/m^3 (or 33%) for Wave-1 cities and 65 μg/m^3 (or 76%) for Wave-2 cities. In column (5), 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 is less of a concern. We find an increase of 57 μg/m^3 for this subgroup, implying that automation indeed increased reported PM_{10} concentration.

We provide several sets of results that together lend additional credibility to the baseline findings. First, we fit equation (1) for AOD, with the time being measured in months instead of days, to investigate whether there is a discrete change in AOD after automation. We find no discontinuity in AOD levels after automation (Figure 1(B) and Table 1(A)), confirming that this measure of true air quality did not deteriorate after the automation.[5] In fact, after adjustment for seasonality and weather, the estimated RD coefficient is precisely zero (column (2)).[6] Second, we find that all the weather variables are continuously distributed across the threshold (reported in Appendix B2), suggesting that the dramatic changes in the reported PM_{10} levels were not driven by weather conditions. Finally, we use alternative kernel weighting methods and the parametric approach to check the sensitivity of our findings. As reported in Appendix B3, both exercises yield similar estimates.

[5] An alternative measure of actual air quality is the $PM_{2.5}$ and AQI readings from US consulates in five Chinese cities. We do not observe significant RD in US consulate data, too.

[6] If we aggregate PM_{10} data to the monthly, we obtain similar results. See Appendices B1 and B3.

B. Medium-run Changes in PM_{10}: Evidence from Difference in Differences (DiD) Designs

The RD approach offers a demanding test of the effect of automation immediately after its implementation. We complement the main analysis by estimating a set of difference in differences (DiD) 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:

$$P_{i,c,t} = \gamma_\tau \sum_{\tau=-4}^{\tau=+3} Auto_{i,c,\tau} + \beta W_{i.c.t} + \alpha_i + month_t + \epsilon_{i,c,t} \quad (2)$$

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 are the station fixed effects and $month_t$ are the month or year-by-month fixed effects (we will show results for both specifications). $Auto_{i,c,\tau}$ indicate different periods before and after the automation, and we set the pollution readings 1–2 months before the automation date as the reference group (τ=-1). Then, $\tau \in \{-4, -3, -2, 0, 1, 2, 3\}$ respectively refers to 7–12 months, 5–6 months, and 3–4 months before automation, and 1–2 months, 3–4 months, 5–6 months, and 7–12 months post automation.

The coefficients of γ_0, γ_1, γ_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} levels 1-2 months before automation).

To avoid any composition change this dynamic analysis, we restrict the sample to cities that automated their monitoring stations only at the deadline of their respective wave and use data from 2012 January 1 to 2013 December 31. Thus, the "treatment" monitors are from cities where automation occurred on 2013 January 1 and the "control" monitors are from cities where automation never occurred during this two-year period. With this set-up, the "control" monitors are never treated during this period and provide a plausibly credible counterfactual for the "treatment" monitors. Further, this is one approach to confronting the challenges associated with the staggered assignment of treatment. [7]

The results are reported in Panel B of Table 1. We find that reported PM_{10} concentrations are substantially higher post-automation. While the increase in levels generally declines over time (see columns (1)–(4)), it is relatively stable when the natural logarithm of PM_{10} is the dependent variable,

[7] An emerging literature shows that estimates from two-way fixed effects models with staggered treatment assignment are difficult to interpret. See discussions in de Chaisemartin and D'HaultfŒuille (2018, 2019); Goodman-Bacon (2018); Imai and Kim (2018).

in which the coefficients are approximations to percentage changes (24% to 32% in column (5)). This difference in results is largely due to seasonality in China's pollution concentrations: 1–2 months post automation occur during the winter, which is the time of year when pollution concentrations are the highest in China. Overall, we conclude that automation led to sustained increases in reported PM_{10} levels.

We note that columns (1) and (2) exhibit some pre-automation differences in pollution concentrations although it is difficult to discern a clear trend. Nevertheless, the columns (3)–(5) specifications aim to mitigate the possibility of confounding using nearest neighbor matching. Specifically, we match each monitoring station in the Wave 1 (deadline) cities to its (geographically) nearest monitoring station in the Wave 2 (deadline) cities with replacements and re-estimate equation (2) using the paired sample. The idea is that geographically adjacent pairs of monitors should have similar pre-automation trends in their reported PM_{10} levels. Indeed, after matching, the coefficients for all the "lead" variables become small in magnitude and statistically insignificant (as plotted in Appendix B4). However, the finding that reported PM_{10} concentrations increased post-automation remains unchanged.

C. Additional Evidence and Correcting Pre-automation PM_{10} Data

We provide additional evidence that data quality improved after automation. First, we examine the variability of PM_{10} under the presumption that manipulated measures are likely to exhibit less variability than true realizations. Indeed, we find that the standard deviation of monthly PM_{10} data increased by around 42% after automation (as presented in Appendix B5).

Second, we examine changes in SO_2 and NO_2, the other two pollutants available before and after automation. The underreporting 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 more than 90% of the time. In Appendix B6, we find a statistically significant post-automation discontinuity of 3.0 ppb or 6.0% in NO_2 concentrations and a statistically insignificant effect on reported SO_2 concentrations. These results are consistent with the hypothesis that underreporting of pollution readings was governed by local officials' incentives.

Third, we examine whether the higher level of reported-PM_{10} after automation is simply driven by higher reporting standards. The results presented in Appendix B7 show that changes in the data collection standard alone do not mechanically generate the RD estimates.

Fourth, we examine whether there is any bunching effect in post-automation PM_{10} data at the critical thresholds in the air quality standards. As shown in Appendix B8, we find no evidence of bunching, which again confirms that automation improves data quality and limits local governments' strategic underreporting.

Finally, we examine the correlation between PM_{10} and the AOD data, treating the latter as an unbiased measure. We find that the correlation/partial correlation between PM_{10} and AOD indeed became stronger after automation, confirming the improvement in data quality. These results are discussed in Appendix B9.

As a by-product of this paper, we attempt to correct the pre-automation PM_{10} data by exploiting the relationship between PM_{10}, AOD, and weather conditions. Specifically, we train an artificial neural network (ANN) and predict the pre-automation PM_{10} levels, assuming that the post-automation relationships between PM_{10}, AOD, and weather conditions can be carried to the pre-automation period (see Mullainathan and Spiess (2017) for method discussion). The ANN is able to explain 81% of PM_{10} variation post automation. The mean of the corrected PM_{10} concentration is 24.4 $\mu g/m^3$ or 29% higher than the pre-automation reported mean and the correction shifts the distribution of the pre-automation PM_{10} data to the right. These corrected PM_{10} data are available to other researchers and are provided as an online appendix (see Appendix B10 for details).

V. Variation in Data Quality and Welfare Implication

A. Data Quality in Different Cities

This section explores the heterogeneity in data quality across different cities. We estimate equation (1) city by city using the non-parametric approach, with the unit of observation being a monitor by day. Before proceeding, we note a change in the sample. Some cities suspended data reporting while they installed and tested the new automatic monitors for $PM_{2.5}$, CO, and O_3. As a result, 49 cities did not report PM_{10} readings for more than two months preceding the initiation of automatic monitor reporting and it is challenging to credibly apply the RD approach to these cities individually. Among them, 5 cities did not report PM_{10} readings for over six months and are dropped from the analysis in this section. Below we start with the 74 cities without missing data problem, although we include the other 44 cities in a robustness check. The results presented in the previous section are robust to including or excluding these cities.

Figure 2 plots the estimated city-specific RD coefficients and their 95% confidence intervals. The RD coefficient is positive for more than 70% of these cities. Among them, 33 cities' 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 17.1 $\mu g/m^3$, as denoted by the red horizontal line. We observe substantial variation in these estimated effects of automation. Particularly noteworthy findings are that there are 12 cities with the estimated discontinuities greater than 75 μg/m^3 and 11 of them would be judged statistically significant at the 1% level. The spatial variation in manipulation by city is plotted in Appendix C1.

Why were some cities more likely to manipulate air pollution data than other cities? Our quantitative city-specific RD estimates provide an opportunity to investigate the underlying incentives of such hidden actions, which we will link to city and leader characteristics.

Here we do not attempt to identify a causal impact but to provide cross-sectional correlations between city-leader characteristics and manipulation. We define manipulation in three ways. The first is a binary indicator of manipulation if the RD estimate is positive and significant at 5% level for the 74 cities in Figure 2. The second is also a binary indicator, but we extend it to include 49 cities that have missing data issues. For these cities, we compare PM_{10} levels between January–June 2013 and January–June 2014: if the reported average increased by 35 μg/m^3 in the city (the average discontinuity in Table 1), we define it a data manipulating city (13 cities meet this criterion).[8] For the other 31 cities, 10 cities have positive RD estimates that are significant at 5% level, and we treat them as data-manipulating cities. Following the second definition, 56 cities are defined as data-manipulating. The third is to directly use the RD coefficients as the measure of manipulation in the 74-city sample, and we weight the regression by the inverse of the standard errors to assign heavy weights to cities with more accurate RD estimates.

We focus on three explanatory variables and examine their correlation with the manipulation measures in Table 2. The first is GDP per capita measured in 2012, as this is likely to capture 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 Table 2, regardless of how we define manipulation (Panels A to C). So, for example, a 1 standard-deviation (SD) increase in the logarithm of GDP per capita is associated with a 12% decrease in the probability of manipulation (column (1) of Panel A and Panel B) and a decrease in the magnitude of the RD estimate of roughly 12 μg/m^3 (column (1) of Panel C). The second is the city's corrected PM_{10} concentration in 2012 derived from the ANN prediction. It is apparent that true pre-automation pollution concentrations are strongly and positively correlated with manipulation, which is consistent with local leaders facing sanctions for allowing high pollution concentrations. The third variable is a corruption index that measures the share of a city's civil servants that were convicted for corruption during China's anti-corruption campaign (Nie et al. 2018). We standardize the index so that a higher value indicates more

[8] Using slightly different thresholds (20, 30, or 40) yields qualitatively similar results.

corruption. We do not find the corruption index to be correlated with any measures of manipulation.

Column (4) reports on "horse race" style regressions that include all three covariates. The findings that GDP per capita and pre-automation PM_{10} concentrations are negatively and positively, respectively, correlated with manipulation remain unchanged. Cities with more corrupt officials seem to be more likely to underreport data, but the evidence is not strong. In Appendix C2, we further present additional results on the correlation between the manipulation measures and the personal characteristics of the city leaders (Party secretaries and mayors). Overall, city characteristics appear to have more explanatory power than leader characteristics.

B. Welfare Implications

This subsection examines whether the improvement in the quality and availability of pollution information had any welfare implications. Specifically, we test whether the number of online searches for "anti-haze face masks" and "air filters" change immediately after automation. We focus on people's online searching behaviors because such data are available in all our sampled cities during the study period and are strong predictors of actual purchases (Appendix D).

The graphical RD estimates are presented in Figures 1(C) and 1(D) for face masks and air filters, respectively. Panel A of Table 3 reports on the fitting of the non-parametric RD version of equation (1). The estimates indicate that monthly online searches for "face masks" immediately tripled (columns (1)–(2)) after automation and searches for "air filters" increased by 17–20%. Column (3) limits the sample to "normal" cities where we fail to detect underreporting and column (4) focuses on data-manipulating cities according to the second definition from the previous subsection (Panel B of Table 2). We find that in data-manipulating cities, the post-automation increase in searches was even larger, more than three times for "anti-haze face masks" and 26% higher for "air filters." It is tempting to interpret these larger estimates in column (4) as being entirely due to individuals' learning that PM_{10} concentrations were higher than they had believed, but it is also possible that the increase in news about air pollution disproportionately increased in these cities at the same time.

Panel B of Table 3 reports on the estimation of equation (2). We find that the higher rates of searches for these two terms were sustained and still event 7-12 months after automation. Based on these results and the positive correlation between searches and purchases, it seems reasonable to assume that automation led to an immediate and sustained increase in the purchases of goods that protect individuals from PM_{10}.

We do not believe that it is possible to isolate whether the post-automation behavioral changes were due to individuals updating their estimates of air pollution concentrations or learning about air pollution more generally or some combination of both. Regardless, it is evident that the biased and imperfect information about air pollution imposed meaningful welfare costs prior to automation.

VI. Conclusion

Governments delegate authority to bureaucratic officials, which makes the principal-agent problem inherent to 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 its "war on pollution" in recent years – it is difficult to imagine an effective policy without reliable information. Besides, we show that the more reliable information post automation appears to have induced more people to take avoidance behaviors against pollution, which implies welfare gains that are of a potentially significant magnitude. That said, new monitoring and surveillance technologies are likely to have other important implications for governance about which there is much to learn. Our study is just one example of the consequences of technological advancement for governance and we believe that this is a rich area for research going forward.

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. From fog to smog: The value of pollution information (No. w26541). National Bureau of Economic Research.

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., 2019. 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.

Cattaneo, M. D., Jansson, M., & Ma, X. (2019). Simple local polynomial density estimators. *Journal of the American Statistical Association*, 1-7.

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. and D'HaultfŒuille, X., 2018. Fuzzy differences-in-differences. *The Review of Economic Studies*, 85(2), 999-1028.

de Chaisemartin, C. and D'HaultfŒuille, X., 2019. Two-way fixed effects estimators with heterogeneous treatment effects (No. w25904). National Bureau of Economic Research.

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.

Ebenstein, A., Fan, M., Greenstone, M., He, G. and Zhou, M., 2017. New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy. *Proceedings of the National Academy of Sciences*, *114*(39), pp.10384-10389.

Ebenstein, A. and Greenstone M. 2020. Childhood exposure to particulate air pollution, human capital accumulation and income: Evidence from China. Working Paper.

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 (No. w25018). National Bureau of Economic Research.

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.

Greenstone, M., He, G., Li, S. and Zou., E., 2020. Pollution in China: Recent trends and empirical evidence. Working Paper.

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, pp.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., 2020. Willingness to pay for clean air: Evidence from air purifier markets in China. *Journal of Political Economy*, *128*(5), pp.000-000.

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

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. http://www.niehuihua.com/a/chuban/487.html.

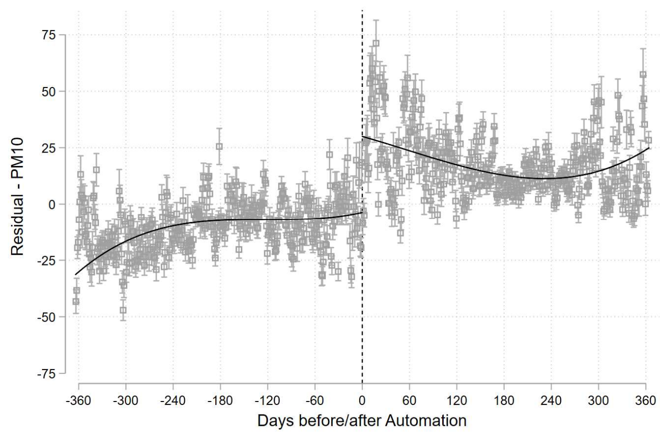
Orphanides, A., 2001. Monetary policy rules based on real-time data. *American Economic Review*, *91*(4), pp.964-985.

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

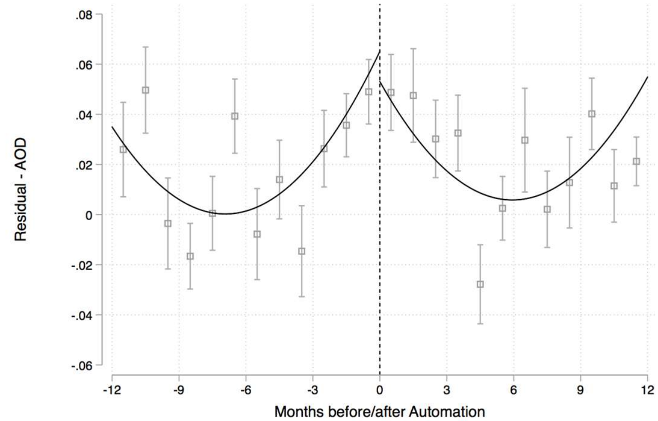
Williamson, O.E., 1996. *The mechanisms of governance*. Oxford University Press.

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

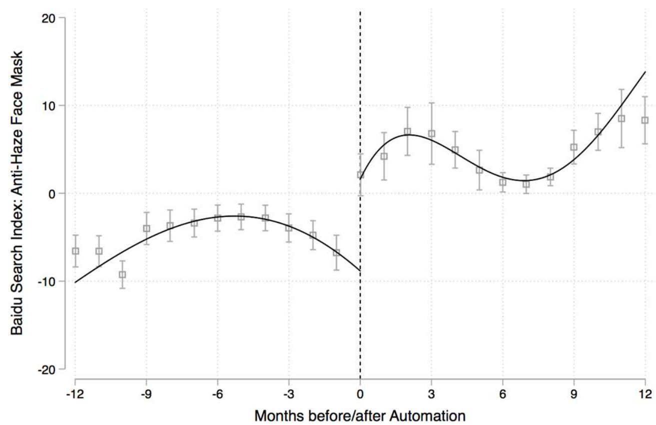
Figure 1. RD Plots for PM_{10}, AOD and Online Search



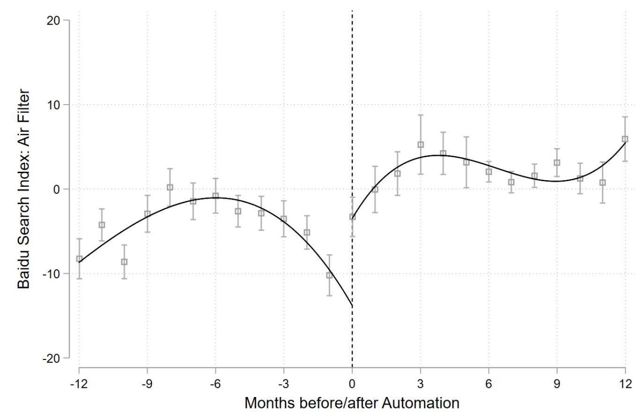
(A). Daily Residual PM_{10}



(B). Monthly Residual AOD



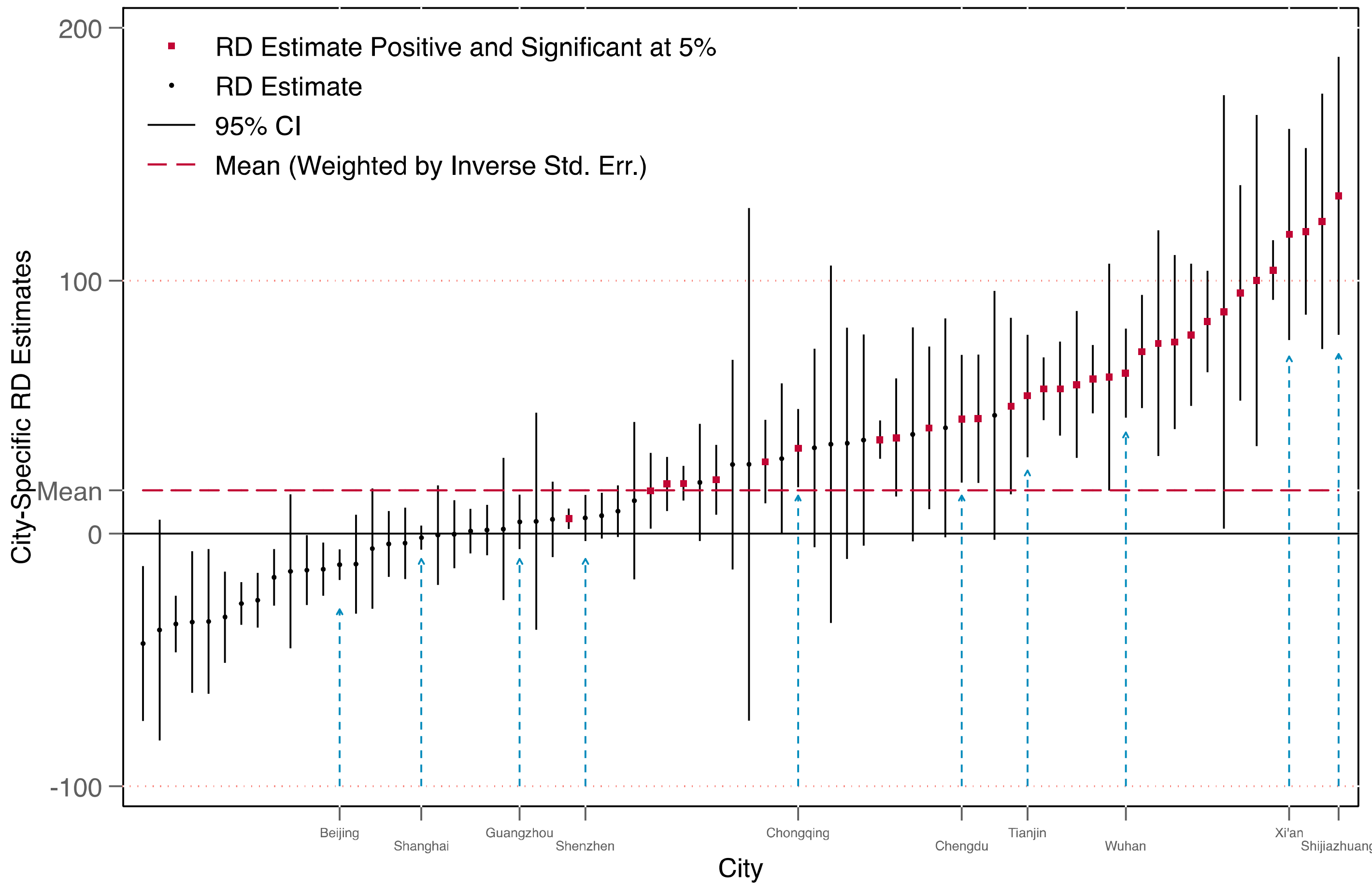
(C). Monthly Residual Face Mask Search



(D). Monthly Residual Air Filter Search

Notes: Panel (A) shows the increase in PM_{10} immediately after automation using daily data. Panel (B) shows no significant change in monthly AOD data. Panels (C) and (D) show the increase in monthly online search data for anti-haze face masks and air filters. 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 RD estimates for PM_{10} ($\mu g/m^3$) for the 74-city sample are plotted with 95% confidence intervals. The weighted average RD estimate is denoted by the red dashed line, with the weights being the inverse of the standard errors of the RD coefficients.

Table 1. Automating Air Quality Monitoring System and PM_{10} Concentrations

	(1)	(2)	(3)	(4)	(5)
Panel A. RD Estimates					
RD in PM_{10} (Daily)	34.7***	34.9***	27.5***	64.7***	57.1***
	(10.7)	(5.8)	(9.8)	(9.9)	(8.6)
RD in AOD	0.065	-0.005	0.026	-0.030	-0.003
	(0.044)	(0.021)	(0.031)	(0.029)	(0.025)
Sample	All	All	Wave 1	Wave 2	Deadline
Station FE		Y	Y	Y	Y
Month FE		Y	Y	Y	Y
Weather Controls		Y	Y	Y	Y
Obs. (Daily)	91,470	232,326	81,950	68,456	86,042
Bandwidth (Days)	109	263	140	234	184
Obs. (Monthly)	5,057	5,851	3,173	2,316	4,894
Bandwidth (Months)	6	7	7	6	10
Panel B. Event-Study Estimates					
	PM_{10}	PM_{10}	PM_{10}	PM_{10}	Log(PM_{10})
7-12 Months before	-8.5*	-17.2**	-10.7	-10.8	-0.13*
	(4.7)	(6.7)	(7.7)	(9.7)	(0.07)
5-6 Months before	6.8	-19.2**	10.5	-2.2	0.02
	(6.0)	(9.3)	(8.5)	(12.1)	(0.11)
3-4 Months before	-6.4	-12.0*	-2.8	-5.2	-0.03
	(5.6)	(6.9)	(7.3)	(9.2)	(0.09)
1-2 Months after	60.3***	31.4***	66.5***	45.6***	0.24***
	(11.0)	(11.1)	(14.3)	(16.3)	(0.09)
3-4 Months after	45.0***	33.6***	47.2***	32.5**	0.32**
	(7.8)	(8.8)	(10.7)	(14.2)	(0.12)
5-6 Months after	28.1***	22.2***	33.4***	29.0**	0.29**
	(6.7)	(8.0)	(9.7)	(13.7)	(0.12)
7-12 Months after	40.0***	9.8	42.9***	15.8	0.24*
	(6.1)	(8.8)	(7.7)	(14.0)	(0.15)
Sample	Deadline	Deadline	+Matching	+Matching	+Matching
Weather Controls	Y	Y	Y	Y	Y
Station FE	Y	Y	Y	Y	Y
Month FE	Y		Y		
Year-Month FE		Y		Y	Y
R-Squared	0.34	0.35	0.33	0.34	0.38
Obs.	176,426	176,426	186,499	186,499	186,469

Notes: In Panel A, each cell represents a separate non-parametric RD estimate. Triangle kernel is used and optimal bandwidth is selected by Calonico et al. (2014)'s method. Columns (1) and (2) use the entire sample to estimate the discontinuities; there are 1,049,325 daily observations before bandwidth selection. Columns (3) and (4) use the Wave 1 and Wave 2 cities. Column (5) uses cities that automated the monitoring system at their deadlines. In Panel B, event-study estimates are reported and "1-2 months before automation" is the reference group. In columns (1) and (2), there are 242 Wave-1 (deadline) stations and 123 Wave-2 (deadline) stations. In columns (3) to (5), each Wave-1 (deadline) station is matched with its nearest Wave-2 (deadline) station (with replacement). Weather controls include temperature, relative humidity, precipitation and wind speed. Standard errors clustered at the city level are reported in parentheses below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

Table 2. Correlations b/w Data Quality and City Characteristics

	(1)	(2)	(3)	(4)
Panel A. Manipulation = 1 if RD > 0 (P < 0.05)				
ln(GDP per capita)	-0.12**			-0.08
(per 1 SD)	(0.06)			(0.06)
Corrected PM_{10}		0.18***		0.22***
(per 1 SD)		(0.04)		(0.06)
Corruption Index			-0.04	0.11**
(per 1 SD)			(0.06)	(0.05)
Observations	74	74	74	74
R-squared	0.05	0.15	0.01	0.19
Panel B. Manipulation = 1 if RD > 0 (P < 0.05) or Diff ≥ 35μg/m³ (P < 0.05)				
ln(GDP per capita)	-0.12**			-0.10***
(per 1 SD)	(0.04)			(0.04)
Corrected PM_{10}		0.15***		0.17***
(per 1 SD)		(0.04)		(0.04)
Corruption Index			-0.01	0.06
(per 1 SD)			(0.05)	(0.05)
Observations	118	118	118	118
R-squared	0.06	0.09	0.00	0.15
Panel C. Y = City Specific RD Estimates Weighted by the Inverse of Std. Err.				
ln(GDP per capita)	-11.66***			-7.94**
(per 1 SD)	(3.56)			(3.79)
Corrected PM_{10}		20.51***		20.95***
(per 1 SD)		(4.02)		(4.49)
Corruption Index			-6.68	4.19
(per 1 SD)			(8.22)	(4.54)
Observations	74	74	74	74
R-squared	0.09	0.29	0.03	0.34

Notes: In Panel A, the dependent variable is a dummy variable indicating manipulation for 74 cities that do not have missing data issues. If the city-specific RD estimate is positive and statistically significant at 5%, manipulation equals to 1. City-level GDP per capita in 2012 is used. The corrected PM_{10} data are obtained from ANN predictions and we also use the average predicted values in 2012 in the regression. The corruption index is standardized based on Nie et al (2018) to reflect the corruption level in a city. In Panel B, we further include 44 cities that have missing data issues. For these cities, we compare PM_{10} levels between January–June 2013 and January–June 2014: if the reported average increased by 35 $\mu g/m^3$ in the city, or if the city's RD estimate is positive and significant, we define it a data manipulating city. In Panel C, the dependent variable is the city-specific RD estimate weighted by the inverse of the standard error. Robust standard errors in parentheses. * significant at 10% ** significant at 5% *** significant at 1%.

Table 3. Automating Air Quality Monitoring System and Avoidance Behaviors

	(1)	(2)	(3)	(4)
Panel A. RD Estimates				
RD in Face Mask Searches	10.10***	11.03***	7.18***	15.43***
(pre-automation mean =0.62)	(1.58)	(1.66)	(2.41)	(2.29)
RD in Air Filter Searches	7.36**	8.73***	6.02**	13.44***
(pre-automation mean =35.5)	(3.60)	(1.86)	(3.00)	(2.15)
RD in Log (1+ Face Mask Searches)	1.06***	1.15***	0.96***	1.43***
(pre-automation mean =0.16)	(0.17)	(0.17)	(0.24)	(0.22)
RD in Log (1+ Air Filter Searches)	0.18*	0.16***	0.12***	0.23***
(pre-automation mean =3.30)	(0.10)	(0.04)	(0.06)	(0.05)
Sample	All	All	Normal	Manipulate
City FE		Y	Y	Y
Month FE		Y	Y	Y
Weather Controls		Y	Y	Y
Panel B. DiD Estimates				
	Mask Searches	Mask Searches	Filter Searches	Filter Searches
7-12 Months before	0.00	-0.05	-0.15	-0.14
	(0.00)	(0.09)	(1.31)	(1.31)
5-6 Months before	0.00	0.12	0.62	0.86
	(0.00)	(0.20)	(1.16)	(1.23)
3-4 Months before	0.00	0.17	0.91	1.16
	(0.00)	(0.11)	(1.12)	(1.15)
1-2 Months after	18.60***	18.52***	2.87	2.80
	(2.78)	(2.75)	(1.90)	(1.87)
3-4 Months after	17.39***	17.31***	6.11***	6.10***
	(2.86)	(2.85)	(1.78)	(1.77)
5-6 Months after	5.43***	5.37***	2.48	2.58
	(1.19)	(1.16)	(1.64)	(1.64)
7-12 Months after	14.45***	14.62***	6.00***	6.22***
	(2.31)	(2.29)	(1.72)	(1.74)
Sample	Deadline	Deadline	Deadline	Deadline
City FE	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y
Weather Controls		Y		Y
R-Squared	0.32	0.32	0.53	0.53
Obs.	51,901	51,900	51,170	51,169

Notes: In Panel A, each cell represents a separate RD estimate. Triangle kernel is used in all RD estimations. Columns (1) and (2) use the entire sample from the 123 cities, consisting of 8,661 mask search observations and 8,590 filter search observations before bandwidth selection. Column (3) limits the sample to "normal" cities where we fail to detect manipulation and column (4) focuses on data-manipulating cities according to the second definition in Panel B of Table 2. In Panel B, each column represents a separate fixed-effects regression. There are 39 Wave-1 (deadline) cities (treatment group) and 32 Wave-2 deadline cities (control group). Weather controls include temperature, relative humidity, precipitation and wind speed. Standard errors clustered at the city level are reported in parentheses below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

Appendix

A. Background and Data

- **A1. Map for Automation**
- **A2. Policy Dates Distribution**
- **A3. Summary Statistics**
- **A4. Descriptive Patterns in the Yearly Data**
- **A5. City-level Cases**

B. Additional Results on Data Quality pre-post Automation

- **B1. RD Using Raw Daily PM_{10} and Monthly PM_{10}**
- **B2. No Discontinuity in Weather Conditions**
- **B3. Additional RD Specifications for the Levels of PM_{10}**
- **B4. DiD Plots for PM_{10}**
- **B5. Variability in PM_{10}**
- **B6. Results for Other Pollutants**
- **B7. Changes in Data Collection Requirement**
- **B8. No Discontinuity at Categorical Cutoffs of PM_{10}**
- **B9. Correlation between PM_{10} and AOD pre-post Automation**
- **B10. Correcting the Pre-Automation PM_{10}**

C. Additional Results on Data Quality pre-post Automation

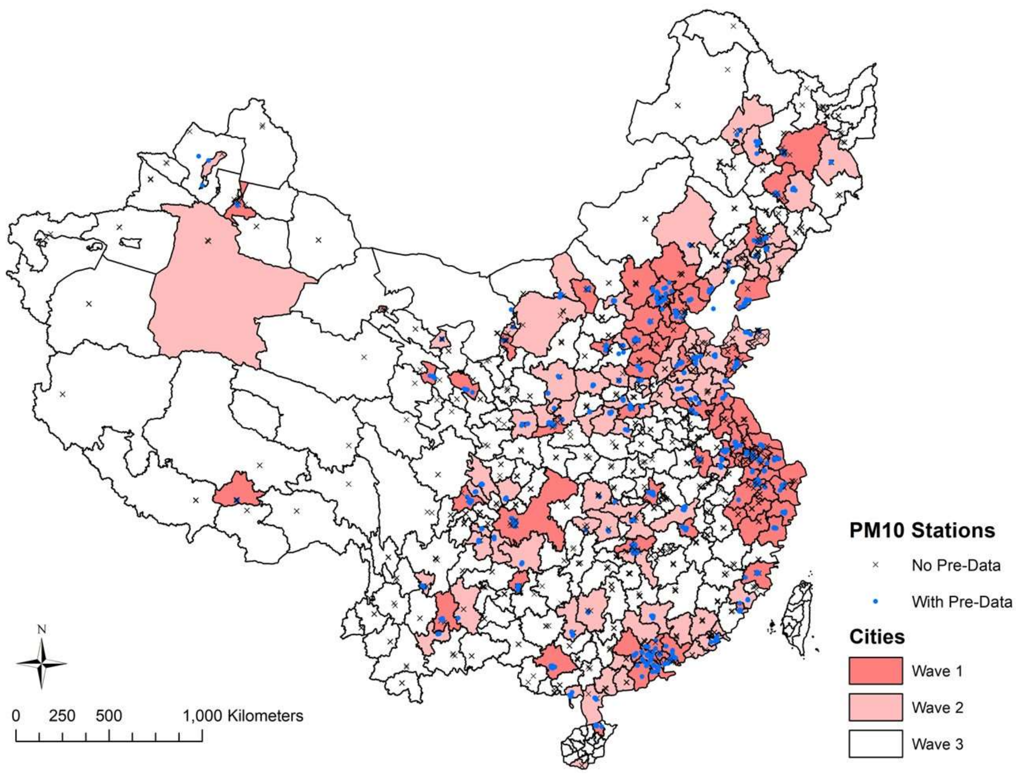
- **C1. Map of Manipulation Status across Chinese Cities**
- **C2. Manipulation and City/Leader Characteristics**

D. Association between Online Search and Sales

A. Background and Data

A1. Map for Automation

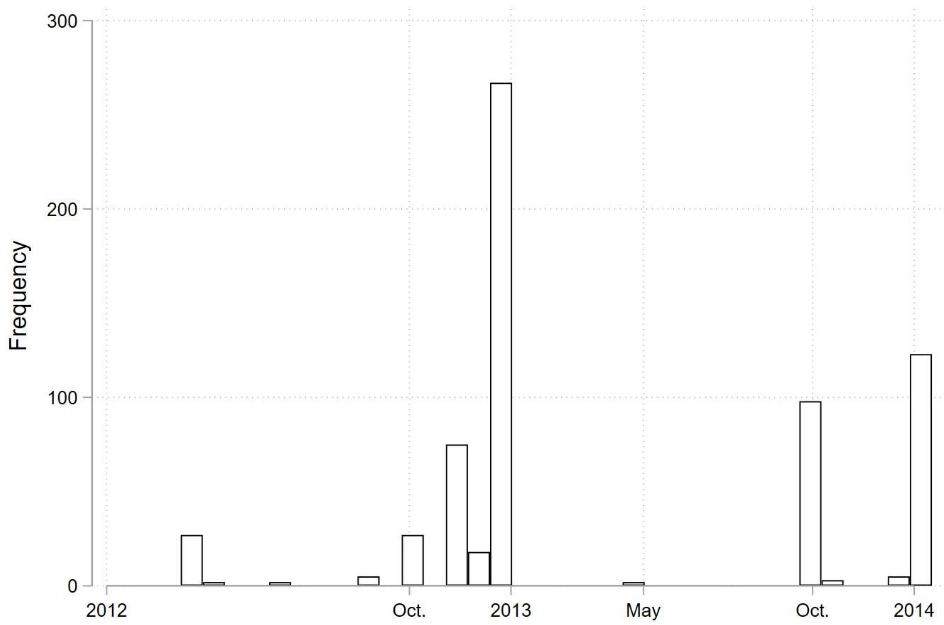
Figure A1. Waves in Automation



Notes: Wave 1, Wave 2 and Wave 3 cities are plotted. The dots represent PM_{10} monitoring stations where pre-automation data are available.

A2. Policy Dates Distribution

Figure A2. 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, which are the deadlines for the two waves.

A3. Summary Statistics

Table A3. 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	0.60	0.56	0.56	0.55	0.51	0.46
	(0.28)	(0.28)	(0.27)	(0.29)	(0.26)	(0.25)
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	14.6	14.7	15.4	15.5	15.6	15.4
(ºC)	(11.2)	(11.5)	(11.2)	(10.6)	(10.4)	(11.0)
Precipitation	2.4	3.5	3.4	3.3	3.7	4.1
(mm)	(7.4)	(10.2)	(11.0)	(10.3)	(11.4)	(12.1)
Relative Humidity	63.8	65.5	64.4	64.9	67.2	67.2
(%)	(18.1)	(19.1)	(18.7)	(19.1)	(19.1)	(19.2)
Wind Speed	2.2	2.6	2.7	2.6	2.7	2.8
(m/s)	(1.0)	(1.5)	(1.5)	(1.4)	(1.4)	(1.4)

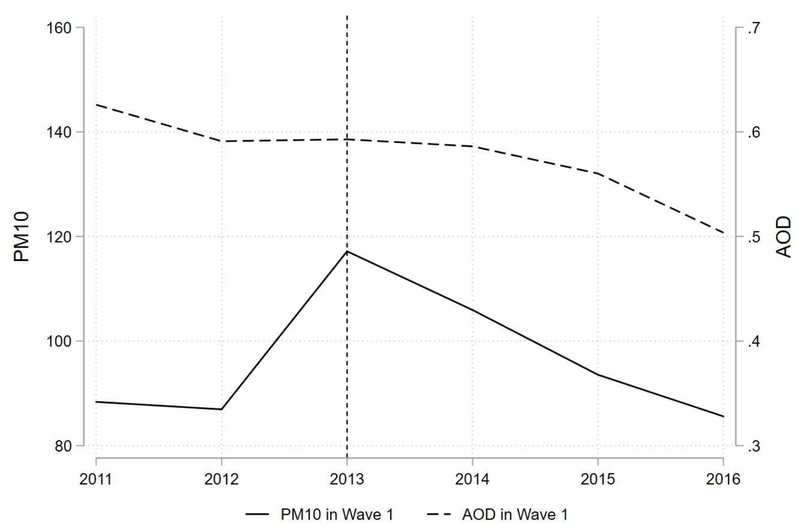
Notes: Daily air quality data are collected from China's air quality monitoring platform. Weather data are collected from local meteorological stations. AOD data are collected from MODIS.

A4. Descriptive Patterns in the Yearly Data

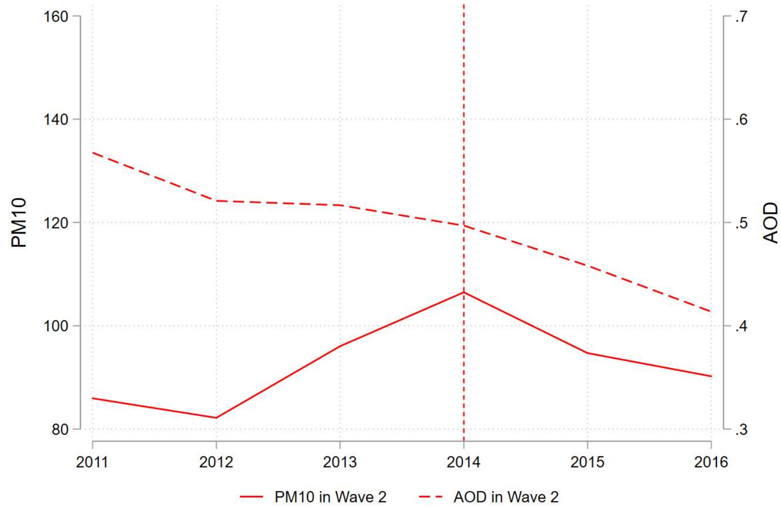
We describe two sets of empirical patterns as motivational evidence. First, even in the yearly data, we observe discontinuity in reported PM_{10} concentrations pre- and post-automation. Second, we present data from a few stations to illustrate the variation in reported PM_{10} levels across stations/cities.

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 Figure A4). 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 μg/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 A3 for the summary statistics of key variables).

Figure A4. Annual PM_{10} and AOD from 2011 to 2016



(A). Wave 1 Cities: PM_{10} and AOD



(B) Wave 2 Cities: PM_{10} and AOD

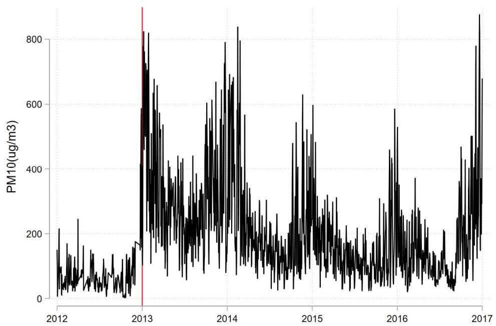
Notes: Annual average PM_{10} concentrations ($\mu g/m^3$) in Wave 1 and Wave 2 are plotted in black and red, respectively. Corresponding AOD levels are shown in dashed lines.

A5. City-Level Cases

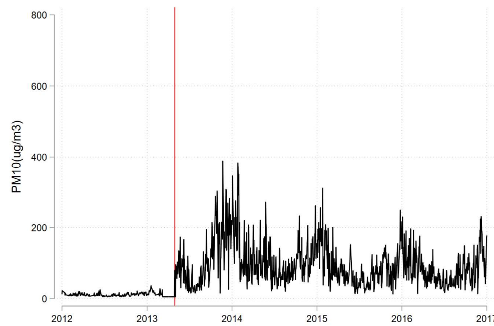
This subsection takes an admittedly selective examination of the reported time series from four stations as a means of highlighting the high geographic and temporal variation of the data and qualitatively previewing the finding of extensive manipulation in some locations before automation. For instance, in the monitoring station in the development zone of Shijiazhuang city (the upper left panel of Figure A5), 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 Figure A5), 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 Figure A5), 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.

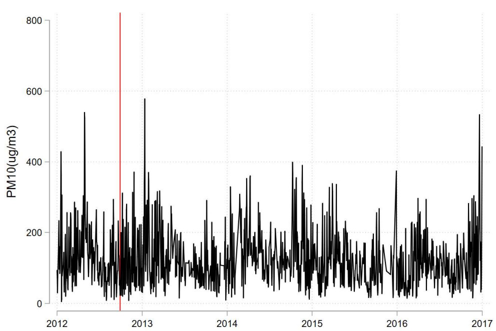
Figure A5. Times Series of PM_{10} Concentrations at Four Stations



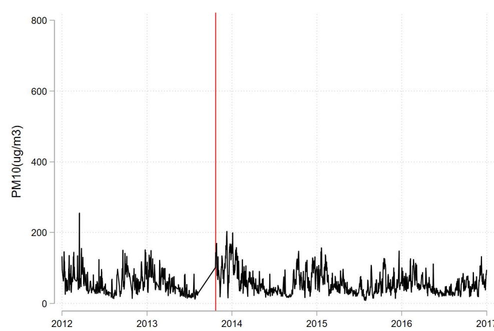
(A). Gaoxin District, Shijiazhuang City, Hebei



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



(C). Gucheng, Beijing



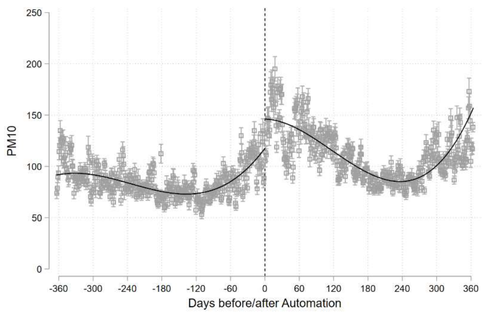
(D). Industrial Park, Beihai City, Guangxi

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.

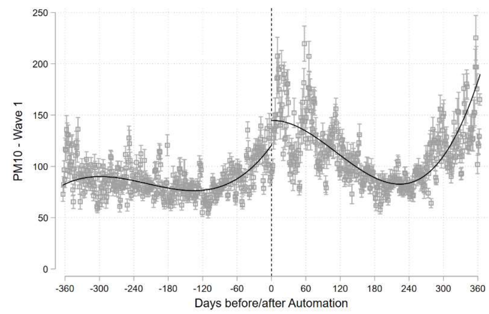
B. Additional Results on Data Quality pre-post Automation

B1. RD Using Raw Daily PM_{10} and Monthly PM_{10}

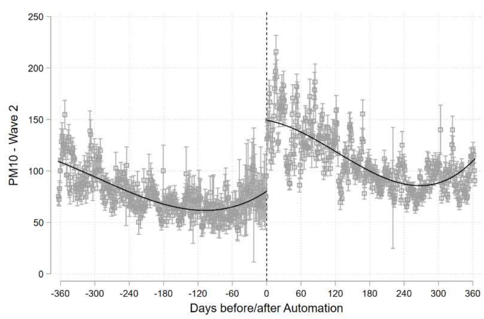
Figure B1. RD Plots Using Raw PM_{10} Data



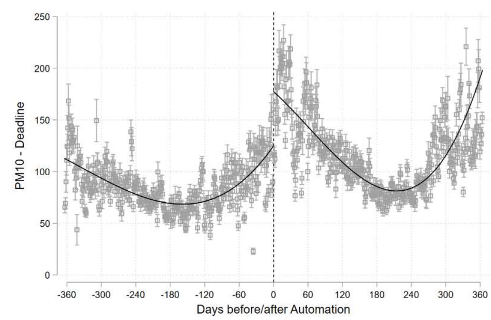
(A). Daily PM_{10}



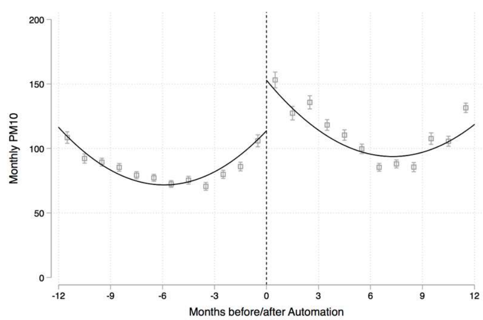
(B). Daily PM_{10} in Wave 1



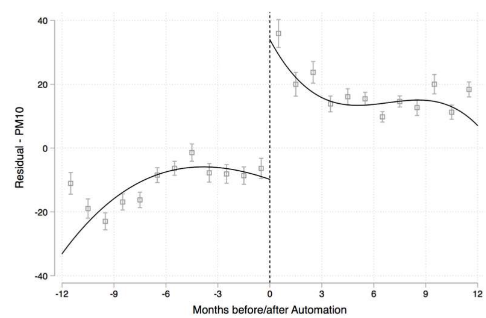
(C). Daily PM_{10} in Wave 2



(D). Daily PM_{10} in Deadline



(E). Monthly PM_{10}



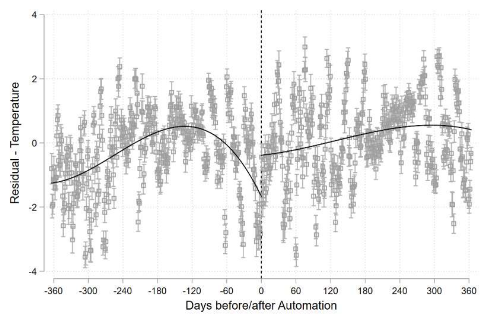
(F). Monthly Residual PM_{10}

Notes: In Panels (A) – (D), the discontinuities are plotted using raw daily PM_{10} concentrations (no controls are included). In Panels (E) and (F), the discontinuities for station-month PM_{10} and the residuals (absorbing station, month fixed effects and weather) are plotted.

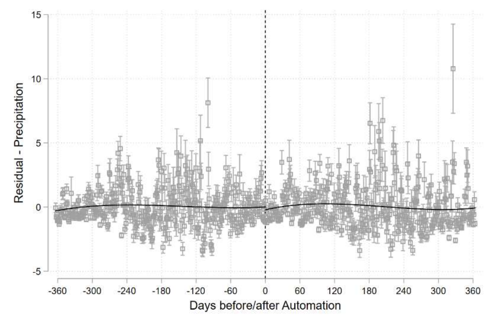
B2. No Discontinuity in Weather Conditions

We conduct additional checks on weather conditions, which lend additional credibility to our findings. 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 (Figure B2 and Table B2), suggesting that the dramatic changes in the air pollution levels across the switching dates were not driven by weather conditions.

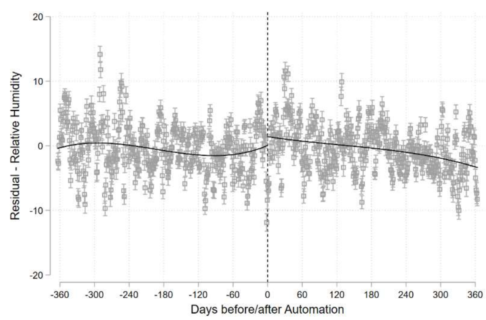
Figure B2. Weather Conditions Before and After the Automation



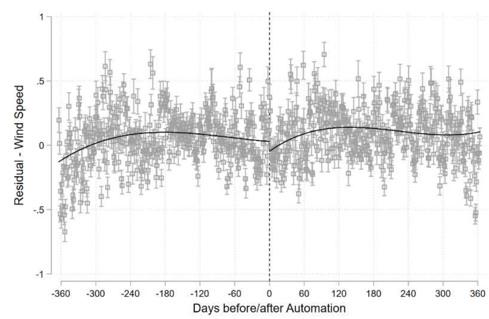
(A). Daily Residual Temperature



(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.

Table B2. Changes in Weather Conditions after Automation

	All Sample			No Missing PM_{10}		
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature	0.92	0.90	0.97	0.55	0.50	0.52
(pre-automation mean =14.56)	(0.65)	(0.65)	(0.66)	(0.77)	(0.77)	(0.78)
Relative Humidity	1.85	2.24*	2.22	2.81*	2.91*	2.00
(pre-automation mean =64.44)	(1.32)	(1.34)	(1.40)	(1.66)	(1.73)	(1.75)
Precipitation	-0.13	-0.13	-0.39*	0.36	0.29	0.23
(pre-automation mean =2.97)	(0.22)	(0.22)	(0.22)	(0.26)	(0.27)	(0.33)
Wind Speed	-0.09	-0.10	-0.11	-0.15*	-0.14*	-0.10
(pre-automation mean =2.41)	(0.06)	(0.06)	(0.07)	(0.08)	(0.08)	(0.09)
Kernel Function	Tri.	Epa.	Uni.	Tri.	Epa.	Uni.
Station FE	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y

Notes: Each cell represents a separate non-parametric RD estimate. The optimal bandwidth is selected by Calonico et al. (2014)'s method. Columns (1) to (3) use all the weather sample. Columns (4) to (6) keep only the sample with PM_{10} data available. Standard errors clustered at the city level are reported in parentheses below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

B3. Additional RD Specifications for the Levels of PM_{10}

We check the sensitivity of the RD estimates using alternative kernel weighting and higher-order global polynomial functions (see Table B3 below). For the local linear RD, using different kernel functions yield similar estimates. The results also remain similar when we use global polynomial RD.

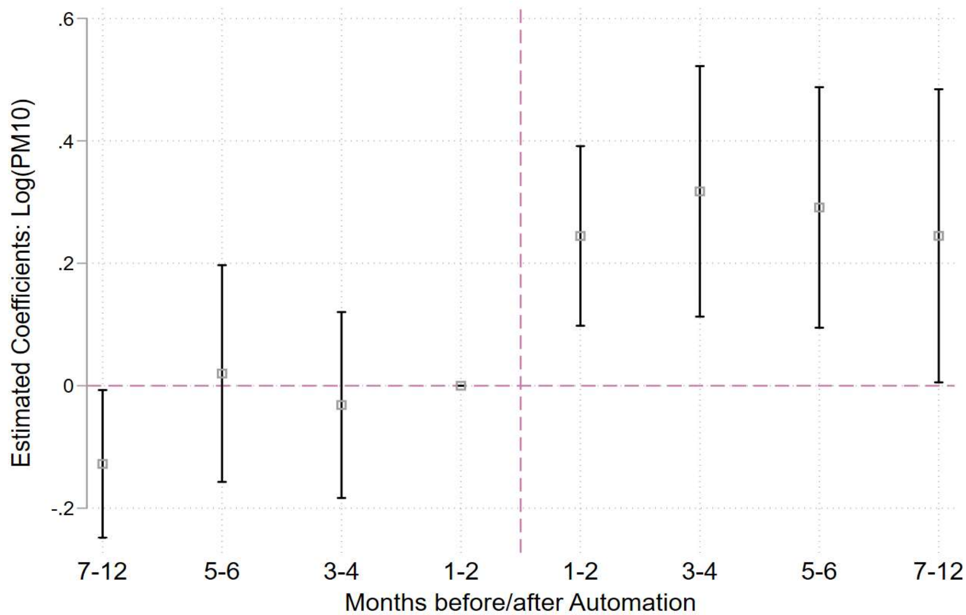
Table B3. RD Estimates Using Alternative Kernel Weightings and Polynomials

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	LLR			Global Polynomial			
Panel A. Station-Day RD							
PM_{10}	34.9***	36.0***	35.7***	32.8***	31.2***	26.6***	31.7***
	(5.8)	(6.4)	(6.6)	(4.1)	(4.4)	(4.6)	(5.3)
Obs. (Daily)	232,326	172,417	131,778	1,049,325	1,049,325	1,049,325	1,049,325
Bandwidth (Days)	263	199	156	All	All	All	All
Panel B. Station-Month RD							
PM_{10}	38.2***	37.6***	35.3***	32.0***	31.1***	24.9***	30.6***
	(5.2)	(5.1)	(5.1)	(4.0)	(4.5)	(5.0)	(5.9)
Obs. (Monthly)	8,389	8,389	8,389	40,964	40,964	40,964	40,964
Bandwidth (Months)	7	7	7	All	All	All	All
AOD	-0.005	-0.007	-0.005	0.036***	0.023**	-0.020	-0.029
	(0.021)	(0.021)	(0.024)	(0.007)	(0.011)	(0.016)	(0.022)
Obs. (Monthly)	5,851	5,851	4,259	26,964	26,964	26,964	26,964
Bandwidth (Months)	7	7	5	All	All	All	All
Station 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	Tri.	Epa.	Uni.	Linear	Quadratic	Cubic	Quartic

Notes: Each cell represents a separate RD estimate. Optimal bandwidth is selected by Calonico et al. (2014)'s method in the non-parametric estimation. Weather controls include temperature, relative humidity, precipitation and wind speed. Standard errors clustered at the city level are reported in parentheses below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

B4. DiD Plots for PM_{10}

Figure B4. Event-Study Estimates of Differences in PM_{10} pre-post Automation



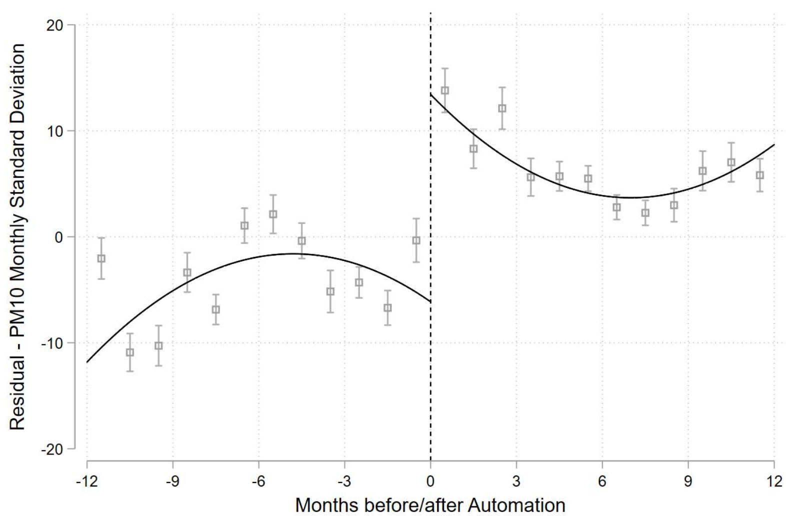
Notes: This figure corresponds to column (5) of Panel B of Table 1, and shows the estimated coefficients for Log(PM_{10}) with 90% CI. The treatment group consists of cities that automated their monitoring stations on January 1st, 2013. The control group consists of cities that automated their monitoring 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.

B5. 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 with the monthly standard deviation of the PM_{10} levels. The monthly standard deviation of PM_{10} is calculated by $\mathrm{SD} = \sqrt[2]{\sum_i^n (\mathrm{P}_{\mathrm{it}} - \overline{\mathrm{P}})^2 / (n-1)}$, where P_{it} is the daily PM_{10} reading at station *i* on day *t*, $\overline{\mathrm{P}}$ is the monthly average, and n is the number of days in a month.

The graphical presentation is illustrated by Figure B5. Like the levels of PM_{10}, we find that automation also significantly increased the variability of the reported PM_{10} concentrations.

Figure B5. RD Plots for PM_{10} Variability



Notes: The discontinuities are plotted using residuals of PM_{10} monthly standard deviations after absorbing station fixed effects, month fixed effects and weather conditions.

Table B5 reports the corresponding estimates. The effect is large in magnitude: when weather and seasonality are controlled, the standard deviation of PM_{10} increased by around 42% after automation (the mean standard deviation before automation is 39.5). This finding adds more evidence on the change in pollution data quality change post-automation.

Table B5. Automating Air Quality Monitoring System and PM_{10} Variability

	All	Wave 1	Wave 2	Deadline
	(1)	(2)	(3)	(4)
Monthly SD in PM_{10}	16.5***	14.5***	27.6***	25.2***
	(2.8)	(4.3)	(5.5)	(4.4)
Station FE	Y	Y	Y	Y
Month FE	Y	Y	Y	Y
Weather Controls	Y	Y	Y	Y
Kernel Function	Tri.	Tri.	Tri.	Tri.
Obs. (monthly)	7,167	4,077	2,811	3,932
Bandwidth (months)	6	5	7	6

Notes: Each cell in the table represents a separate RD estimate from local linear regression. The bandwidth is selected by applying Calonico et al. (2014)'s method to the full sample of 41,920 monthly observations (Column 1) or the relevant subsample. Weather controls include temperature, relative humidity, precipitation and wind speed. Standard errors clustered at the city level are reported in parentheses below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

B6. Results for Other Pollutants

Table B6. Automating Air Quality Monitoring System and Reported Pollutants

	All	Wave 1	Wave 2	Deadline
	(1)	(3)	(4)	(5)
SO_2	1.55	3.25	-0.70	2.40
(ppb)	(2.08)	(2.97)	(2.30)	(3.04)
NO_2	2.98***	3.48***	2.99**	4.68***
(ppb)	(0.87)	(1.11)	(1.37)	(1.28)
Station FE	Y	Y	Y	Y
Month FE	Y	Y	Y	Y
Weather Controls	Y	Y	Y	Y
Kernel Function	Tri.	Tri.	Tri.	Tri.
SO_2 Obs.	160,852	105,030	77,402	91,074
SO_2 Bandwidth	177	169	250	182
NO_2 Obs.	152,685	85,271	89,696	79,334
NO_2 Bandwidth	169	137	284	161

Notes: Each cell in the table represents a separate RD estimate from local linear regression. The bandwidth is selected by applying Calonico et al. (2014)'s method to the full sample of 1,106,783 (1,103,215) daily SO_2 (NO_2) readings or to the relevant subsample. Weather controls include temperature, relative humidity, precipitation and wind speed. Standard errors clustered at the city level are reported in parentheses below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

B7. Changes in Data Collection Requirement

As mentioned in Section II, the automation of air quality monitoring was accompanied by higher standards for data collection. This would make it harder for the local governments to cherry-pick data to report. We address this concern by comparing cities with different degrees of pre-automation missing data issues.

Specifically, we include the samples with the share of missing PM_{10} readings smaller than 10/15/20/25/30% in the year before automation. The thresholds follow the new reporting criteria closely and also allow for additional occasional or random missing PM_{10}. We find robust RD estimates for these samples, suggesting that changes in the data collection standard alone do not mechanically generate the RD estimates. Importantly, it is the automation technology that enables/facilitates the increase in the required reporting frequencies consistently in real time.

Table B7. RD Estimates for Stations with Fewer Pre-Automation Missing PM_{10}

	(1)	(2)	(3)	(4)	(5)
RD in PM_{10}	55.5***	36.6***	29.0***	26.7***	31.4***
	(20.1)	(11.5)	(9.8)	(8.7)	(9.3)
Observations	49,769	227,318	369,125	466,336	512,418
Pre-Missing PM_{10}	≤10%	≤15%	≤20%	≤25%	≤30%
Effective Obs.	13,496	35,368	50,027	60,552	73,220
Bandwidth	278	160	141	136	152
Station FE	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y
Weather Controls	Y	Y	Y	Y	Y

Notes: This table reports the RD estimates for samples with less severe issues in missing PM_{10} readings in the year before automation. Weather controls include temperature, relative humidity, precipitation and wind speed. Standard errors clustered at the city level are reported in parentheses below the estimates. * significant at 10% ** significant at 5% *** significant at 1%.

B8. No Discontinuity at Categorical Cutoffs of PM10

Local officials' incentives to underreport air pollution can be discontinuous, as continuous changes of concentrations within a pollution category may have less payoff than changes at the cutoff to fall into a lower category. Ghanem and Zhang (2014) show that the distribution of the reported PM_{10} over the period 2001–2010 is not well behaved, and that there exists a significant bunching effect around the critical threshold defining the "blue-sky" days (the Air Pollution Index = 100 or the PM_{10} = 150 $\mu g/m^3$).

We examine whether similar bunching patterns can still be observed using post-automation data. Following Cattaneo, Jansson and Ma (2019), we conduct data manipulation tests using local polynomial density estimation at different categorical cutoffs in AQI in Table B8. We find no evidence of bunching at different cutoffs after automation, suggesting the new system significantly limits the room for strategic underreporting.

Table B8. Data Manipulation Tests at Different AQI Thresholds

AQI	PM_{10} ($\mu g/m^3$)	Statistics	(1)	(2)	(3)
50	50	T	0.04	-0.03	0.32
		P-Value	(0.97)	(0.97)	(0.75)
100	150	T	0.39	0.40	0.40
		P-Value	(0.70)	(0.69)	(0.69)
150	250	T	-0.83	-0.86	-0.83
		P-Value	(0.41)	(0.39)	(0.41)
200	350	T	-0.75	-0.85	-0.83
		P-Value	(0.45)	(0.39)	(0.41)
300	420	T	0.84	0.91	0.92
		P-Value	(0.40)	(0.36)	(0.36)
400	500	T	-1.05	-1.06	-1.01
		P-Value	(0.29)	(0.29)	(0.31)
500	600	T	-0.41	-0.46	-0.12
		P-Value	(0.68)	(0.65)	(0.90)
		Kernel	Tri.	Epa.	Uni.

Notes: This table reports the density tests of post-automation PM_{10} distribution at different AQI thresholds using the local-linear density estimation method proposed by Cattaneo, Jansson and Ma (2019). T-statistics of the RD density and corresponding P-values in parentheses are reported.

B9. Correlation between PM_{10} and AOD pre-post Automation

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 a non-manipulated measure. The observation is at the station-month level and we standardized both the PM_{10} and AOD data for this analysis.

Table B9. Partial Correlation between AOD and Reported PM_{10}

	AOD			
	(1)	(2)	(3)	(4)
Panel A. Pre-Automation				
Reported PM_{10}	0.087	0.221	0.225	0.120
Obs.	8,972	8,972	8,972	8,972
Panel B. Post-Automation				
Reported PM_{10}	0.138	0.407	0.389	0.121
Obs.	14,595	14,595	14,595	14,595
Increase in Explanatory Power	59%	85%	73%	1%
Weather Controls		Y	Y	Y
Year-Month FE			Y	Y
Station FE				Y

Notes: Column (1) reports the correlation coefficient between monthly AOD and PM_{10}. Columns (2) to (4) report the partial correlation coefficients after the control variables are partialled out (weather and fixed effects). All correlations are significant at the 0.1% level.

Table B9 summarizes the findings. In column (1), we present the correlations between PM_{10} and AOD. We find that the correlation became stronger after automation, suggesting an improvement in PM_{10} data. In columns (2) and (3), we further include weather controls and time fixed effects. Again, we find that the correlation between PM_{10} and AOD became significantly stronger after automation and the explanatory power increased by over 70% post automation.

Column (4) includes station fixed effects, so this test relies on within-station 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 finds 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.

B10. Correcting the Pre-Automation PM_{10}

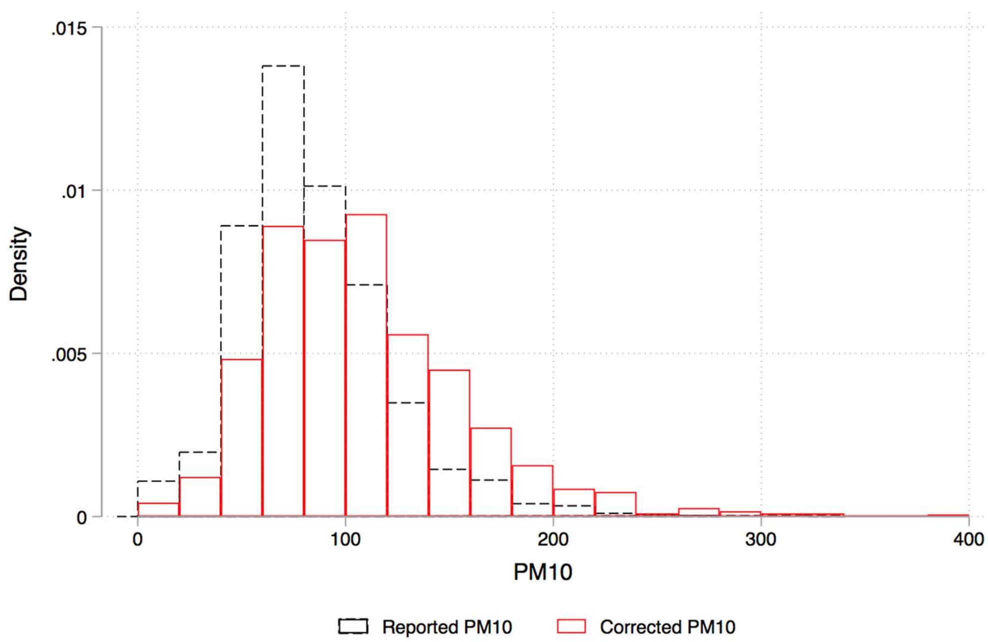
In light of the results in columns (1) to (3) of Table B7, we attempt to correct the pre-automation PM_{10} data by exploiting the relationship between PM_{10}, AOD and weather conditions (temperature, relative humidity, precipitation and wind speed). To increase our predictive power, we use an artificial 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 to train a multi-layered neural network (Doherr, 2018). 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 polynomial functions of AOD and weather conditions aggregated at city level, as well as a rich set of dummies indicating location and month. We use two hidden layers with 20 nodes each, and train the model using a random 70% subset of the post-automation data with 300 iterations.

The trained neural network can explain 81% of the variation in PM_{10} in a held-out test subset of the post-automation sample. As a basis of comparison, this model outperforms polynomial regression models; a regression of PM_{10} on polynomial functions of AOD and weather conditions, conditional on city and month fixed effects, has an R-squared of 0.59 on the same left-out test set. We thus use the trained network to predict PM_{10} concentrations for each pre-automation month in each city.

The correction shifts the distribution of the pre-automation PM_{10} data to the right (see Figure B10 for data-manipulating cities following the second definition in Section V). The mean of PM_{10} in this corrected distribution is 24.4 μg/m^3 or 29% 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.

Figure B10. Correction of Pre-Automation PM_{10} Data

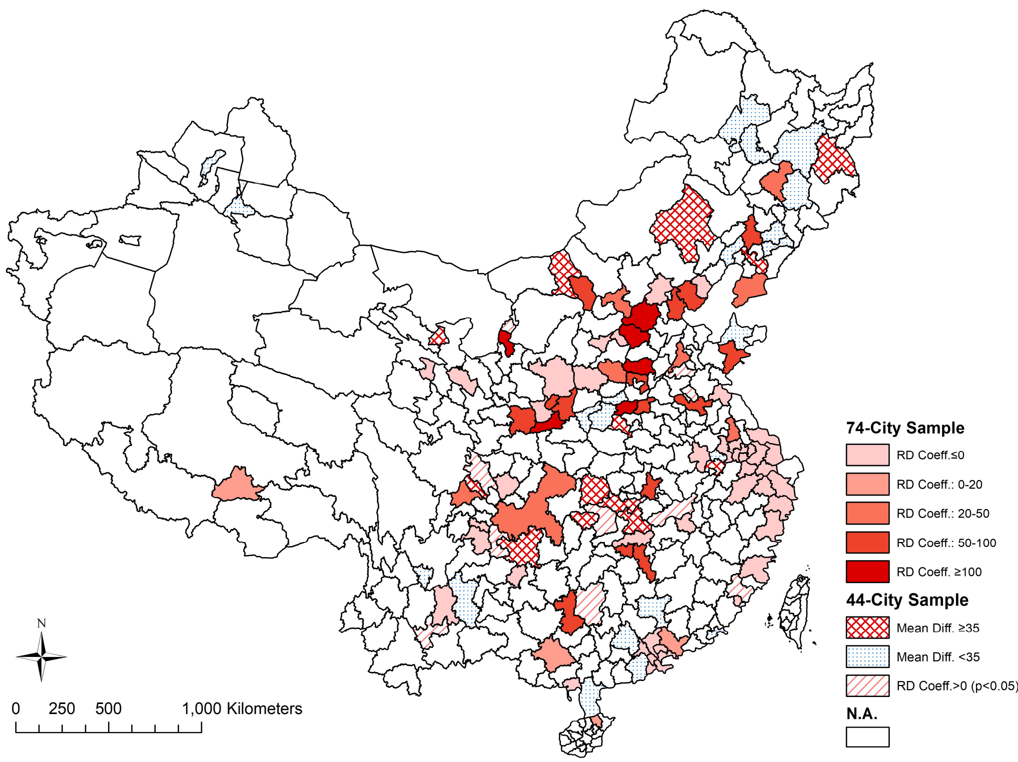


Notes: This figure shows daily PM_{10} before automated monitoring in data-manipulating cities as defined in Panel B of Table 2. The distribution of reported PM_{10} data before automation is plotted in black, and the corrected PM_{10} data using ANN are plotted in red.

C. Additional Results on City-level Variation in Manipulation

C1. Map of Manipulation Status across Chinese Cities

Figure C1. Manipulation Status across Chinese Cities



Notes: The PM_{10} manipulation status in Chinese cities are plotted. For the 74-city sample, manipulation is defined by whether the local linear RD estimate is positive and statistically significant at 5% level. For the other 44 cities with some missing data: (1) if we are unable to obtain a RD estimate using Calonico et al. (2014)'s method, we define a city as a data-manipulating city when the difference in average PM_{10} between January–June 2013 and January–June 2014 is greater than 35 $\mu g/m^3$; if we are still able to obtain a RD estimate using Calonico et al. (2014)'s method, we define a city as a data-manipulating city if the RD estimate is positive and statistically significant at 5% level.

C2. Manipulation and City/Leader Characteristics

Table C2. Correlations b/w Data Quality and City/Leader Characteristics

	Dummy: RD>0 (p<0.05)		Dummy: RD>0 (p<0.05) or Diff $\geq$ 35μg/m³		RD Estimate: Weighted	
	(1)	(2)	(3)	(4)	(5)	(6)
ln(GDP per capita)	-0.07	-0.11*	-0.11***	-0.12***	-6.83**	-6.89
(per 1 SD)	(0.06)	(0.07)	(0.04)	(0.04)	(3.13)	(4.35)
Corrected PM_{10}	0.24***	0.21***	0.18***	0.17***	23.33***	22.75***
(per 1 SD)	(0.06)	(0.07)	(0.05)	(0.05)	(4.17)	(4.15)
Corruption Index	0.12**	0.10*	0.06	0.06	4.13	3.68
(per 1 SD)	(0.05)	(0.06)	(0.05)	(0.05)	(3.70)	(4.37)
Local: PS	-0.09	-0.08	-0.10	-0.11	-8.49	-3.93
	(0.11)	(0.12)	(0.09)	(0.09)	(7.09)	(8.63)
Youth League: PS	0.08	0.10	0.08	0.07	10.38	9.82
	(0.12)	(0.12)	(0.10)	(0.10)	(7.47)	(7.37)
Age: PS	0.01	0.01	0.01	0.01	0.13	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.52)	(0.60)
Science: PS	-0.21**	-0.23**	-0.02	-0.05	-17.31**	-20.39***
	(0.10)	(0.10)	(0.09)	(0.09)	(7.10)	(7.17)
Local: Mayor		0.14		0.12		-8.67
		(0.12)		(0.10)		(9.74)
Youth League: Mayor		0.00		-0.11		-6.78
		(0.12)		(0.11)		(6.78)
Age: Mayor		0.02		0.00		0.31
		(0.02)		(0.01)		(0.94)
Science: Mayor		-0.07		-0.01		-7.51
		(0.13)		(0.10)		(9.46)
Observations	74	74	118	118	74	74
R-squared	0.23	0.28	0.11	0.14	0.38	0.40

Notes: The dependent variables are dummy variables indicating manipulation in columns (1) to (4), and are the regression discontinuity (RD) estimates in columns (5) – (6). In correspondence to Figure 2, columns (1) and (2) define manipulation as cities with RD estimates that are significant at 5% or above. Columns (3) and (4) further include 44 cities that have missing data issues. For these cities, we compare PM_{10} levels between January–June 2013 and January–June 2014: if the reported average increased by 35 μg/m³ in the city, or if the RD coefficient for the city is estimable, positive, and significant, we define it a data manipulating city. In columns (5) and (6), the dependent variable is the city-specific RD estimate weighted by the inverse of the standard error. Leader characteristics include if the Party secretary (PS) or mayor is born in the same province, has experience in the Youth League, has a science or engineering degree, and age. Robust standard errors in parentheses. * significant at 10% ** significant at 5% *** significant at 1%.

D. Association between Online Search and Sales

Table D. Association between Baidu Search Index and Taobao Sales Index

	(1)	(2)	(3)	(4)
	Log (Face Mask Sales Index+1)		Log (Air Filter Face Mask Sales Index +1)	
Log (Search+1)	0.64***	0.31**	0.82**	0.60*
	(0.14)	(0.13)	(0.33)	(0.33)
Observations	467	467	467	467
R-squared	0.86	0.94	0.84	0.88
Weather	Y	Y	Y	Y
City FE	Y	Y	Y	Y
Month FE		Y		Y

Notes: The outcome variables are the log of monthly Taobao Sales Indices for face masks and air filters. The independent variables are the corresponding log of Baidu Search Index. Weather controls include temperature, relative humidity, precipitation and wind speed. Standard errors in parentheses are clustered by city. * significant at 10% ** significant at 5% *** significant at 1%.