Can Technology Solve the Principal-Agent Problem?
Evidence from China's War on Air Pollution

MICHAEL GREENSTONE, GUOJUN HE, RUIXUE JIA, TONG LIU*

We examine the introduction of automatic air pollution monitoring to counter suspected tampering at the local level, 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 an immediate and lasting increase of 35% in *reported* PM_{10} concentrations post–automation. Moreover, automation's introduction increased online searches for face masks and air filters that are strong predictors of purchases. Overall, our findings suggest that the biased and imperfect information prior to automation led to suboptimal investments in defensive measures, plausibly imposing meaningful welfare costs.

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

---

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

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

Ruixue Jia, UC San Diego and London School of Economics and Political Science, rxjia@ucsd.edu

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

Acknowledgment: We thank the editors, Amy Finkelstein and Pete Klenow, and the three anonymous referees for constructive comments. We also thank Christopher 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. We appreciate the financial support from the EPIC-UCAS Joint Center for Energy Policy Research in China.

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 also 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 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 statistical cross-validation 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 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 two key findings. First, there is striking evidence of the underreporting of air pollution concentrations before automation and improvement in data quality after automation. The RD estimates 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, immediately after monitoring was automated.

By several measures, the estimated increase in reported concentrations is large and economically meaningful. It is almost equivalent to China's remarkable improvement in air quality in the first five years of its "war on pollution" (Greenstone et al. 2020). Based on Ebenstein et al. (2017), a permanent 35 $\mu g/m^3$ difference in PM_{10} concentration implies a loss in life expectancy of approximately 2.2 years for an average person. Additionally, the estimated increase in the *reported* PM_{10} concentrations meaningfully exceeds the 2019 annual average PM_{10} *levels* for Los Angeles (20 $\mu g/m^3$) and Boston (10 $\mu g/m^3$).

We also find that the increase in *reported* PM_{10} concentrations post-automation is evident in difference-in-differences (DiD) designs that exploit the variation in the timing of automation across cities, indicating that the higher recorded concentrations were a longer-run phenomenon. Further, city-specific RD estimates produce quantitative measures of the degree of government misconduct for 76 cities. It is rare to have such measures and the variation across cities is striking: underreporting is indicated in 33 cities and 12 of them could have under-reported the PM_{10} levels by more than 75 $\mu g/m^3$.

The second key finding is that the introduction of automated monitoring led to an immediate and lasting increase in online searches for face masks and air filters. Further, the search responses are stronger in cities that are suspected to be "data-manipulating." With respect to the mechanism, it is not possible to isolate whether the changes in internet searches were due to individuals updating their estimates of air pollution concentrations or learning about air pollution more generally due to greater efforts to inform the public about pollution or some combination of both.

The absence of city-level data on purchases from the automation period means that we cannot directly measure automated monitoring's impact on defensive expenditures, but we show using data from a different period that searches of face masks and air filters are strongly correlated with their purchases. Overall, our findings suggest that the biased and imperfect information prior to automation led to suboptimal investments in defensive measures, plausibly imposing 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, Hana, and Ryan 2012; Muralidharan, Niehaus, and Sukhtankar 2016). We provide 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; He, Wang, and Zhang 2020). 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. (2020) also examine the introduction of China's new monitoring system by studying important behavior differences pre-post 2013. The key difference between Barwick et al. (2020) and ours is that we focus on change in information *quality* brought by automated monitoring, while Barwick et al. (2020) treat the introduction of the new system as if pollution information was almost non-existent before automation and its accessibility was largely the same across China. However, pollution data were available in China before 2013, and we complement their paper by demonstrating a wide variation in pollution data quality across cities. We further show that the increase in internet searches was greater in cities with manipulated data, suggesting that biased information prevented people from optimally protecting themselves against pollution. At the same time, Barwick et al. (2020) explore how the new monitoring system affected a wider set of outcomes than this study, most notably the pollution-mortality relationship in China and the pollution-housing price relationship in Beijing.

Third, our study is also related to the literature that assesses the reliability of pollution data from China (e.g., Andrews 2008; Chen et al. 2012; Ghanem and Zhang 2014; and Stoerk 2016). In this literature, researchers find that the distribution of the pre-automation air quality data is not smooth at Air Pollution Index (API) = 100, implying data are manipulated at the politically-important

threshold. One challenge for these studies is that it is unclear whether data manipulation is local and only exists around the threshold. We show that this concern is more general and researchers should be cautious about China's pre-automation air pollution data.

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 quality has increased. As recently as the early 2000s, China only provides readings of the opaque API, rather than concentrations of individual pollutants used to construct the API (i.e., 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 $PM_{2.5}$ 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, Fan, and Zhou 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 China's air quality monitoring network consists of purchasing new monitoring equipment and establishing a new real-time reporting system. New monitors were purchased to monitor $PM_{2.5}$, CO, and O_3, whose information was not available in the past. The equipment and method measuring PM_{10}, SO_2, NO_2 were unchanged, assuring that differences in PM_{10}, SO_2, NO_2 if any, were not due to changes in equipment or measurement method. Instead, the existing equipment was integrated into the new monitoring system. The primary feature of the new system is real-time reporting, which enables online validation and high-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. After automation, opportunities for selective reporting are greatly mitigated as air quality data are sent to the central government in real time. Further, the improvement in surveillance technology allows remote control of data measurement and transmission, as well as quality assurance and 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. Additionally, with the availability of real-time data, a higher standard applies to how to measure pollutants. The minimum requirement for calculating daily, monthly, and yearly PM_{10} are respectively increased from 12 to 20 hours per day, 5 to 27 days per month, and 60 to 324 days per year.

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 cities in three waves, as planned by the central government (see Appendix A1). The variation in timing roughly follows the economic importance of the cities. In the first wave, 74 key polluting cities (496 stations) were required to finish the upgrade by January 1st, 2013. These are major cities in China's Capital Economic Zone, the Yangtze River Delta, the Pearl River Delta, and provincial capitals. In the second wave, another 116 cities (449 stations) were ordered to automate by January 1st, 2014. These cities include the national key environmental cities pre-determined in 2007 and the national model cities for environmental protection (which are evaluated every five years based on their environmental and economic performances since 1997). The third wave further required the remaining 177 cities, which previously did not have systematic air quality monitoring, to build 552 new stations by November 2014.

We focus on 123 cities with 654 monitoring stations that have pollution data both before and after automation. Among these cities, 60 were automated in the first wave and 63 in the second wave. Within a city, all monitors were automated on the same day.

III. Data and Summary Statistics

Our data include measures of monitored air pollution concentrations, weather, aerosol optical depth (AOD), automation dates, and behavioral responses. This section describes these variables and Appendix A2 reports the summary statistics. We also present descriptive evidence that the quality of air pollution data improved after automation in Appendices A3 and A4.

A. Station-Daily Data on Pollutants and Weather

Before automation, the station-level air pollution data were reported on local environmental bureaus' websites and were continuously collected by us. After automation, real-time pollution data can be accessed through China's Real-Time Air Quality Reporting and Analysis Platform.[3] Meteorological data are collected from 403 weather stations, which include daily average temperature, precipitation, relative humidity, and wind speed. We geocoded all the datasets and matched each pollution monitor with its closest weather station.

B. Monthly Data on AOD

AOD data are 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 (10*10 km) 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}.[4] 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), so we use monthly AOD data.[5] We match the ground-level monitors to AOD grid cells based on coordinates.

C. Automation Dates across Cities

We collected news reports on the automation date for each city. As we discussed above, the deadline for automation dates was assigned by the central government; however, cities were allowed to implement the policy before the deadline. Appendix A5 plots the distribution of actual

[3] http://106.37.208.233:20035/

[4] 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.

[5] AOD at the daily level is a much noisier proxy for PM concentration, because clouds can significantly affect its measurement (e.g., too many missing values).

automation dates. Among the 60 (63) Wave-1 (2) cities, 39 (32) automated monitoring at their respective January 1st deadline in 2013 and 2014; in total, 58% of Wave 1 and 2 cities automated their monitoring system at their respective deadline. One concern is that the local governments might strategically choose the automation dates to hide underreporting. In the subsequent analysis, we will conduct a separate analysis that focuses on cities that automate at their deadline to assess the possibility that some cities chose to automate before the deadline to hide manipulation.

D. Behavioral Responses

We measure 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 search index measures city-level search volume for a specific keyword during a specific period. The key advantages of using search indices as the outcomes are that (1) they are available for all the sampled cities both before and after the automation, and (2) they are strong predictors for actual purchasing behaviors. In fact, the internet search advertising business model is built on the idea that searches are a predictor of sales.

IV. Results and Implications

A. Short-Run 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:

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

where P_{ict} indicates the pollution levels reported by station *i* of city *c* at time *t* (daily/monthly). $I(t \geq Auto_{ict})$ is an indicator variable that equals one if station *i* at time *t* is automated. $t - Auto_{ict}$ represents the number of days from the automation and is the running variable. We include a "control function," $\mathrm{f}(t - Auto_{ict})$, and allow it to differ pre and post automation. The station fixed effects (FEs), α_i, account for time-invariant confounders specific to each station. Month FEs, $month_t$, control for seasonality. Weather conditions, $\mathrm{W}_{\mathrm{ict}}$, include temperature, precipitation, relative humidity, and wind speed. u_{ict} is the error term. Since a city can have multiple stations, we cluster our standard errors at the city level.

The parameter of interest, β_1, estimates 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 FEs 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, by running an OLS regression in which the dependent variable is *reported* PM_{10} and the explanatory variables are station FEs, month FEs, and weather controls, and then conduct and RD analysis on the residualized PM_{10}. This procedure provides a consistent estimate of the same RD parameter of interest (Lee and Lemieux 2010).[6]

We start by visualizing the patterns in the data. In Figure 1(A), we plot the raw daily *reported* PM_{10} concentration data. Figure 1(B) plots the residualized concentrations after adjustment for monitoring station FEs, month FEs, and meteorological conditions. In both panels, we observe a striking increase in *reported* PM_{10} immediately after automation. Similar pattern can be observed using monthly PM_{10} data, as shown in Figure 1(C). The estimated increase in *reported* PM_{10} concentrations on the day of automation is 35 μg/m^3 in Figure 1(B) with daily data and 38 μg/m^3 with monthly data in Figure 1(C).

We present RD estimates from equation (1) in Table 1(A). Columns (1) and (2) report the results from the local linear RD with and without covariates (i.e., $\mathrm{W_{ict}}, \alpha_i, month_t$). The bandwidths are 109 days and 263 days for PM_{10} in the two columns, respectively, which are the optimally selected by Calonico, Cattaneo, and Titiunik (2014)'s method. Both columns use a triangle kernel weighting function. The estimated discontinuity is around 35 μg/m^3, corresponding to a 35% increase, relative to the overall post-automation mean (99.5 μg/m^3).

The remaining columns analyze subsets of the monitors. In columns (3) and (4), the *reported* 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 (see Appendix B1 for the RD plots). In column (5), we restrict the sample to the 71 cities (58% of the sample) that implemented the policy at the deadline (drawn from both waves) to explore the possibility that local governments strategically choose their automation dates to hide underreporting. It is at least plausible that cities that automated at their deadline are less likely to have strategically chosen the automation dates to hide manipulation. We find an increase of 57 μg/m^3 in *reported* PM_{10} concentrations for this subgroup, which is larger than in the full sample but

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

qualitatively similar. We conclude that although there may have been some strategic choice of automation dates, the influence of such behavior on the overall results is at most modest.

We provide several sets of results that together lend additional credibility to the baseline findings. First, 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. Second, alternative kernel weighting methods and the parametric RD approach yield similar estimates, as reported in Appendix B3. Finally and most importantly, we fit equation (1) for monthly AOD and find no discontinuity in AOD levels after automation (Figure 1(D), Table 1(A), and Appendix B3), confirming that this measure of true air quality did not deteriorate after the automation.[7]

We then explore the heterogeneity in data quality across different cities, which provides a rare opportunity to learn about the degree of manipulation at the city-level. Specifically, we estimate equation (1) city by city using the non-parametric approach, with the unit of observation being a monitor by day but restricting the effect of automation to be equal across monitors within a city. We note a change in the sample, relative to the 123-city (654-monitor) primary 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, 47 cities did not report PM_{10} readings for more than two months preceding the initiation of automatic monitor reporting and it is therefore challenging to credibly apply the RD approach to these cities individually. We drop these cities for this analysis, leaving a sample of 76 cities (464 monitors).[8]

Figure 2 plots the city-specific RD estimates and their 95% confidence intervals. The RD coefficient is positive for more than 70% of these cities. Among them, 33 estimates are positive and statistically significant at the 5% level. These results suggest that our baseline findings in Table 1(A) are driven by widespread manipulation, rather than a few cities. The average of all the RD coefficients is 28.2 $\mu g/m^3$ (the red line). We observe substantial variation in the estimated discontinuities. Particularly noteworthy findings are that 12 cities have discontinuities greater than 75 $\mu g/m^3$ and 11 of them are statistically significant at the 1% level. Appendix B5 plots the spatial variation in manipulation.

Appendices B6–B10 provide a battery of additional evidence that data quality improved after automation. Specifically, we find that the variability of *reported* PM_{10} became larger post automation (Appendix B6), the discontinuities in *reported* SO_2 and NO_2 (both of which play only small roles in determining promotion for local officials) are small (Appendix B7), the data quality change is not

[7] 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.

[8] Excluding the 47 cities does not affect our baseline RD estimates, as shown in Appendix B4.

simply driven by tighter data collection requirement (Appendix B8), the distribution of post-automation PM_{10} data is smooth at different critical thresholds in the air quality standards (Appendix B9), and the correlation/partial-correlation between *reported* PM_{10} and AOD became stronger after automation (Appendix B10).

B. Medium-run Changes in reported PM_{10}: Difference-in-Differences Estimates

The RD approach offers a demanding test of the effect of automation immediately after its implementation. We complement the main analysis using a set of difference-in-differences models that provide less strict tests but offer the potential to estimate the medium-run effects of automation. Specifically, we conduct event-study analyses using the following equation:

$$P_{ict} = \gamma_\tau \sum_{\tau=-4}^{\tau=+3} Auto_{ic\tau} + \beta W_{ict} + \alpha_i + month_t + \epsilon_{ict} \quad (2)$$

where P_{ict} and W_{ict} are defined in the same way as the RD set-up above. α_i are the station FEs and $month_t$ are the month or year-by-month FEs (we will show results for both specifications). $Auto_{ic\tau}$ 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 to γ_3 allow us to examine whether the automation increases *reported* 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} to γ_{-2} further tell us if the PM_{10} readings months ago were comparable to the baseline readings (PM_{10} readings 1–2 months before automation).

To avoid composition changes in this dynamic analysis, we restrict the sample to cities that automated their monitoring stations only at the deadline of their respective wave and focus on the period from 2012 January 1st, to 2013 December 31st. Thus, the "treatment" monitors are from Wave-1 cities where automation occurred on 2013 January 1st, and the "control" monitors are from Wave-2 cities where automation never occurred during this two-year period. With this set-up, the "control" monitors are never treated and provide a plausibly credible counterfactual for the "treatment" monitors. Further, this is one approach to confronting the challenges associated with the staggered assignments of treatment.[9]

The results are reported in Table 1(B). There is some evidence of increasing *reported* PM_{10} concentrations prior to automation in columns (1) and (2) specifications, which raise questions

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

about the DiD's identifying assumption. To mitigate the influence of any differences in pre-trends, the columns (3)–(5) specifications 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 this paired sample. The idea is that geographically adjacent pairs of monitors should have similar pre-automation trends in PM_{10} readings. Indeed, after matching, the coefficients for all the "lead" variables become small in magnitude and statistically insignificant.

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 *reported* PM_{10} is the dependent variable, 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: the first two months post automation are in the winter season, which is the time of year when pollution concentrations are the highest in China. Figure 3(A) presents the column (5) results graphically.

Overall, we conclude from the previous subsection and this one that automation led to immediate and sustained increases in *reported* PM_{10} concentrations.

C. Correcting Pre-automation PM_{10} Data

As a by-product of this paper, we attempt to correct the pre-automation PM_{10} data using machine learning. The intuition is that we can train prediction models based on the post-automation AOD-PM_{10} relationship and use them to predict pre-automation PM_{10} levels. In Appendix C, we train an artificial neural network to correct the pre-automation PM_{10} data and provide the data as an appendix for other researchers (see Mullainathan and Spiess (2017) for the method). The correction shifts the distribution of the pre-automation PM_{10} data to the right. The mean of PM_{10} in this corrected distribution is 27.3 $\mu g/m^3$ or 32% higher than the mean of the reported pre-automation distribution.

D. Behavioral Responses

This subsection tests for behavioral responses to the improvements in the quality and availability of pollution information. We focus on city-level search indices for "anti-haze face masks" and "air filters" in the period before and after automation. Comparable data on purchases would allow for a revealed preference estimate of the welfare consequences of automation, but such data were unavailable during our study period. However, we present evidence of a strong relationship

between searches and purchases in Appendix D1, albeit from a different period and set of cities.[10] Here, we document that a 10% increase in the city-level anti-haze face mask (air filter) search index corresponds to a 3.1% (6.0%) increase in city-level sales using Taobao sales data.

Turning to the search data, Figures 1(E) and 1(F) present the RD plots for the two outcomes, respectively. Table 2(A) summarizes the RD estimates using equation (1). We find that monthly online searches for "face masks" immediately tripled (columns (1)–(2)) after automation and searches for "air filters" increased by 17–20%.

We then examine the differences in people's responses between data-manipulating cities and normal cities. For the 76 cities with RD estimates, we define a city as a data-manipulating one if its RD estimate is positive and statistically significant at 5% level. For the remaining cities that have missing data issues, we compare *reported* PM_{10} levels between January–June 2013 and January–June 2014: if the *reported* average PM_{10} concentrations increased by 35 $\mu g/m^3$ in the city, we define it a data manipulating city (13 cities meet this criterion).[11] Column (3) reports the results using data from the "normal" cities and column (4) focuses on "data-manipulating" cities. In "data-manipulating" cities, the post-automation increase in searches was substantially larger: more than four times for "face masks" and 28% higher for "air filters." While it is tempting to interpret these larger estimates in column (4) as being entirely due to individuals' learning that actual PM_{10} concentrations were higher than they had believed, it is also possible that the increase in news about air pollution disproportionately increased in these cities at the same time.

Table 2(B), Figures 3(B) and 3(C) report the event-study estimates using equation (2). We find that the higher rates of searches for these two terms were sustained and still event 7-12 months after automation.[12] It is reassuring that the RD and event-study estimates are similar in size. These results imply automation leads to an immediate and sustained increase in the purchases of goods that protect individuals from PM_{10}.

Because most cities adopted the new technology just on two dates (January 1st of 2013 and 2014), a legitimate concern is that there might be other policies launched at the beginning of a new year that could drive the behavioral changes. This concern is mitigated by the fact that the event-study estimates persist. We investigate this issue in two additional ways. First, we conduct similar analyses around January 1, 2015 in Appendix D2 and find no significant behavioral changes.

[10] The Taobao sales data are available in 34 Wave-1 cities (4 municipalities and 27 provincial capitals) from April 2013 to April 2014.

[11] We use 35 $\mu g/m^3$ as the threshold to define the data-manipulating status for these cities because 35 $\mu g/m^3$ is the average discontinuity in Table 1. In reality, it is almost impossible for a city to reduce its PM_{10} concentration by 35 $\mu g/m^3$ within a year. Slightly change the definition of data-manipulating city, for example, using a more/less stringent definition, does not affect our results in Table 2.

[12] The searches for anti-haze masks were very low before automation, since anti-haze became an issue only after automation.

Second, we report the RD estimates separately for cities that automated their systems at their deadline and cities that automated their systems before their deadline in Appendix D3. We find that the behavioral responses can be observed in both groups, even though the effect size is greater in the deadline cities.[13]

We do not believe that it is possible to neatly 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 seems reasonable to conclude that the biased and imperfect information about air pollution imposed meaningful welfare costs prior to automation.

V. 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 offer 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, the more reliable information post automation seems to have induced more people to take avoidance behaviors against pollution, which implies welfare gains that are of a potentially significant magnitude. That being 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.

[13] A limitation of this comparison is that pollution levels are higher in winter and the deadline is January 1st, which is in the midst of winter. Consequently, it is possible that the larger response in internet searches in deadline-complying cities reflects greater awareness of pollution and responsiveness to information when concentrations are the highest.

References

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

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

Barwick, Panle Jia, Shanjun Li, Liguo Lin, and Eric Zou, 2020. 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 non-parametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6), pp.2295-2326.

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.

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.

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. China's war on pollution: Evidence from the first five years. Working Paper.

He, G., Wang, S., & Zhang, B. (2020). Watering down environmental regulation in China. *The Quarterly Journal of Economics*.

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.

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.)

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.

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.

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*

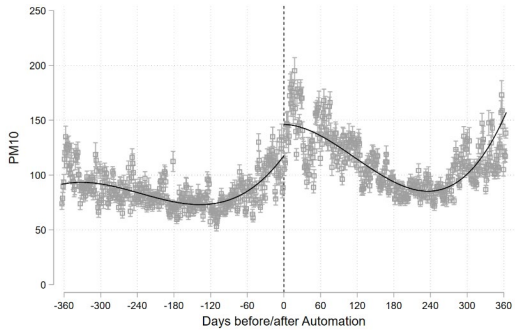
Review of Resource Economics, 6(1), pp.339-360.

Stoerk, T. 2016. Statistical corruption in Beijing's air quality data has likely ended in 2012. *Atmospheric Environment*, 127, 365-371.

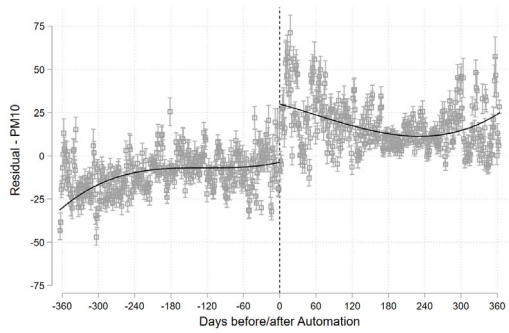
Williamson, O.E., 1996. *The mechanisms of governance.* Oxford University Press.

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

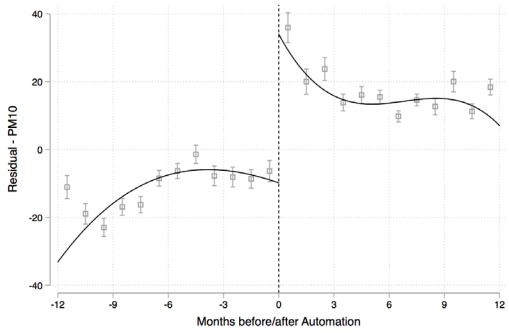
Figure 1. RD Plots for Reported PM_{10}, AOD, and Online Searches



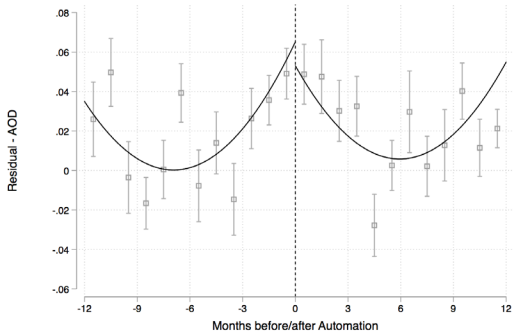
(A). Daily Raw PM_{10}



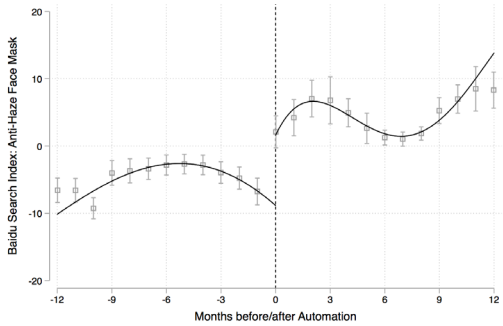
(B). Daily Residual PM_{10}



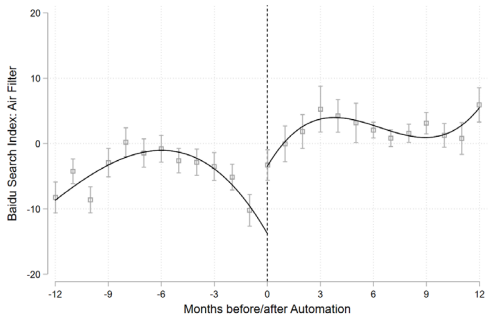
(C). Monthly Residual PM_{10}



(D). Monthly Residual AOD



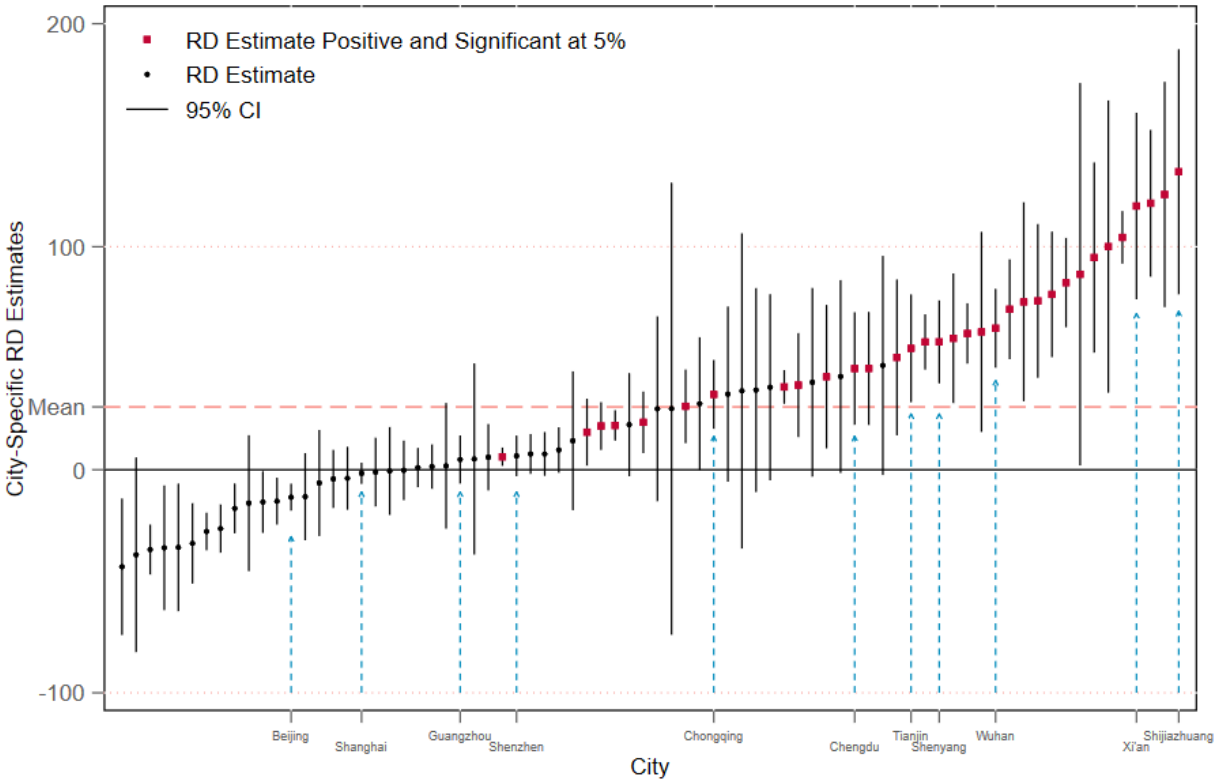
(E). Monthly Residual Face Mask Search



(F). Monthly Residual Air Filter Search

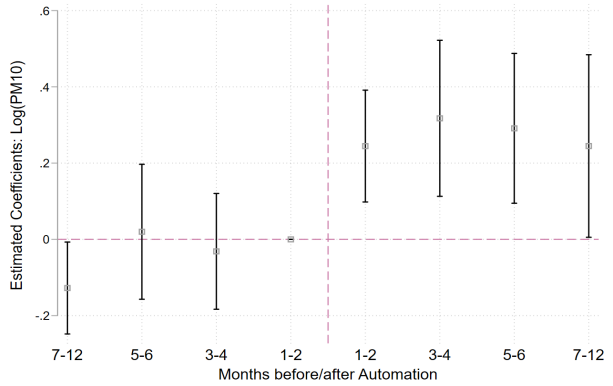
Notes: Panels (A) – (C) show the increase in reported PM_{10} immediately after automation using raw daily data, daily residuals and monthly residuals (absorbing station, month FEs and weather), respectively. Panel (D) shows no significant change in monthly AOD data. Panels (E) and (F) show the increase in monthly online searches for anti-haze face masks and air filters after absorbing city FEs, month FEs, and weather conditions.

Figure 2. Magnitudes of Manipulation in Chinese Cities

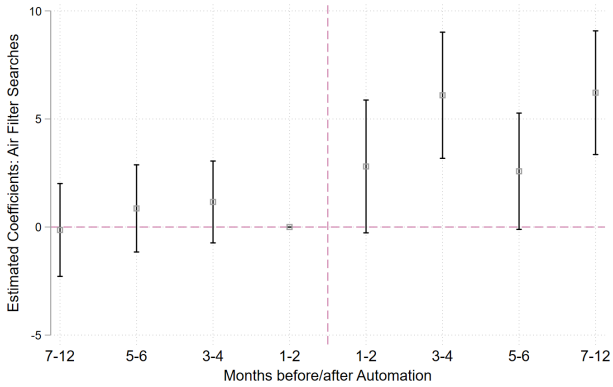


Notes: The RD estimates and their 95% confidence intervals are plotted for 76 cities. The average RD estimate is denoted by the red dashed line.

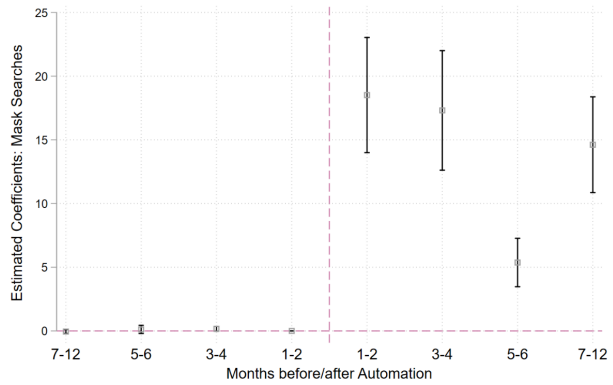
Figure 3. Event-Study Estimates for Reported PM_{10} and Online Searches



(A). Reported PM_{10}



(B). Air Filter Search Index



(C). Anti-Haze Face Mask Search Index

Notes: This figure shows the event-study estimates and their 90% confidence intervals for reported PM_{10} and online search indices. Panel (A) corresponds to the regression results reported in column (5) of Panel B in Table 1. Panels (B) and (C) correspond to the regression results reported in columns (2) and (4) of Panel B in Table 2. In Panels (B) and (C), the treatment group consists of cities that automated the system on January 1st, 2013, and the control group consists of cities that automated the system on January 1st, 2014. Panel (A) matches the treatment stations with the nearest control stations. We keep data from January 1st 2012 to December 31st, 2013 for the estimations and omit the "1–2 months before automation" period in the event-study regressions. Location FEs, month FEs, and weather conditions are controlled.

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

	(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, Cattaneo, and Titiunik (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, we use data from January 1st 2012 to December 31st 2013 for the event-study estimates. The omitted period is "1-2 months before automation." 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.

Table 2. 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	6.31	18.77
(pre-automation mean =0.62)	(1.58)	(1.66)	(2.12)	(2.63)
RD in Air Filter Searches	7.36	8.73	5.48	16.30
(pre-automation mean =35.5)	(3.60)	(1.86)	(2.64)	(2.39)
RD in Log (Face Mask Searches+1)	1.06	1.15	0.86	1.72
(pre-automation mean =0.16)	(0.17)	(0.17)	(0.21)	(0.25)
RD in Log (Air Filter Searches+1)	0.18	0.16	0.13	0.25
(pre-automation mean =3.30)	(0.10)	(0.04)	(0.05)	(0.06)
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)	(2.86)	(1.78)	(1.78)
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 and optimal bandwidth is selected by Calonico, Cattaneo, and Titiunik (2014)'s method. 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 defined in Section IV. In Panel B, we use data from January 1st 2012 to December 31st 2013 for the event-study estimates. The omitted period is "1-2 months before automation." The 39 Wave-1 deadline cities are the treatment group and the 32 Wave-2 deadline cities are the 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.

Online Appendix

Can Technology Solve the Principal-Agent Problem?

Evidence from China's War on Air Pollution

MICHAEL GREENSTONE, GUOJUN HE, RUIXUE JIA, TONG LIU

A. Background and Data

- **A1. Different Automation Waves**
- **A2. Summary Statistics**
- **A3. Descriptive Patterns in the Yearly Data**
- **A4. City-Level Cases**
- **A5. Distribution of Automation Dates**

B. Additional Results on Data Quality pre-post Automation

- **B1. RD Using Raw Reported Daily and Monthly PM_{10}**
- **B2. No Discontinuity in Weather Conditions**
- **B3. Additional RD Specifications for the Levels of PM_{10}**
- **B4. Dropping 47 Cities that Have Missing Data Problem**
- **B5. Map of Manipulation Status in Chinese Cities**
- **B6. Variability in Reported PM_{10}**
- **B7. Results for Other Pollutants**
- **B8. Changes in Data Collection Requirement**
- **B9. No Bunching Effect of Reported PM_{10} Post Automation**
- **B10. Correlation between Reported PM_{10} and AOD pre-post Automation**

C. Correcting the Pre-Automation Reported PM_{10}

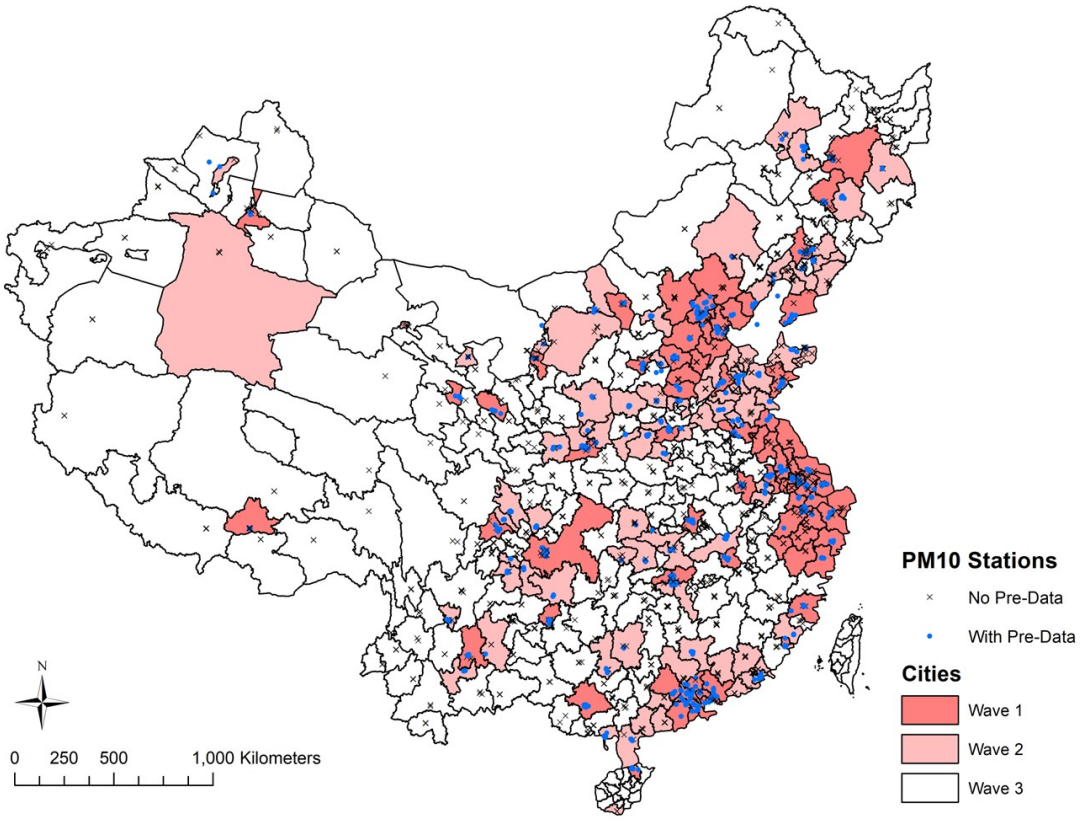
D. Additional Results on Online Searches

- **D1. Association between Online Searches and Sales**
- **D2. RD Plots for Online Searches around January 1st 2015**
- **D3. Automation and Online Search in Deadline and Non-Deadline Cities**

A. Background and Data

A1. Different Automation Waves

Figure A1. Different Automation Waves



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. Summary Statistics

Table A2. Summary Statistics

	Mean and Std. Dev.					
	2011	2012	2013	2014	2015	2016
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Pollution and AOD						
Reported 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)
Reported 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)
Reported 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.

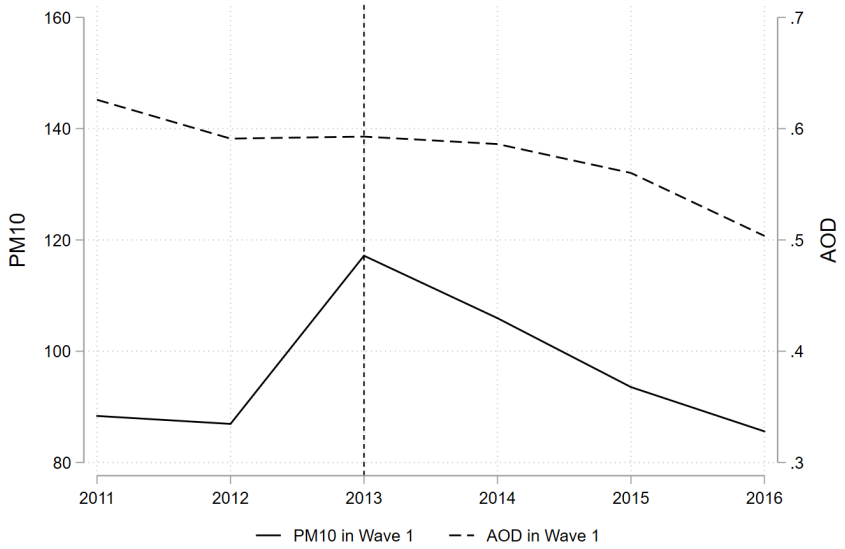
A3. Descriptive Patterns in the Yearly Data

Figure A3 plots the yearly *reported* PM_{10} levels. 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. 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.

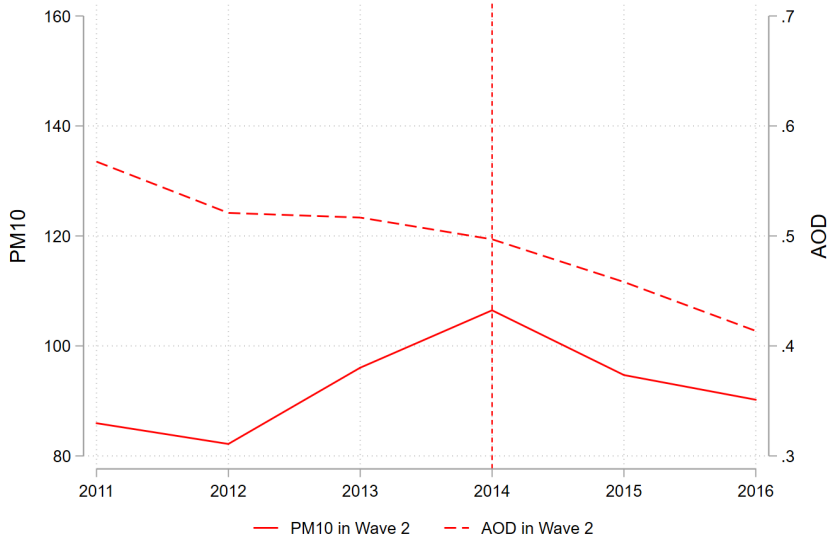
For cities in the first wave, for example, *reported* annual PM_{10} levels increased by more than 30 $\mu g/m^3$ from 2012 to 2013, which was about the same magnitude as the total improvement in PM_{10} reduction in the following four years (see Appendix A3 for the summary statistics of key variables).

Importantly, we observe that the trends in AOD and PM_{10} for both waves of cities became similar after automation. This result suggests the automation improved the pollution data quality.

Figure A3. Annual Reported PM_{10} and AOD from 2011 to 2016



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



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

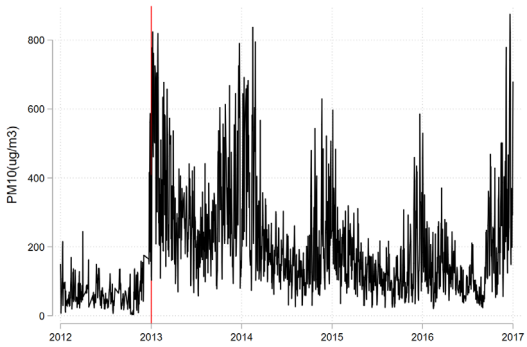
Notes: Annual average reported PM_{10} concentrations ($\mu g/m^3$) in Wave-1 and Wave-2 cities are plotted in black and red, respectively. Corresponding AOD levels are shown in dashed lines.

A4. City-Level Cases

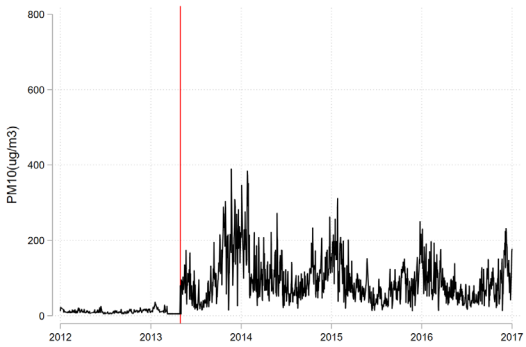
This subsection takes an admittedly selective examination of the reported time series from four stations as a means of highlighting the high geographic and temporal variation of the data and qualitatively previewing the finding of extensive manipulation in some locations before automation. For instance, in the monitoring station in the development zone of Shijiazhuang city (the upper left panel of Figure A4), 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 A4), the *reported* 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 *reported* 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 A4), the *reported* 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.

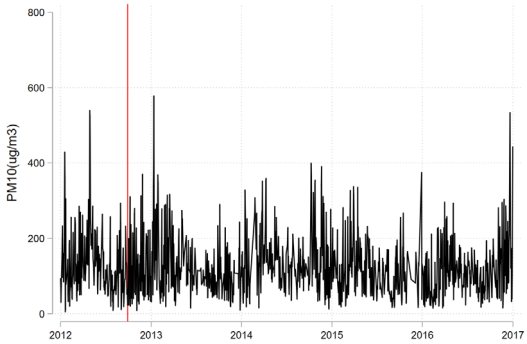
Figure A4. Times Series of Reported PM_{10} Concentrations at Four Stations



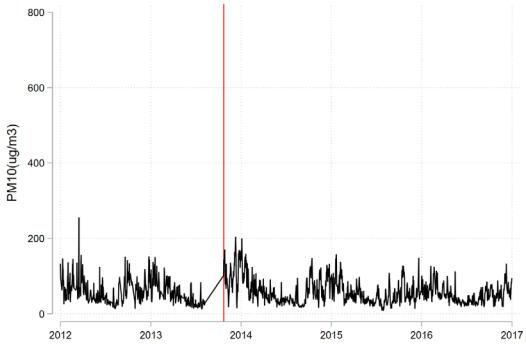
(A). Gaoxin District, Shijiazhuang City, Hebei



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



(C). Gucheng, Beijing

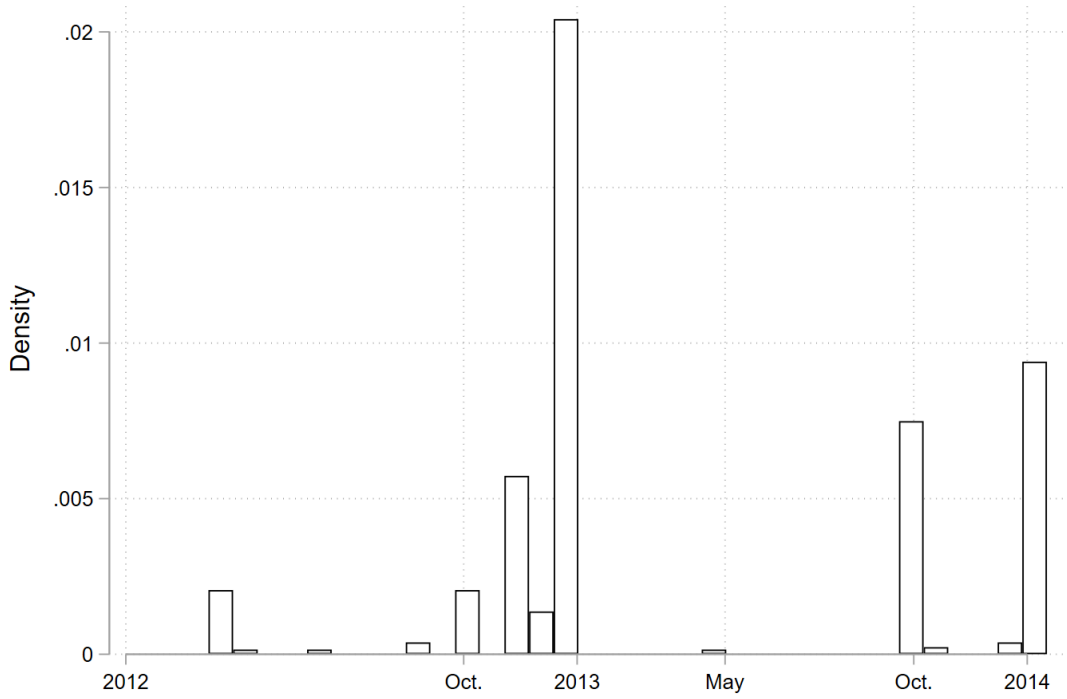


(D). Industrial Park, Beihai City, Guangxi

Notes: The time series of reported 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.

A5. Distribution of Actual Automation Dates

Figure A5. Distribution of Actual Automation Dates

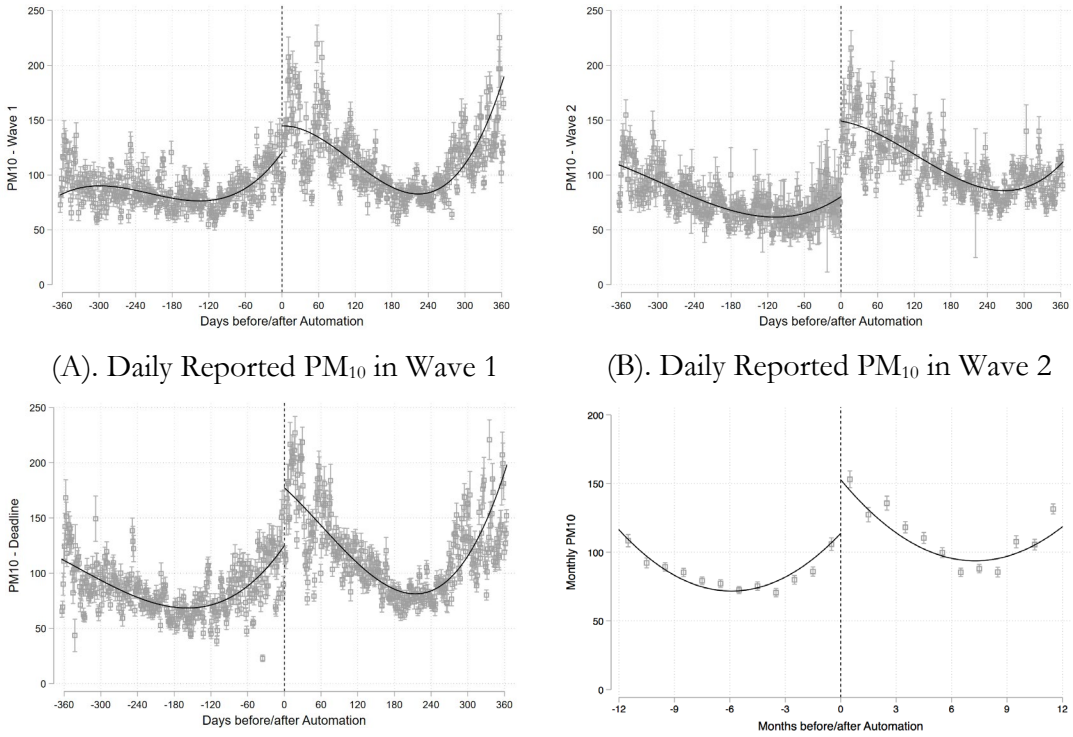


Notes: This figure summarizes the distribution of the automation dates across different cities. The majority of them automated the air quality monitoring stations on January 1st, 2013 and January 1st, 2014, which are the deadlines for the two waves.

B. Additional Results on Data Quality pre-post Automation

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

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



(A). Daily Reported PM_{10} in Wave 1

(B). Daily Reported PM_{10} in Wave 2

(C). Daily Reported PM_{10} in Deadline

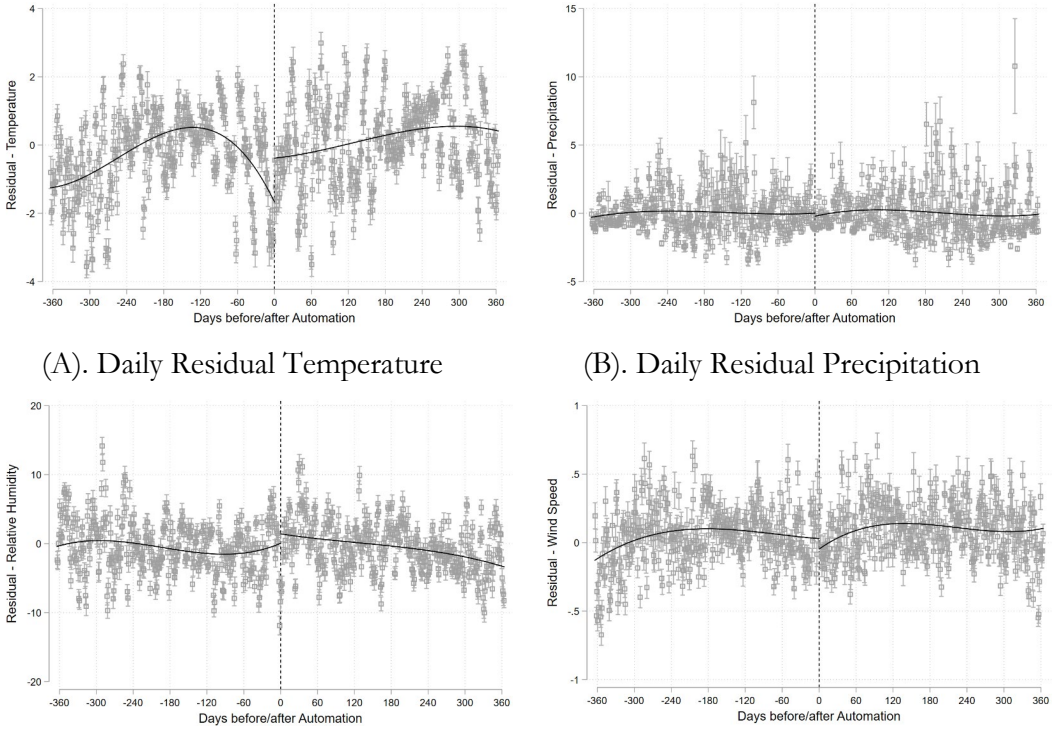
(D). Monthly Reported PM_{10}

Notes: In Panels (A)–(C), the discontinuities are plotted using raw reported daily PM_{10} concentrations (no controls are included). In Panel (D), the discontinuity is plotted using station-by-month reported PM_{10} data.

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 FEs and month FEs 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 in which PM_{10} data are available. Standard errors clustered at the city level are reported in parentheses below the estimates.

B3. Additional RD Specifications for the Levels of Reported 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.

B4. Dropping 47 Cities that Have Missing Data Problem

Out of the 123-city sample, 47 cities suspended PM_{10} reporting for more than two months when installing and testing the new monitors prior to automatic reporting. Our results remain largely unchanged after we drop these cities and keep only the other 76 cities (464 monitors).

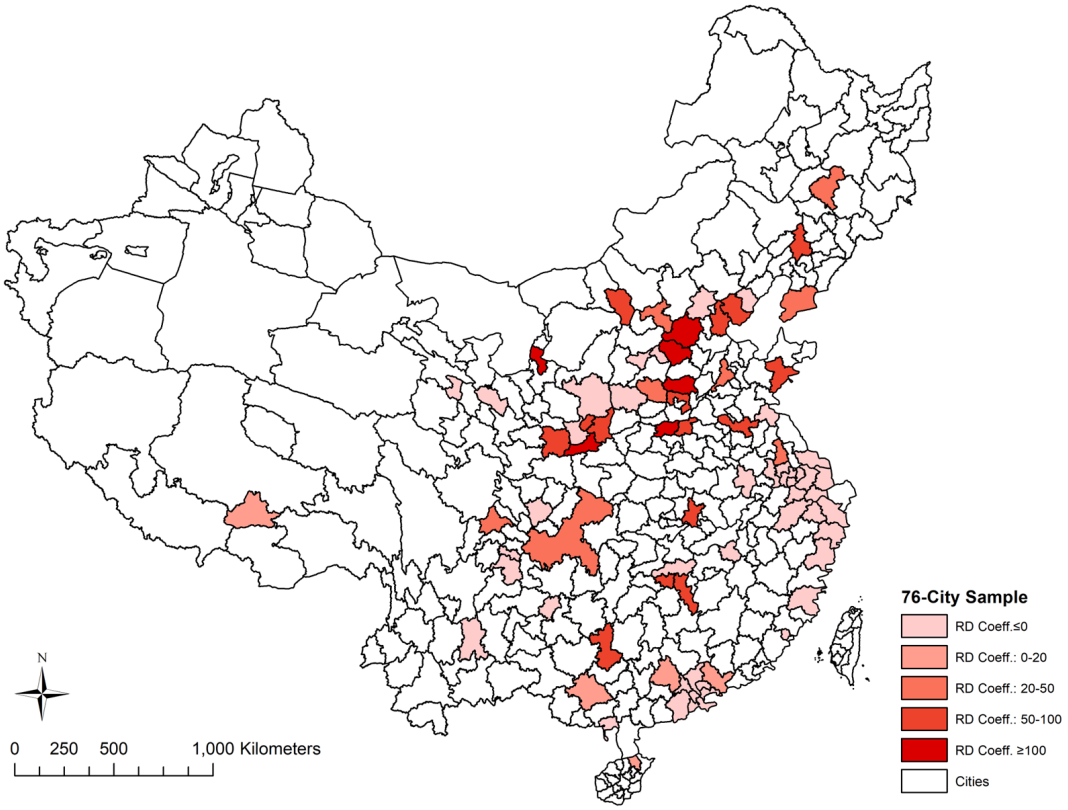
Table B4. Automation and Reported PM_{10} in 76 Cities

	(1)	(2)	(3)	(4)	(5)
RD in PM_{10} (Daily)	32.2	33.6	28.6	71.3	64.3
	(12.9)	(8.9)	(10.1)	(15.2)	(12.1)
Sample	All	All	Wave 1	Wave 2	Deadline
Station or City FE	N	Y	Y	Y	Y
Month FE	N	Y	Y	Y	Y
Weather Controls	N	Y	Y	Y	Y
Obs. (Daily)	77,143	116,867	83,003	16,632	71,130
Bandwidth (Days)	113	172	144	161	207

Notes: Each cell represents a separate non-parametric RD estimate. We focus on the 76-city sample (464 monitors), which does not have missing-data problem. 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. Columns (3) and (4) use the Wave-1 and Wave-2 cities to estimate the discontinuities. Column (5) uses cities that automated the monitoring system at their deadlines to estimate the discontinuities. Weather controls include temperature, relative humidity, precipitation and wind speed. Standard errors clustered at the city level are reported below the estimates.

B5. Map of Manipulation Status in Chinese Cities

Figure B5. Map of Manipulation Status in Chinese Cities



Notes: The map describes the geographical locations for 76 Chinese cities (with city-specific RD estimates). In the paper, manipulation is defined by whether the local linear RD estimate is positive and statistically significant at 5% level.

B6. Variability in Reported PM_{10}

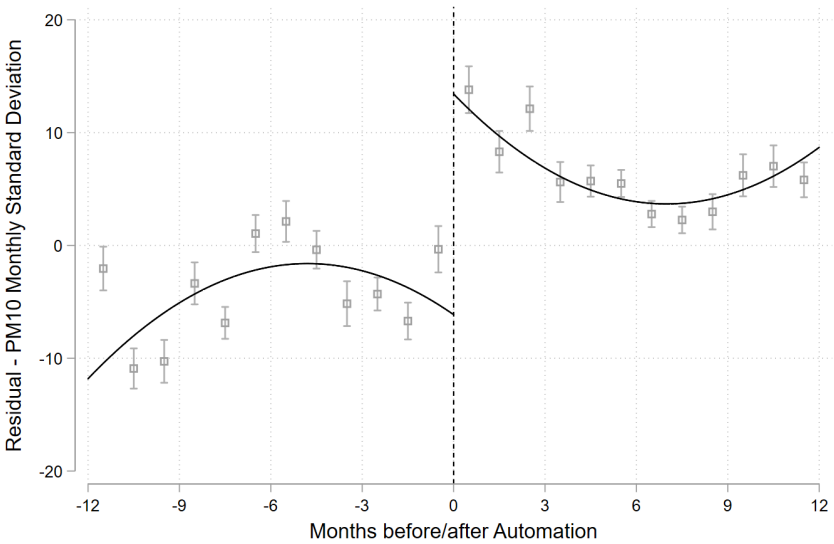
As another measure of data quality, we examine the variability of *reported* PM_{10} under the presumption that manipulated measures are likely to exhibit less variability than true realizations. We fit equations (1) and (2) by replacing the outcome variable with the monthly standard deviation of the *reported* PM_{10} levels, which is calculated the following equation:

$$\mathrm{SD} = \sqrt[2]{\sum_{i}^{n} (\mathrm{P_{it}} - \overline{\mathrm{P}})^2 / (n-1)}$$

where $\mathrm{P_{it}}$ is the *reported* 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 B6. We find that automation also significantly increased the variability of the *reported* PM_{10} concentrations.

Figure B6. RD Plots for PM_{10} Variability



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

Table B6 reports the corresponding estimates. The effect is large in magnitude: when weather and seasonality are controlled, the standard deviation of *reported* PM_{10} increased by around 42% after automation (the mean standard deviation before automation is 39.5).

This finding adds additional evidence on the change in pollution data quality post-automation.

Table B6. Automating Monitoring System and Reported PM_{10} Variability

	All	Wave 1	Wave 2	Deadline
	(1)	(2)	(3)	(4)
Monthly SD in Reported 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.

B7. Results for Other Pollutants

Table B7. Automating Monitoring System and Reported SO_2 and NO_2

	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.

B8. 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 local governments to cherry-pick data to report. One concern is that the increase in reported PM_{10} post-automation may be simply driven by the higher reporting standards. We address this concern by comparing cities with different degrees of pre-automation missing data issues.

Specifically, we apply the RD method to different subsamples with varying degrees of missing data and examine whether the increase in *reported* PM_{10} levels is larger among cities with more server missing-data issue. Table B8 reports the results. We find the discontinuity exists in all the sub-samples, suggesting that changes in the data reporting standards alone do not mechanically generate the RD estimates.

Table B8. 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.

B9. No Bunching Effect of Reported PM_{10} Post Automation

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 pollution 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 μg/m^3).

We examine whether similar bunching patterns can 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 B9. We find no evidence of bunching at different cutoffs after automation, suggesting the new system significantly limits the room for strategic underreporting.

Table B9. Data Manipulation Tests at Different AQI Thresholds

AQI	PM_{10} (μg/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.

B10. Correlation between Reported 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 B10. 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 FEs). All correlations are significant at the 0.1% level.

Table B10 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 FEs. 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 FEs, 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, Rubin, and Walker (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. 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.

C. Correcting the Pre-Automation PM_{10}

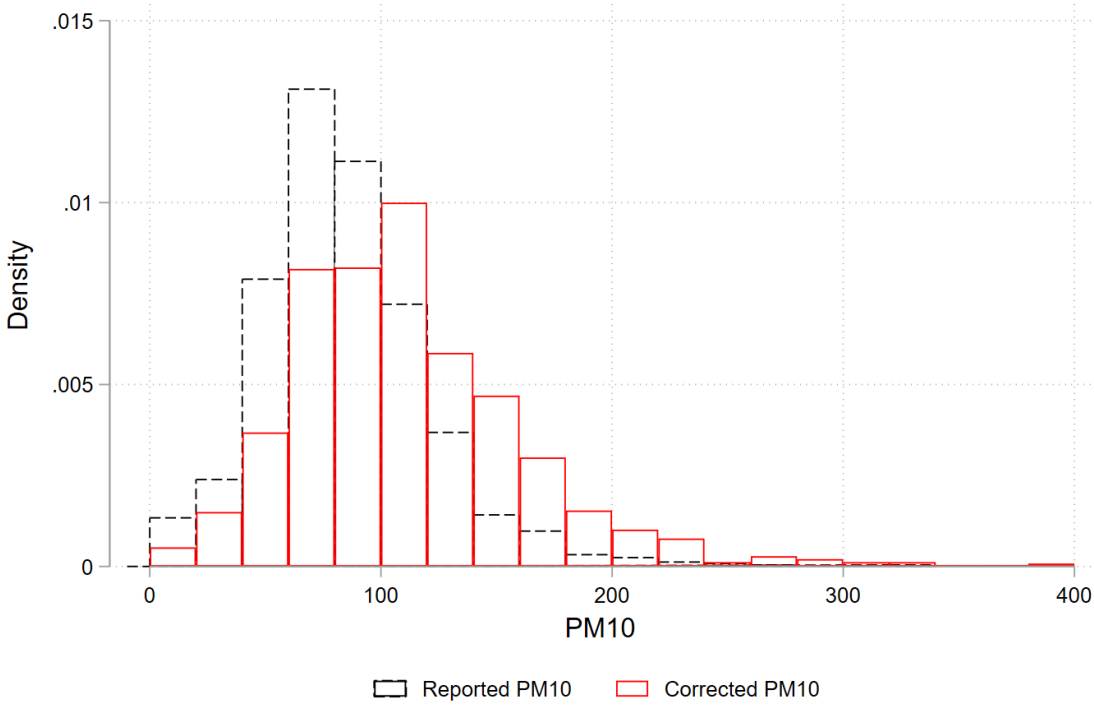
In light of the results in columns (1) to (3) of Table B10, we attempt to correct the pre-automation PM_{10} data by exploiting the relationship between PM_{10}, AOD and weather conditions (temperature, relative humidity, precipitation and wind speed). To increase our predictive power, we use an artificial 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 FEs, 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 (the definition of data-manipulating cities is discussed in Section IV of the paper). The mean of PM_{10} in this corrected distribution is 27.3 $\mu g/m^3$ or 32% 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 C. Correction of Pre-Automation PM_{10} Data



Notes: The distribution of reported PM_{10} data in the data-manipulating cities (defined in Section IV of the paper) before automation is plotted in black, and the corrected PM_{10} data using ANN are plotted in red.

D. Additional Results on Online Searches

D1. Association between Online Searches and Sales

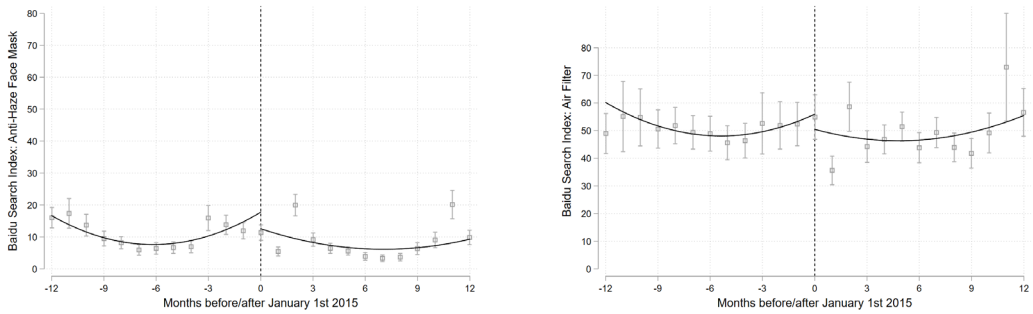
Table D1. Association between Baidu Search Index and Taobao Sales Index

	(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 sales data are available for 34 Wave-1 cities from April 2013 to April 2014. 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.

D2. RD Plots for Online Searches around January 1st 2015

Figure D2. RD Plots for Online Searches: January 1st 2015



(A). Anti-Haze Face Mask (B). Air Filter

Notes: This figure plots the discontinuities in search indices of anti-haze face masks (A) and air filters (B) before and after 2015 January. We use the 123-city sample to plot this figure.

D3. Automation and Online Search in Deadline and Non-Deadline Cities

Table D3. Automation and Online Searches in Deadline/Non-Deadline Cities

	Deadline Cities		Non-Deadline Cities	
	(1)	(2)	(3)	(4)
RD in Face Mask Searches	18.72	18.60	4.35	4.31
	(2.03)	(2.06)	(1.86)	(1.86)
RD in Air Filter Searches	20.78	21.18	-2.11	-2.18
	(1.99)	(2.04)	(3.33)	(3.33)
RD in Log (Face Mask Searches+1)	1.67	1.63	0.78	0.87
	(0.21)	(0.23)	(0.21)	(0.20)
RD in Log (Air Filter Searches+1)	0.21	0.20	0.11	0.12
	(0.05)	(0.05)	(0.05)	(0.05)
City FE	Y	Y	Y	Y
Month FE	Y	Y	Y	Y
Weather Controls		Y		Y

Notes: This table reports the RD estimates of online searches for stations that are automated on the deadline dates and the non-deadline dates, respectively. Weather controls include temperature, relative humidity, precipitation and wind speed. Standard errors clustered at the city level are reported below the estimates.

References

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

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

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.

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