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}

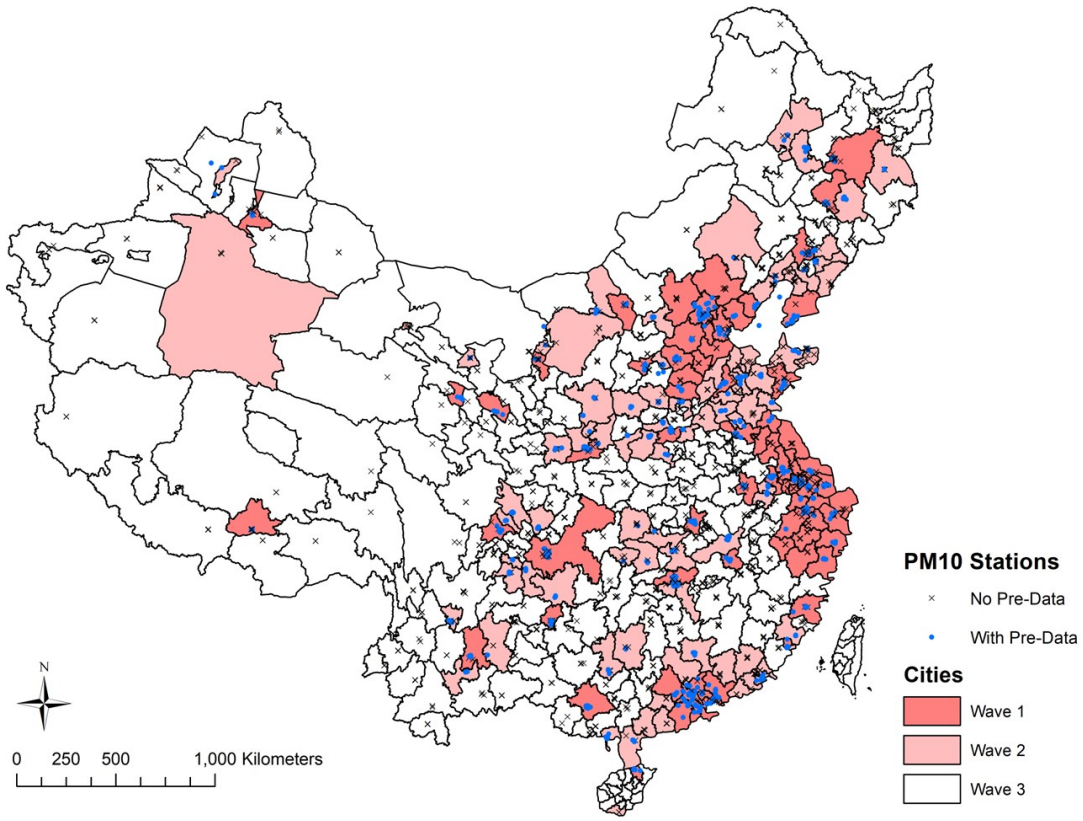
D. Additional Results on Online Searches

- **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.

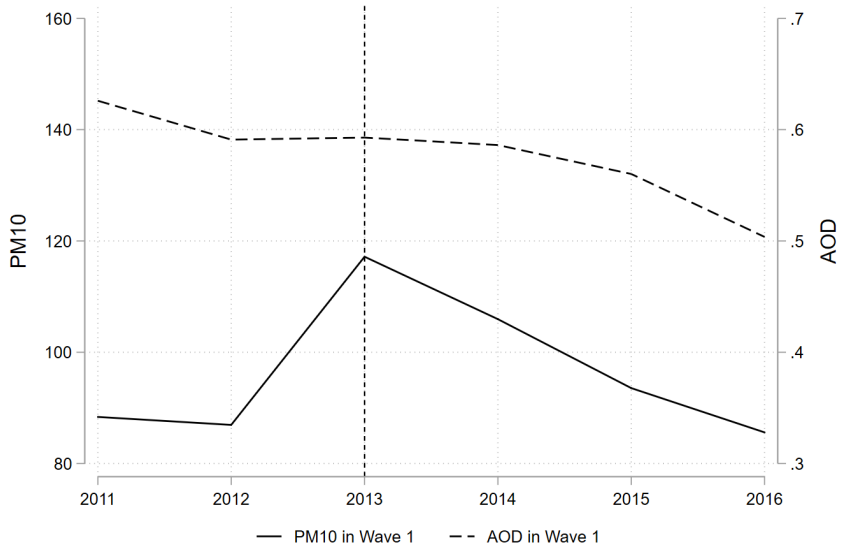
A3. Descriptive Patterns in the Yearly Data

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.

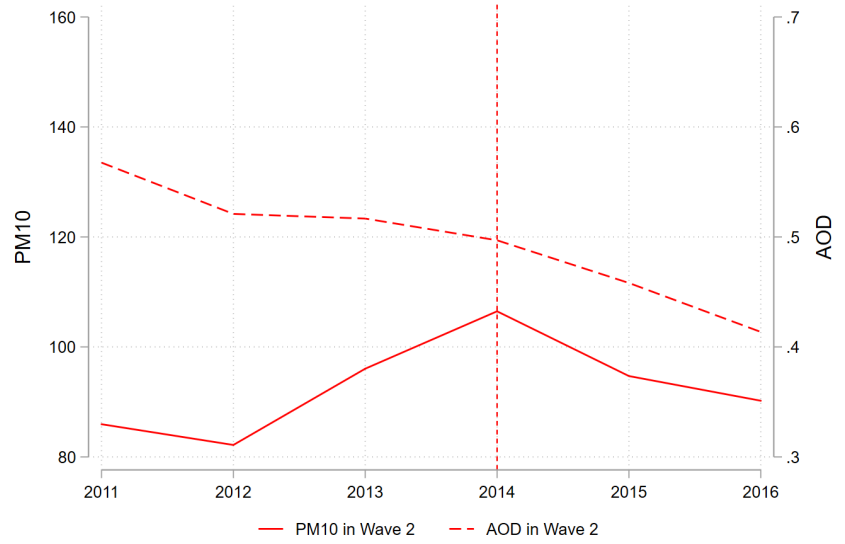
For cities in the first wave, for example, *reported* annual PM_{10} levels increased by more than 30 $\mu g/m^3$ from 2012 to 2013, which was about the same magnitude as the total improvement in PM_{10} reduction in the following four years (see Appendix A3 for the summary statistics of key variables).

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

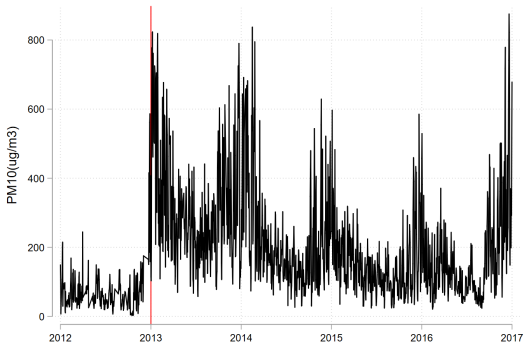
Notes: Annual average reported PM_{10} concentrations ($\mu g/m^3$) in Wave-1 and Wave-2 cities are plotted in black and red, respectively. Corresponding AOD levels are shown in dashed lines.

A4. City-Level Cases

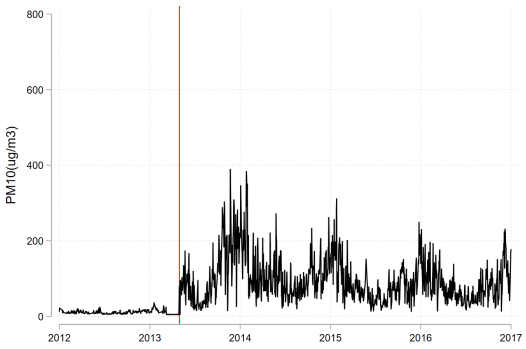
This subsection takes an admittedly selective examination of the reported time series from four stations as a means of highlighting the high geographic and temporal variation of the data and qualitatively previewing the finding of extensive manipulation in some locations before automation. For instance, 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.

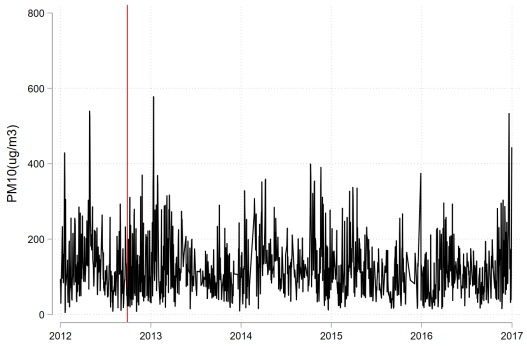
Figure A4. Times Series of Reported PM_{10} Concentrations at Four Stations



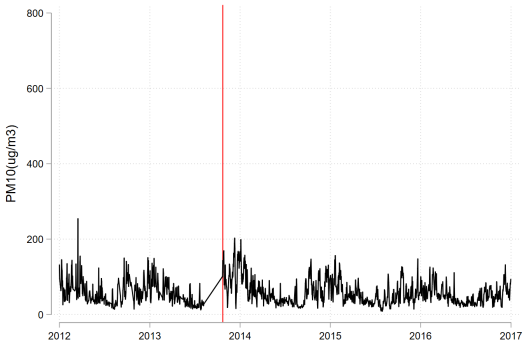
(A). Gaoxin District, Shijiazhuang City, Hebei



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



(C). Gucheng, Beijing

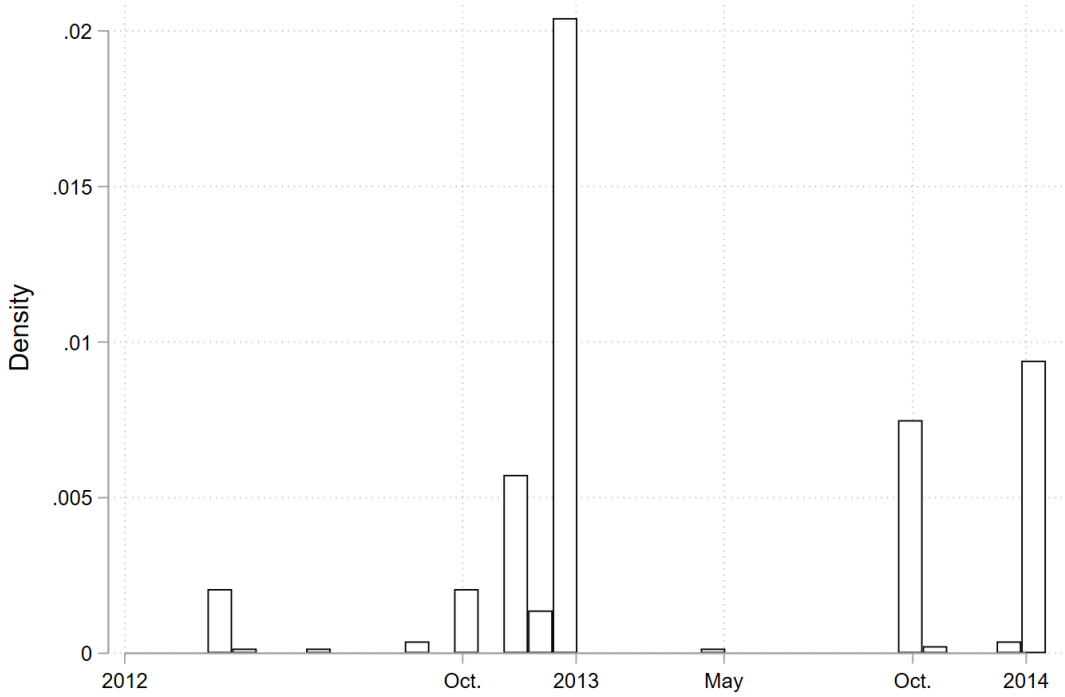


(D). Industrial Park, Beihai City, Guangxi

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

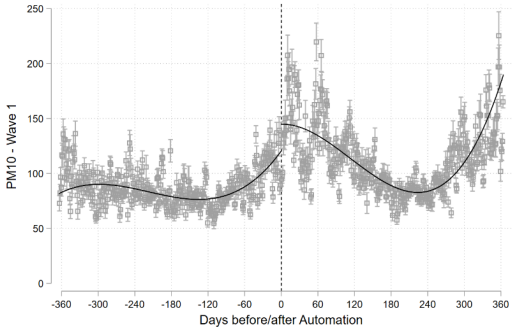


Notes: This figure summarizes the distribution of the automation dates across different cities. The majority of them automated the air quality monitoring stations on January 1st, 2013 and January 1st, 2014, which are the deadlines for the two waves.

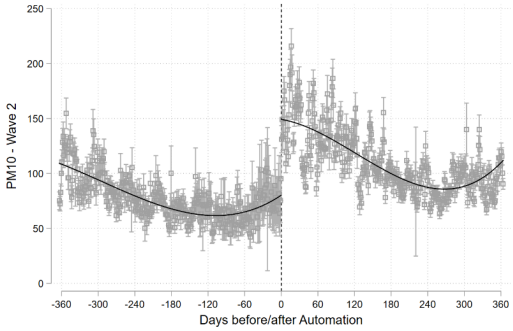
B. Additional Results on Data Quality pre-post Automation

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

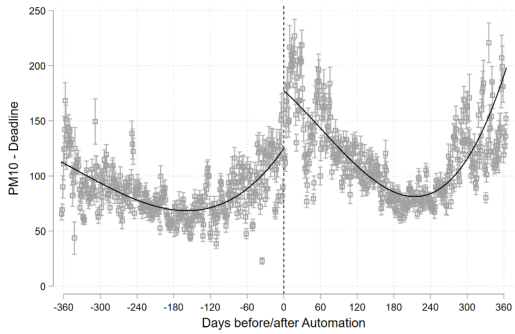
Figure B1. RD Plots Using Raw Reported PM_{10} Data



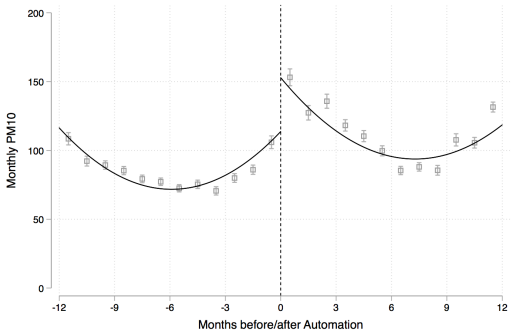
(A). Daily Reported PM_{10} in Wave 1



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



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



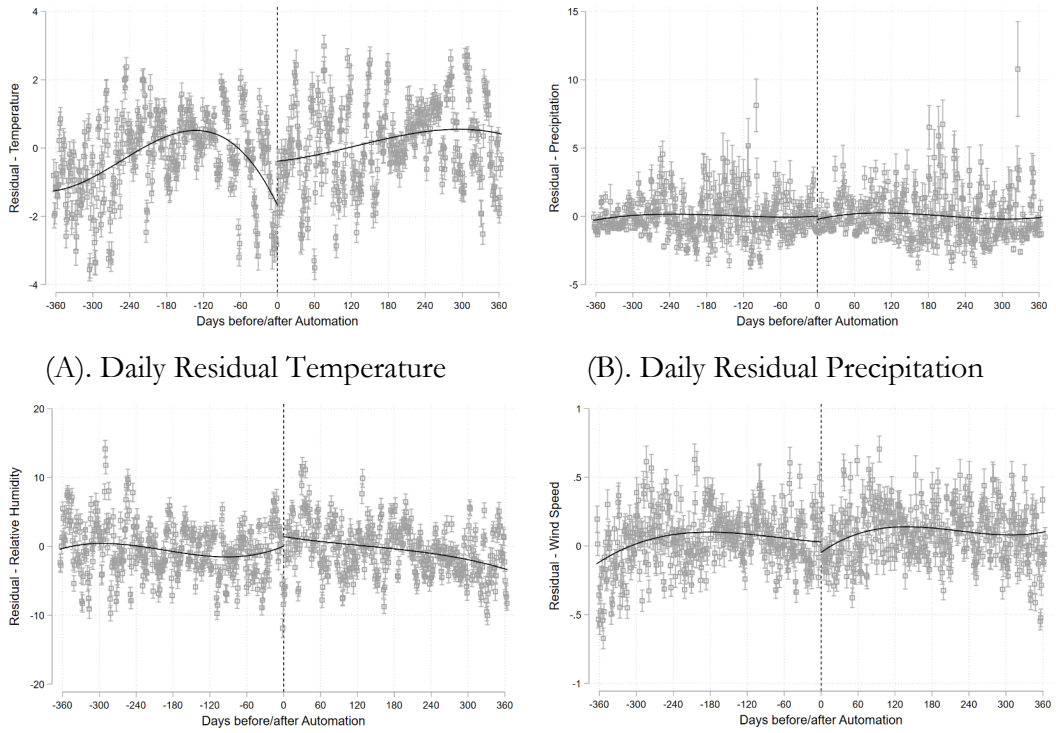
(D). Monthly Reported PM_{10}

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).

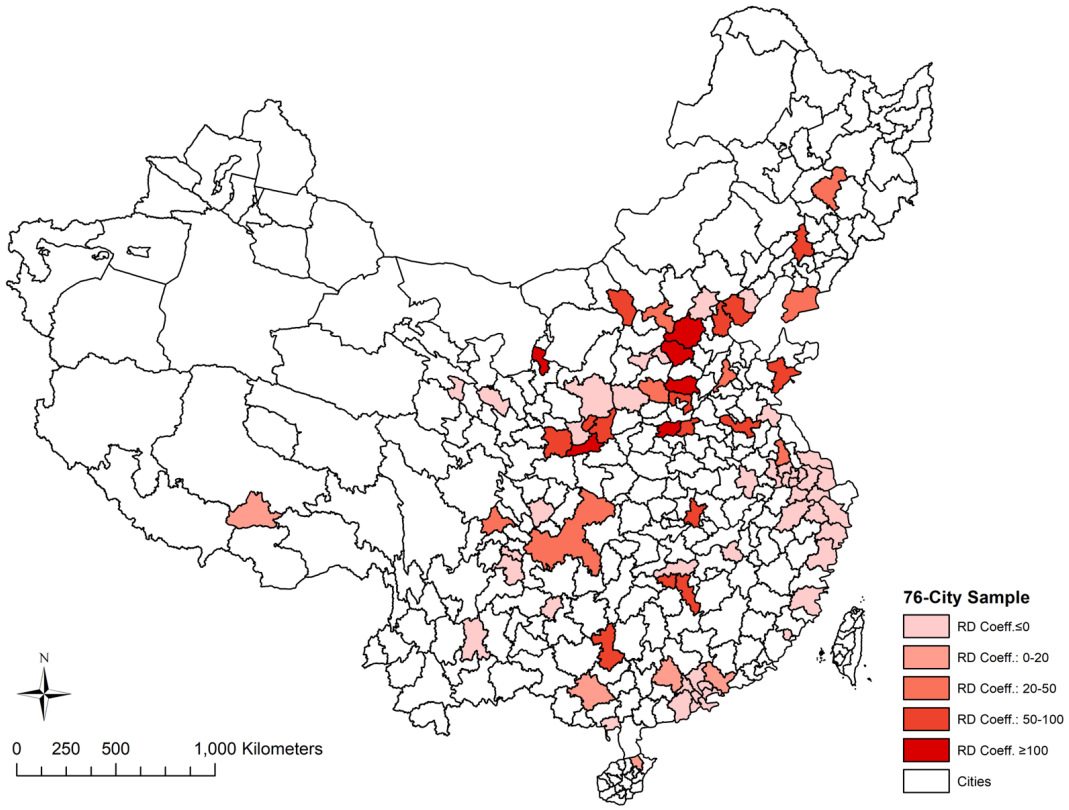
Table B4. Automation and Reported PM_{10} in 76 Cities

	(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: *E*ach 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}

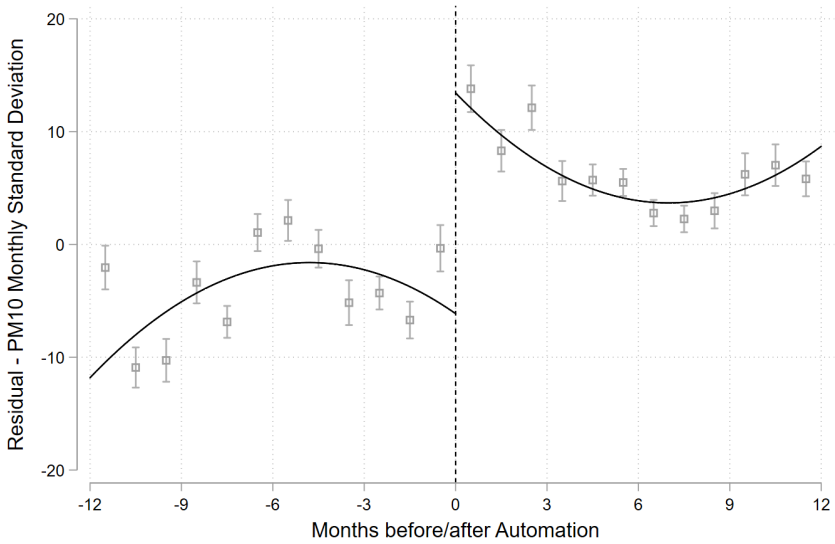
As another measure of data quality, we examine the variability of *reported* PM_{10} under the presumption that manipulated measures are likely to exhibit less variability than true realizations. We fit equations (1) and (2) by replacing the outcome variable with the monthly standard deviation of the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

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

where P_{it} is the *reported* daily PM_{10} reading at station i on day t, $\overline{P}$ is the monthly average, and n is the number of days in a month.

The graphical presentation is illustrated by 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 $\mu g/m^3$).

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} ($\mu g/m^3$)	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	0.23	0.22	0.19
		P-Value	(0.81)	(0.82)	(0.85)
		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}

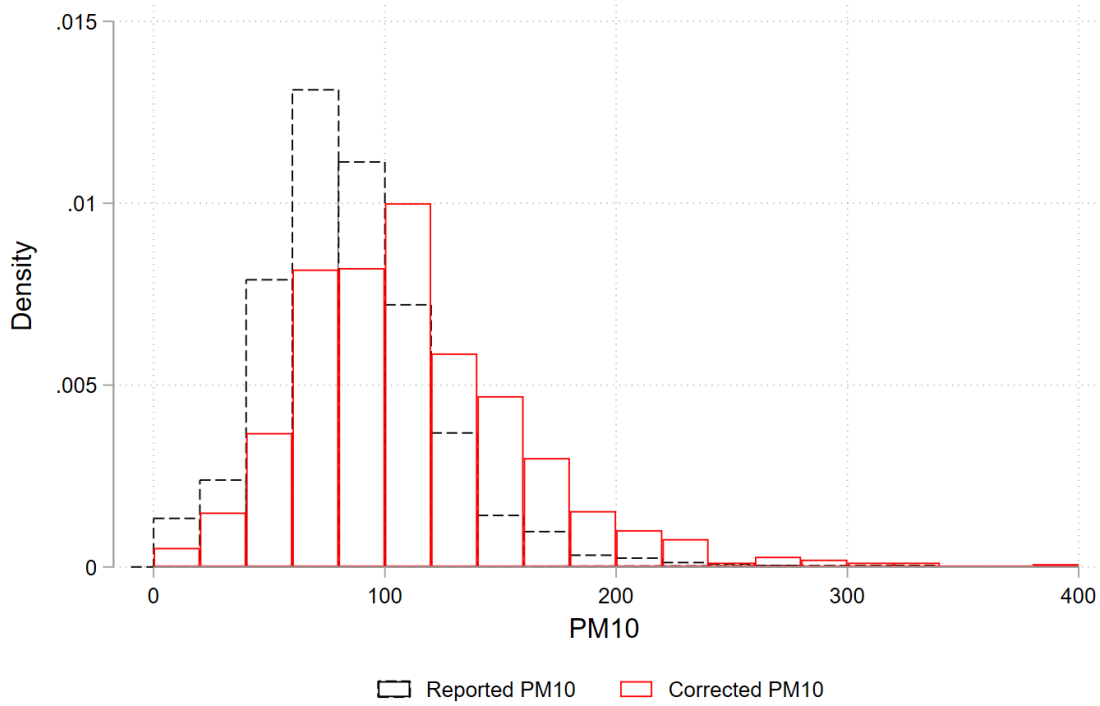
In light of the results in columns (1) to (3) of Table B10, we attempt to correct the pre-automation PM_{10} data by exploiting the relationship between PM_{10}, AOD and weather conditions (temperature, relative humidity, precipitation and wind speed). To increase our predictive power, we use an artificial 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

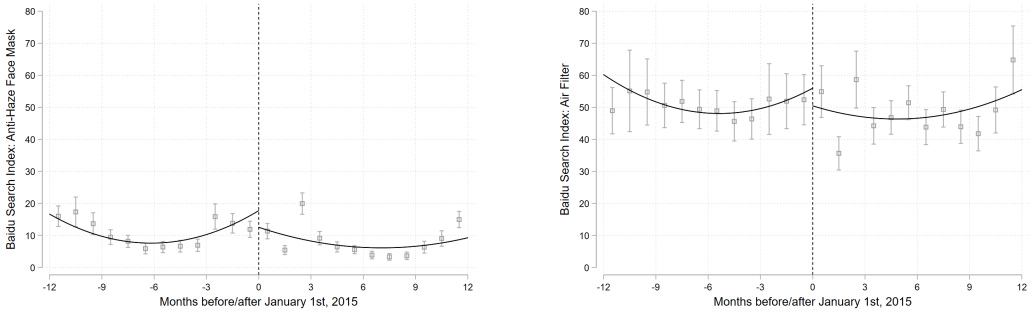
Table D1. Association between Baidu Search Index and Taobao Sales Index

	(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.