# Quantifying coal power plant responses to tighter $SO_2$ emissions standards in China

Valerie J. Karplus[a,1], Shuang Zhang[b,1,2,3], and Douglas Almond[c]

[a]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]Department of Economics, University of Colorado, Boulder, CO 80309; and [c]School of International and Public Affairs, Columbia University, New York, NY 10027

Edited by Maureen L. Cropper, University of Maryland, College Park, MD, and approved May 8, 2018 (received for review January 11, 2018)

**We evaluate the impact of China's new air pollution standards on sulfur dioxide ($SO_2$) emissions by comparing newly available data from Continuous Emissions Monitoring Systems (CEMS) at coal power plants with satellite measures. First, we show that following the July 2014 deadline for implementing tighter emissions standards, stack concentrations of $SO_2$ reported by CEMS declined by 13.9%. Second, on average the ratios of the declines of $SO_2$ measures in the satellite data and the CEMS data are about 0.5. However, the degree of correspondence between the two data sources varies by policy stringency, with weak correspondence found in key regions facing the toughest new limits. Third, large plants achieved compliance earlier than small (typically) power and heat cogeneration plants. To achieve continued air quality improvement, our results suggest a need for increased scrutiny of emissions data quality and monitoring practices and clear long-term targets.**

air pollution | satellite | high-frequency monitoring | China | policy

Severely polluted air has become pervasive in industrializing economies. While regulations increasingly incorporate the most stringent international standards, weak and uneven implementation complicates efforts to reduce emissions and improve human health (1). Data quality is also a major challenge. Reliable, high-frequency information on industrial emissions by source is essential given that acute health effects often depend on emissions timing (2–6). In developing countries, the present norm—sporadic, often manually collected, data—provides an incomplete picture of environmental performance and may be particularly susceptible to manipulation, as ref. 1 found in India. There is a growing need for studies that assess the impact of policy on polluting behavior at the firm level over time with attention paid to the plausibility of underlying data.

Our study contributes to addressing this gap by using high-frequency data from Continuous Emissions Monitoring Systems (CEMS) to evaluate how coal power plants responded to an increase in the stringency of air pollution standards in China. In 2007, the Ministry of Environmental Protection (MEP) in China required that certain (mainly high-emitting) plants install and operate CEMS (7). Early analysis of data collection efforts identified several challenges to implementation (8): inadequate local Environmental Protection Bureau (EPB) involvement in the initial testing of CEMS performance, large variation in the EPB response to data submitted, and insufficient capacity for comprehensive field inspections. From the end of 2013, 14,410 firms were required to upload hourly, automatically recorded pollutant-specific concentration data to a publicly available, online platform for each province.

Existing coal-fired power plants in China were required to comply with new emission standards (GB13223-2011) by the deadline of July 1, 2014. For power plants in relatively less polluted, nonkey regions, the limit on the maximum concentration of $SO_2$ declined from 400 to 200 $mg/m^3$. This limit was further reduced to 50 $mg/m^3$ for plants in highly polluted and populous key regions, including 47 prefecture-level cities in 19 provinces primarily located in the greater Beijing–Tianjin–Hebei area, the Pearl River Delta, and the Yangtze River Delta.

The CEMS data allow us to study the overall impact of tighter air pollution standards on plant-level emissions near the policy deadline and to probe heterogeneity in compliance behavior as a function of policy stringency and firm characteristics. Furthermore, we compare changes in plant-level $SO_2$ emissions concentrations recorded by CEMS with changes in the Ozone Monitoring Instrument (OMI) satellite $SO_2$ measures from the US National Aeronautics and Space Administration (NASA) over an area around each plant following the policy deadline. The satellite data provide an objective source for assessing changes in plant-level emissions that is not susceptible to manipulation. The OMI data have been used in many settings to measure and verify changes in air pollution emissions from point sources (see, for instance, refs. 9–13).

We report three primary findings. First, we find a large reduction (13.9%) in average plant $SO_2$ concentrations in the CEMS data in the months following the policy deadline. Satellite data also show a corresponding reduction in $SO_2$ column amounts. Second, while postpolicy reductions in CEMS and satellite measurements correspond closely in nonkey regions facing less stringent new standards, we find no such correspondence in key regions where new standards are more aggressive. Third, we explore heterogeneity in firm responses and find that the firms that fail to comply after the policy deadline tend to have smaller, older boilers and may thereby find it technically difficult or costly to introduce the end-of-pipe controls required for compliance. Our results suggest steps to improve both the quality of emissions data upon which evaluation must rely and the effectiveness of policy.

We test for confounding effects due to neighboring point sources of $SO_2$ and find that our main results are robust. Raising the capacity share threshold used to define a power plant as

## Significance

**We observe large reductions in the concentration of sulfur dioxide ($SO_2$) from coal power plants in China following the implementation of a tougher national air emissions standard using a high-frequency plant-level data source. We find a corresponding decline in $SO_2$ measures in satellite observations. However, correspondence between these two measures is lower in areas that faced a sharper increase in standard stringency.**

Author contributions: V.J.K., S.Z., and D.A. designed research, performed research, contributed new data and analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.



[1]V.J.K. and S.Z. contributed equally to this work.

[2]Present address: Department of Economics, New York University Shanghai, 200122 Shanghai, China.

[3]To whom correspondence should be addressed. Email: shuang.zhang@colorado.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1800605115/-/DCSupplemental.

Published online June 18, 2018.

isolated does not change our main results. We find no evidence that our results are driven by the operation of combined heat and power (CHP) systems during the winter heating season. Finally, we use the CEMS data to show that the behavior of nearby nonpower firms likewise does not explain our results.

## Methods

**Data Sources.** We assemble CEMS monitor-level hourly emissions data for coal-fired power plants in four Chinese provinces: Guangdong, Hubei, Shandong, and Shanghai. In each province, the provincial EPB provides a publicly available data platform to disseminate hourly emissions concentration data from monitored firms. These four provinces provided CEMS data from November 2013, 9 mo before the July 2014 policy deadline, allowing us to compare plant-level emission performance before and after the policy deadline. Our data include 256 power plants, with 43 plants whose generating capacity is larger than 1,000 MW. While plant-level configurations vary, multiple monitors are often installed on stacks associated with one or more generating units and perform hourly measurements of the concentration of $SO_2$ ($mg/m^3$) in emitted stack gases. A plant's average daily $SO_2$ concentration is measured as the average of all observed hourly values in a 24-h period from all monitors. Monthly average $SO_2$ concentration data are aggregated from daily measures by plant and are used in our analysis.

We use $SO_2$ satellite observations from NASA's dataset OMSO$_2$e: OMI/Aura Sulfur Dioxide ($SO_2$) Total Column L3 1 d Best Pixel in 0.25° × 0.25° V3. For each power plant, we draw a circle with radius of 35 km centered at the plant. We take the average of the $SO_2$ readings of all of the OMSO$_2$e grids that are partially or fully covered by the 35-km circle to capture $SO_2$ emissions from the power plant on each day. Monthly $SO_2$ measures for our analysis are aggregated from daily values. To benchmark the CEMS data to satellite data, we restrict our analysis to a sample of relatively isolated plants for which generating capacity represents at least 50% of all surrounding plants' total generating capacity within 35 km. (Our definition of relatively isolated plants is based on the distribution of plant-level capacity shares. Most power plants are located in dense industrial areas in China, and only 16% of all power plants in the CEMS data have a capacity that represents at least 50% of all surrounding plants' total capacity within 35 km.) Additional description of both the CEMS data and the satellite data are provided in *SI Appendix*.

The CEMS records $SO_2$ concentration at the point of emission from a plant, whereas the satellite column amounts capture local changes in $SO_2$ from all sources. In our analyses below, we test the robustness of our main results by taking into account $SO_2$ emissions from sources other than power plants in the CEMS data.

Summary statistics are reported in Table 1. The average stack $SO_2$ concentrations reported in CEMS are 93.6 $mg/m^3$ for all 256 plants in our sample. In key regions (113 plants), the average stack $SO_2$ concentration before July 2014 is 89.4 $mg/m^3$, and it decreases to 64.9 $mg/m^3$ after July 2014, which is still higher than the new standards at 50 $mg/m^3$. In nonkey regions (143 plants), the average $SO_2$ concentration is 170.3 $mg/m^3$ before July 2014, and it decreases to 102.4 $mg/m^3$ after the policy deadline, below the new standard. (Subsequent tables will consider the subset of geographically isolated and large-capacity power plants.)

**Literature.** Methods for comparing in situ measurements of air pollutants with satellite retrievals are well established for a range of air pollutants (see, for instance, refs. 12 and 14 for $SO_2$, ref. 9 for $NO_x$, and ref. 11 for $PM_{2.5}$). Prior studies have compared in situ measures of $SO_2$ from point sources to satellite observations over North America (12, 14–16), Mexico City (10), China (13), and globally (17, 18). The methods used here most closely follow refs. 10, 13, 14, and 17.

This study contributes to prior research in several ways. For the ground (or other direct, in situ) measures of $SO_2$, this study is the first to use plant-level $SO_2$ concentration data recorded on an hourly basis by the newly installed CEMS network in China. This data source distinguishes our study from ref. 13, which did not observe high-frequency $SO_2$ emissions data from monitors at emitting facilities. (Ref. 13 used information on power plant technology and operation to estimate monthly $SO_2$ emissions.) For the satellite measures of $SO_2$, we use the OMI data release described in ref. 14, which allows detection of smaller point sources of $SO_2$ emissions compared with earlier editions. Unlike ref. 17, which focused on quantifying long-term changes in pollutants over industrial areas, our study uses a regression approach to study correspondence in plant and satellite measurements of $SO_2$ emissions changes around the July 2014 policy deadline.

Furthermore, because China's CEMS reports hourly $SO_2$ emissions concentration, not emissions amounts, our study also differs from prior studies in that we estimate the correspondence between plant $SO_2$ concentrations and satellite measurements. As we control for conditions affecting the plant operation and any annual variation, plant-level $SO_2$ concentration is a good proxy for $SO_2$ emission amounts. We discuss the validity of this assumption in greater detail in *SI Appendix, Materials and Methods: Comparing CEMS $SO_2$ Concentration and Satellite Column Amounts*.

We expect that changes in the plant and satellite $SO_2$ measures due to the policy follow a linear relationship, as discussed in refs. 19 and 13. As a short-lived pollutant, $SO_2$ is more amenable to such comparisons than ambient $PM_{2.5}$ or ozone, which include contributions from substances formed in the atmosphere from emissions precursors. Our analysis of the satellite data controls for climate conditions from the same satellite source, as described below.

**Econometric Model.** We separately estimate changes in CEMS and satellite $SO_2$ emissions around the July 2014 policy deadline using the following specification:

$$E_{it} = \alpha P_t + \gamma_{iy} + \theta_m + X_{it}\beta + \epsilon_{it}. \quad [1]$$

Depending on the dataset we use, $i$ denotes either the plant in the CEMS data or the 35 km area surrounding the plant in the satellite data. $t$ denotes time in months ($y$ calendar year together with $m$ calendar month). The dependent variable, $E_{it}$, denotes the $SO_2$ concentration of plant $i$ in month $t$ in the CEMS data or $SO_2$ measure of the area $i$ in month $t$ in the satellite data. The key independent variable is the policy indicator, $P_t$, which equals 1 starting July 2014 and 0 otherwise. We include plant-by-year fixed effects $\gamma_{iy}$ to absorb plant or surrounding area-specific characteristics that change by year. We also include calendar month dummies, $\theta_m$, to capture seasonal variation common to all plants/areas. For analysis of the satellite observations, we also include additional controls $X_{it}$, specifically climate controls and satellite data missing counts. The climate controls include temperature, precipitation, humidity, wind speed, and wind direction. Our specification allows flexibility to capture the substantial unobserved heterogeneity in plant characteristics and circumstances, isolating common changes in the neighborhood of the policy deadline.

## Results

Fig. 1 shows the trends in demeaned $SO_2$ concentration (controlling for province and calendar month fixed effects) in the CEMS data. There is a clear trend of reductions in $SO_2$ concentrations before the July 2014 compliance deadline, and the decline

**Table 1. Summary statistics for plants included in both the CEMS and satellite datasets**

| | | Key regions | | Nonkey regions | |
|---|---|---|---|---|---|
| Measure | All | Before July 2014 | After July 2014 | Before July 2014 | After July 2014 |
| Average CEMS $SO_2$ concentration, mg/m$^3$ | 93.6 | 89.4 | 64.9 | 170.3 | 102.4 |
| Standard deviation | (78.2) | (59.8) | (46.4) | (152.7) | (66.4) |
| Average OMI $SO_2$ total column, mol/cm$^2$ | 0.347 | 0.414 | 0.336 | 0.396 | 0.331 |
| Standard deviation | (0.210) | (0.283) | (0.206) | (0.229) | (0.184) |
| $SO_2$ concentration standard, mg/m$^3$ | | 400 | 50 | 400 | 200 |

CEMS $SO_2$ concentration and OMI $SO_2$ column amounts are averages of monthly values. Standards in place by region and time period are provided for reference.

ECONOMIC SCIENCES

SUSTAINABILITY SCIENCE

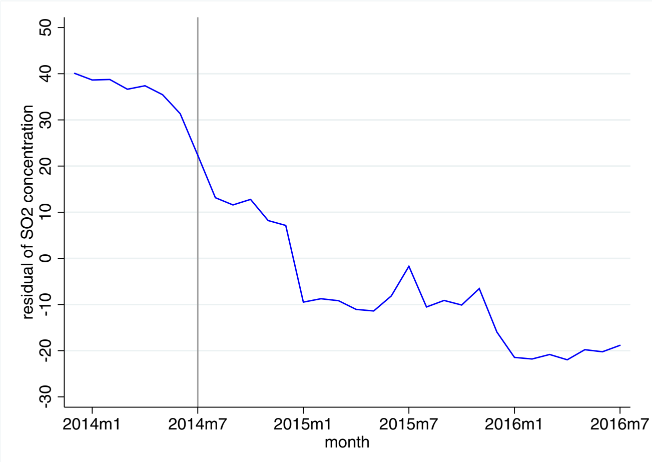


**Fig. 1.** Demeaned $SO_2$ concentrations reported in plant CEMS data.

persists after the deadline. To benchmark CEMS data against the satellite data, in Fig. 2 we focus on emissions measures for the largest plants (capacity larger than 1,000 MW) because they would be expected to generate the strongest signal in the satellite observations. The general patterns in both measures are similar, with substantial declines in the months leading up to July 2014.

We estimate changes in plant and satellite $SO_2$ measures around the policy deadline in Table 2. (All changes are estimated in log points.) Column 1 of "The policy" shows that the average decline in $SO_2$ concentration in the CEMS data after July 2014 is about 13.9%. In columns 2 and 3, to compare post-policy reductions in the CEMS and the satellite data, we restrict the analysis sample to 35 relatively isolated power plants whose capacity is at least 50% of all power plants' total capacity within a 35 km radius of the plant. In this sample, we find that enforcing the new emissions standards reduces both $SO_2$ measurements. $SO_2$ concentrations in the CEMS data fell by 36.8% after the policy, and the $SO_2$ measure from satellite fell by 18.3%. (Omitting the control variable for missing data counts in the satellite data does not change estimates of postpolicy reductions, reducing concern about potential bias from an endogenous control.) The ratio of the estimated declines of $SO_2$ measures in the satellite and the CEMS data are about 0.5 (0.183/0.368). In columns 4 and 5, we further restrict the sample to large isolated power plants (capacity larger than 1,000 MW) and find similar results.

**Falsification Test.** To address the possibility that the results in Table 1 (the policy section) were not due to the policy deadline, we repeat the analysis using the hypothetical policy timing of July 1, 2015, which serves as a falsification test as this date did not correspond to any changes in standard stringency. We use data from July 2014 to July 2016 for this test. In the falsification test section of Table 2, as expected, estimated effects on $SO_2$ concentrations from power plants are much smaller, have the opposite sign, and are less precisely estimated. The estimated changes in the $SO_2$ column amount from the satellite are very small and statistically insignificant.

**Heterogeneous Responses in Key vs. Nonkey Regions.** Given more stringent new standards in key regions, power plants may have used different approaches to adjust their emissions concentrations. Plants can alter emissions concentrations mainly by running end-of-pipe control equipment, changing the operational profile of the plant, or purchasing higher cost, low sulfur coal. Plants also differ by size and technology on the extent of reductions possible at a given cost. To compare plant responses in key regions versus nonkey regions, we first ensure that the average share of local power capacity that the isolated plants represent is very similar between key and nonkey regions: On average, plants in nonkey regions represent 82% of total capacity in 35 km, while plants in key regions account for an average of 81%.

Fig. 3 shows compliance with the prevailing standards in key regions and nonkey regions before and after the July 2014 policy deadline. Compliance is defined as the share of nonmissing hourly observations in a month in which average $SO_2$ concentrations measured by CEMS fell below the standard. In key regions, firms' compliance rate fell substantially from 100% to around 50% after the policy due to the stricter new standards. The fall in compliance after standards were implemented was on average smaller in magnitude and shorter in duration for larger plants. Larger plants constructed since 2003 (when prior standards were issued) tend to be more efficient and installed with end-of-pipe controls that remove $SO_2$ and other air pollutants. In *SI Appendix*, we show that smaller, typically CHP, plants controlled by county governments exhibited the lowest compliance rates in key regions (see *SI Appendix*, Figs. S1 and S2). In contrast, in nonkey regions where the new standards were less strict, the compliance rate decreased only slightly from 90% for a few months after the policy and then increased to close to 100% in early 2015.

Table 3 reports estimation results using Eq. **1** for key regions and nonkey regions separately. The decline in the CEMS data associated with the policy deadline is statistically significant at the 1% level. In columns 1 and 2, among all isolated plants, $SO_2$ concentrations in the CEMS data fell by 32.5% after the policy, and the $SO_2$ measure from satellite fell by 20.8%. The ratio of the estimated declines in the satellite and CEMS $SO_2$ measures is 0.64 (0.208/0.325). In columns 3 and 4, the ratio of the estimated declines is greater among plants larger than 1,000 MW, about 0.77 (0.227/0.293). In contrast, in the falsification section, we do not find such correspondence in key regions. Column 1 shows that the estimated decline after the policy in the CEMS data is 51%, larger than that in nonkey regions. If plants indeed reduced emissions as suggested by the CEMS data, we would expect the satellite data to capture the reduction to some extent. However, column 2 shows that the estimated change in the satellite data is close to zero and statistically insignificant. Similar results are found among large plants.

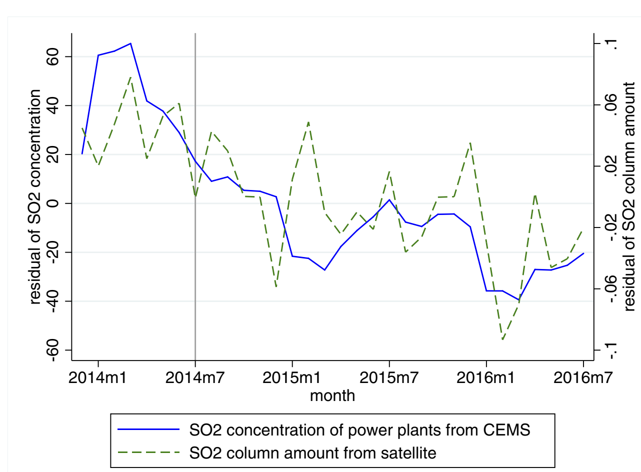


**Fig. 2.** Demeaned $SO_2$ concentrations from CEMS on power plants with capacity above 1,000 MW versus $SO_2$ column amounts from satellite observations in surrounding areas.

Table 2. **Estimated changes in CEMS and satellite $SO_2$ measures after the July 2014 policy deadline for implementing new $SO_2$ emissions standards on existing power plants**

| | All plants | All isolated plants | | Isolated plants with capacity ≥1,000 MW | |
|---|---|---|---|---|---|
| Measure | ln(plant $SO_2$) | ln(plant $SO_2$) | ln(satellite $SO_2$) | ln(plant $SO_2$) | ln(satellite $SO_2$) |
| The policy | | | | | |
| Post-July 2014 | −0.139*** | −0.368*** | −0.183*** | −0.372*** | −0.183** |
| | (0.042) | (0.100) | (0.053) | (0.111) | (0.074) |
| Observations | 5,754 | 901 | 901 | 506 | 506 |
| $R^2$ | 0.81 | 0.84 | 0.62 | 0.88 | 0.64 |
| Falsification test | | | | | |
| Post-July 2015 | 0.083* | 0.107 | −0.129 | 0.202 | −0.105 |
| | (0.043) | (0.082) | (0.154) | (0.134) | (0.243) |
| Observations | 4,539 | 697 | 697 | 384 | 384 |
| $R^2$ | 0.83 | 0.87 | 0.61 | 0.90 | 0.61 |
| Plant/Area × Year fixed effects | Y | Y | Y | Y | Y |
| Month fixed effects | Y | Y | Y | Y | Y |
| Additional controls | | | Y | | Y |

Estimated changes post-July 2015 are included as a falsification test. Column 1 uses all power plants from November 2013 to July 2016. Columns 2–5 use relatively isolated power plants whose capacity is at least 50% of all plants' total capacity in 35 km. The time period for estimating the policy effect, November 2013 to July 2016; Falsification test time period, July 2014 to July 2016 (data before July 2014 are not used in the falsification test section to avoid confounding changes following July 2014). Y, yes. Standard errors are clustered at the plant level. *$P < 0.10$; **$P < 0.05$; ***$P < 0.01$.

A potential explanation for the discrepancy in the two data sources in key regions is that plants overstated or falsified reductions. The stricter new standards and greater pressure to comply may have generated incentives for plant managers to falsify or selectively omit concentration data. In *SI Appendix*, we provide additional evidence suggestive of the potential incentives to misreport. We investigate plant emissions before and after an episode of missing data (on average 19 d) in which firms report that emissions control facilities did not operate properly. In *SI Appendix*, Fig. S3, there is little change in concentration data trends in the 15 d before and after missing data episodes in nonkey regions. In key regions after the policy, however, plants improve their reported compliance substantially after a few weeks of missing data. In addition to the possibility of falsification, plants may have reallocated production across units in ways that increased operation of dirtier units without CEMS monitors within the same facility. Both of these possibilities are consistent with our results. We also investigate alternative explanations in the next section.

**Robustness Tests.** We report the results of robustness tests in *SI Appendix*. First, because we focus on power plants whose capacity represents at least half of all power plants' capacity within 35 km, changes in $SO_2$ emissions in surrounding power plants or heating boilers that are part of CHP plants might generate noise when we compare changes in emissions in the CEMS data with those in the satellite data. Second, a potential concern about our findings in Table 3 is that if $SO_2$ emissions at nonpower firms in key regions increased after July 2014, it could have offset a possible decrease in $SO_2$ emissions of power plants nearby in the key regions, and therefore, little change in the $SO_2$ measure from the satellite is observed.

To address the issue posed by surrounding power plants, we further restrict the sample to power plants whose capacity is at least 70% of all power plants' total capacity within 35 km. (Eleven percent of all power plants in the CEMS data have a capacity that represents at least 70% of all surrounding plants' total capacity in 35 km.) Confounding changes from surrounding power plants, if any, should be smaller in this subsample. Results are reported in columns 1 and 2 in *SI Appendix*, Table S2. We find results similar to those in Table 3. Notably, in nonkey regions, we find a 26.8% reduction in $SO_2$ concentration in the CEMS data and a 27.0% reduction in the satellite data. The ratio of the estimated postpolicy declines of $SO_2$ measures from the satellite data and the CEMS data are nearly proportional (0.268/0.270, or approximately 1). This represents a much stronger correspondence between the two data sources where possible errors from surrounding power plants are further reduced.

To address the concern about winter heating, we note that heating boilers operate only in winter months (November to March), and the deadline for complying with the new emission standards was in July 2014. Therefore, we would not expect the change in emissions before and after the July deadline to have been affected by emissions from heating boilers. We also directly examine this concern by dropping winter months in the sample and report results in columns 3 and 4 of *SI Appendix*, Table S2. We find that results are robust to those shown in Table 3.

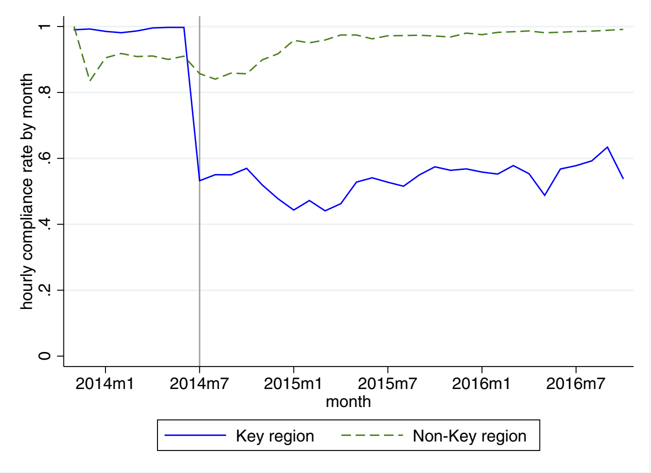


**Fig. 3.** Plant-level hourly compliance rates by month in key regions (blue solid line) and in nonkey regions (green dashed line) before and after the policy deadline in July 2014 (indicated by a vertical line).

**Table 3. Estimated changes in CEMS and satellite $SO_2$ measures after the July 2014 policy deadline shown separately for key and nonkey regions**

| | All isolated plants | | Isolated plants with capacity ≥1,000 MW | |
|---|---|---|---|---|
| Measure | ln(plant $SO_2$) | ln(satellite $SO_2$) | ln(plant $SO_2$) | ln(satellite $SO_2$) |
| Nonkey regions | | | | |
| Post-July 2014 | −0.325*** | −0.208*** | −0.293*** | −0.227** |
| | (0.070) | (0.074) | (0.104) | (0.105) |
| Observations | 711 | 711 | 360 | 360 |
| $R^2$ | 0.70 | 0.61 | 0.68 | 0.59 |
| Key regions | | | | |
| Post-July 2014 | −0.509** | 0.004 | −0.536* | 0.016 |
| | (0.239) | (0.100) | (0.311) | (0.104) |
| Observations | 190 | 190 | 146 | 146 |
| $R^2$ | 0.88 | 0.68 | 0.84 | 0.73 |
| Plant fixed effects | Y | Y | Y | Y |
| Area × Year fixed effects | Y | Y | Y | Y |
| Month fixed effects | Y | Y | Y | Y |
| Additional controls | | Y | | Y |

All columns use power plants whose capacity is at least 50% of all plants' total capacity in 35 km. Time period is from November 2013 to July 2016. Standard errors are clustered at the plant level. Y, yes. *$P < 0.10$; **$P < 0.05$; ***$P < 0.01$.

Finally, to address the concern about nonpower facilities, we directly investigate the change in $SO_2$ concentration of nonpower, high-emitting facilities before and after the July 2014 deadline for power plants. For each power plant in our data, we obtain CEMS data for nonpower firms within 35 km. In *SI Appendix*, Table S3, we do not find statistically significant changes in $SO_2$ concentration at nonpower plants in both key and nonkey regions, consistent with the fact that the July 2014 deadline did not apply to nonpower facilities. The sign of the change is negative in key regions, which further reduces the concern about a potential increase in $SO_2$ concentration at nonpower firms in key regions. Results suggest that the limited correspondence between CEMS and satellite measures of $SO_2$ cannot be explained by changes in the emitting behavior of neighboring nonpower firms. To summarize, these additional findings suggest that our main results are robust and that manipulation of the CEMS data in regions facing the toughest emission standards is plausible.

## Conclusion

We find that tighter environmental standards targeting power plants' emissions concentrations prompted substantial $SO_2$ reductions at coal power plants in China. Satellite data corroborate these reductions in nonkey regions but not in key regions, which faced deeper reduction requirements. Reductions recorded by CEMS were large: Average $SO_2$ emissions concentrations at power plants were 13.9% lower after the policy deadline. The absence of a change in satellite measurements at key regions' plants around the policy deadline is suggestive of misreporting. Prior studies documenting efforts to improve $SO_2$ policy enforcement in China have raised questions about the quality of the CEMS data (see, for example, refs. 7 and 20). Penalties for falsifying data have historically been lower than those for violating the standard. Regulations initially offered neither detailed guidance on CEMS reporting requirements, instructions on how to identify falsification cases, nor a schedule of associated penalties for reporting violations (see *SI Appendix*, Table S1 for a description of penalties for standard violations and data falsification). Our findings based on the satellite comparison further suggest that manipulation is far from universal but may have increased as compliance grew more costly or difficult.

The drop in $SO_2$ concentrations around July 2014 suggests that the policy deadline prompted a substantial change in the emitting behavior of most firms, such as increasing utilization of end-of-pipe control. However, potential reductions from these short-term levers are constrained largely by the vintage of installed equipment. Many plants, especially in key regions, likely required substantial changes to their pollution removal technology or operational practices to meet the tighter standards. Despite a large drop in $SO_2$ emissions concentrations, many firms ultimately fell short. As of June 2016, nearly half of the plants in key regions had not complied with the strictest new standard, and many smaller plants did not have CEMS installed.

Our results suggest substantial room to strengthen incentives for accurate and comprehensive reporting as part of China's national air pollution control efforts. An important first step involves clarifying reporting requirements and strengthening penalties for data inaccuracies or falsification. For instance, officials could mandate that CEMS measure concentrations in the stack gases of all (and not just a subset of) operational boilers and that CEMS record emissions during all of a plant's operational hours. A second step could involve allowing more time for emitters to comply with tougher standards but signaling strong enforcement. Emissions trading systems could help to alleviate high costs for some plants and provide compliance flexibility; however, for these systems to work, high-quality emissions reporting is essential. As a third step, where feasible, plants that contribute disproportionately to local pollution and lack cost-effective pollution control options could be shut down as part of efforts to reduce substantial existing excess capacity in China's power sector. Increasing the reach and accuracy of CEMS monitoring can help to support air quality improvement.

The method applied in this paper could be extended to assess cleanup efforts for other short-lived industrial air pollutants. $NO_x$ was added as a regulatory target during the Twelfth Five-Year Plan; both CEMS and satellite data are available for $NO_x$. Comparisons of CEMS and satellite measures could also be used to alert policy makers to major instances of CEMS data manipulation, extending the growing application of remote sensing data in a regulatory setting (19). Finally, an approach like ours could be applied in other developing countries that seek low-cost tools to ensure data quality and policy compliance and thereby improve air quality.

**ACKNOWLEDGMENTS.** We thank Yi Cheng, Ruinan Liu, Lucy Lu, Yuxuan Mei, Jing Qian, Minghao Qiu, Lin Shi, Mandy Wu, Frank Yang, and Michelle Zheng for research assistance. V.J.K. acknowledges the support of a seed grant from the Samuel Tak Lee MIT Real Estate Entrepreneurship Laboratory. S.Z. and D.A. thank the National Science Foundation for support through Award SES-1658888, "Collaborative Research: Market Based Emissions Policies," in China.

1. Duflo E, Greenstone M, Pande R, Ryan N (2013) Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India. *Q J Econ* 128:1499–1545.
2. Cheng N, et al. (2017) Spatio-temporal variations of PM2.5 concentrations and the evaluation of emission reduction measures during two red air pollution alerts in Beijing. *Sci Rep* 7:8220.
3. Shen Y, et al. (2017) Non-linear increase of respiratory diseases and their costs under severe air pollution. *Environ Pollut* 224:631–637.
4. Davis LW, et al. (2017) Saturday driving restrictions fail to improve air quality in Mexico city. *Nat Sci Rep* 7:41652.
5. Liu Z, et al. (2015) Reduced carbon emission estimates from fossil fuel combustion and cement production in China. *Nature* 524:335–338.
6. Dockery D, Pope A (1994) Acute respiratory effects of particulate air pollution. *Annu Rev Public Health* 15:107–132.
7. Schreifels JJ, Fu Y, Wilson EJ (2012) Sulfur dioxide control in China: Policy evolution during the 10th and 11th five-year plans and lessons for the future. *Energy Policy* 48:779–789.
8. Zhang X, Schreifels J (2011) Continuous emission monitoring systems at power plants in China: Improving SO2 emission measurement. *Energy Policy* 39:7432–7438.
9. Bechle MJ, Millet DB, Marshall JD (2013) Remote sensing of exposure to NO2: Satellite versus ground-based measurement in a large urban area. *Atmos Environ* 69:345–353.
10. de Foy B, et al. (2009) Hit from both sides: Tracking industrial and volcanic plumes in Mexico city with surface measurements and OMI $SO_2$ retrievals during the MILAGRO field campaign. *Atmos Chem Phys* 9:9599–9617.
11. van Donkelaar A, et al. (2016) Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environ Sci Technol* 50:3762–3772.
12. Fioletov VE, McLinden CA, Krotkov N, Moran MD, Yang K (2011) Estimation of $SO_2$ emissions using OMI retrievals. *Geophys Res Lett* 38:1–5.
13. Wang S, et al. (2015) Satellite measurements oversee China's sulfur dioxide emission reductions from coal-fired power plants. *Environ Res Lett* 10:114015.
14. Fioletov VE, McLinden CA, Krotkov N, Li C (2015) Lifetimes and emissions of SO2 from point sources estimated from OMI. *Geophys Res Lett* 42:1969–1976.
15. Fioletov V, et al. (2017) Multi-source $SO_2$ emission retrievals and consistency of satellite and surface measurements with reported emissions. *Atmos Chem Phys* 17:12597–12616.
16. Kharol SK, et al. (2017) OMI satellite observations of decadal changes in ground-level sulfur dioxide over North America. *Atmos Chem Phys* 17:5921–5929.
17. Krotkov NA, et al. (2016) Aura OMI observations of regional SO2 and NO2 pollution changes from 2005 to 2015. *Atmos Chem Phys* 16:4605–4629.
18. Fioletov VE, et al. (2016) A global catalogue of large SO2 sources and emissions derived from the Ozone Monitoring Instrument. *Atmos Chem Phys* 16:11497–11519.
19. Duncan BN, et al. (2014) Satellite data of atmospheric pollution for U.S. air quality applications: Examples of applications, summary of data end-user resources, answers to FAQs, and common mistakes to avoid. *Atmos Environ* 94:647–662.
20. Xu Y (2011) Improvements in the operation of SO2 scrubbers in Chinas coal power plants. *Environ Sci Technol* 45:380–385.